Gains from bilateral cooperation –
A tentative research agenda

One of the most prominent themes in the study of international relations in modern times has been the issue of how to assess cooperation between sovereign states. Is the achievement of cooperation between sovereign states reflections of altruism or pursuit of self-interest? Under what circumstances should we expect to find cooperation, and when should we instead predict a breakdown of negotiations and the loss of potential mutual gains? In this paper, an attempt is made to map out the potential for cooperation between two or more sovereign states. The emphasis is put on theoretical and general factors influencing cooperative behaviour, rather than on the particularities of interactions between Poland, Sweden, and the European Union. By pursuing this theoretical approach, at least three important advantages are achieved. First, by using a more general framework for analysis, theoretical insights gained by other researchers become useful instruments in the identification of factors that might not be easily observable by studying only empirical data from the Polish-Swedish case, or from the Polish EU accession case. Second, a general framework for the identification of potential gains from bilateral cooperation makes it possible to use the same model for analysing the pursuit of joint gains in different empirical areas. Third, by elaborating a framework in close relation with contemporary theoretical models of interstate interaction, spurious correlations and invalid ad-hoc explanations are easier to identify and discard from further analysis. The framework selected in this article is the rational choice approach. This approach is the most general approach used in international politics, which makes it particularly well
solved for a broad analysis where different aspects of gains and losses from international cooperation can be integrated.

It is assumed in this article that the actors (i.e. the states) can be assumed to be unitary actors. This is a heroic assumption, since the fundamental unit of analysis in rational choice theory is the individual. The rationality of collectives such as states does not follow from the assumed rationality of the individual, since preferences of collectives might very well be intransitive (Hassler 2000). If, however, the preferences of the sub-national actors can be assumed to be reasonably homogeneous, than the unitary actor assumption can be defended, since the larger the homogeneity in preferences, the smaller the probability for intransitivity. In cases with a low level of homogeneity, the domestic situation must be modelled as a game before the interaction between states can be analysed. The outcome of the game in the domestic arena will then be the input in the international game, thus giving us a so-called two-level game (Tsebelis 1990). In this article, however, the benefits and costs from cooperation are only described in general and very abstract terms, which means that the issue whether the actor can be assumed to be a unitary actor will not be considered. The states are here simply assumed to be unitary actors.

It should be emphasized that the theoretical discussion of potential gains from bilateral cooperation is not limited to cooperation between two countries only. The framework laid out here could be used for gauging potential gains from collaboration between any numbers of actors. Unless otherwise explicitly stated, the arguments presented below are valid irrespective of the size of the set of countries. This framework could be used for assessing the

---

1 In the early work on collective action, some general propositions were made regarding the relation between the number of actors and the achievement of collective goals. Mancur Olson, for example, suggested that collective action would be easier with smaller sets of actors, ceteris paribus (Olson 1965). Later research has showed, however, that such simplistic propositions are not generally valid. The key aspect in this regard is rather the individual incentives facing the actor, and these incentive patterns are in most cases not related to the size of the set. In other words, findings on models with only two actors is in most cases possible to generalize also to larger sets of actors.
potential gain from bilateral cooperation between Sweden and Poland, but it could equally well be used to uncover rationales behind observations of European Union cooperation in general. In order to increase parsimony, however, most arguments will be offered in terms of interaction between only two actors, since this makes the discussion more lucid.

After an assessment of the relevance of the collective good concept in bilateral relations, the Prisoner’s Dilemma will serve as a starting point for further elaboration. It is then shown that parts of the cognitive prospect theory approach can be used as a supplement to the traditional theory of rational choice.

From hegemony to cost-benefit assessment

In all situations where joint gains are attainable through interaction between two or more actors, issues of collective choice are relevant. The accomplishment of goals that requires action from others as well as oneself always has a strategic component. Suppose, for example, that a cleaner Baltic Sea is positively valued by Poland, as well as by Sweden. It is quite clear then, provided that the actors can be analytically perceived as mainly interested in their own welfare and that the costs of the needed mechanisms to curb pollution are borne by the actors individually, both Poland and Sweden would prefer that the other party would decrease pollution as much as possible, irrespective of the other’s costs. The reason is that benefits from decreased pollution of the Baltic Sea are collective. Although these benefits in most cases are unevenly distributed between the actors, action taken by one of the actors alas influences the other. If this were not the case, this issue would be an internal affair, in the strict sense of the word, of the actor undertaking the initiative. In other words, the existence of a collective good is a precondition for cooperation to be of any interest to the actors in this context, provided
that they are only interested in their own well-being.\footnote{The mirror image of joint gains (collective goods) is collective bads. Since collective goods and bads from the perspective of incentives for choosing cooperative strategies are identical, however, only the collective good label is used.} And where there is a potential gain from cooperation between self-interested actors, there are also incentives for strategic behaviour. Gains and costs from cooperation can always be distributed in different ways, a fact that profoundly influences the actors’ choice of optimal strategies (Knight 1992).

Strategic behaviour related to the potential for mutual gains from cooperation abounds in relationships between states. A few examples from various issue-areas are trade negotiations, where all actors want the others to reduce their trade barriers, while not necessarily wanting to do this themselves, foreign security issues, where increased strength of an allied is typically positive, while net gains from increasing its own strength might very well be negative, environmental cooperation, where positive achievements from others are always positive. Despite the many differences between such diverse issue-areas, the fact that they can all be viewed through the twin concepts of collective goods and strategic choice makes it possible to analyse them with a single, theoretical model.

Dating back to the classical works of Thomas Hobbes and David Hume respectively, the nexus of the cooperation problem has been centred around the issue whether self-interested, rational, and sovereign actors need an external enforcer to realize mutual gains, or if self-enforcing institutions without such an enforcer might evolve or be created (Hobbes 1968; Hume 2000). The Hobbesian argument, stating that an external enforcer is required, was in the context of international relations picked up by the realist school, and later by the early neo-realism school. A leading power, a hegemon, was, depending on one’s interpretation, needed either to compel the actors to choose cooperative strategies, or to make cooperative outcomes sufficiently likely to make the actors perceive cooperative strategies as furthering their own self-interest (Keohane 1984). The Humean argument emphasizing the potential for cooperation without an external monitoring agent was elaborated by, for
example, Michael Taylor, Elinor Ostrom, James Morrow, and Bruce Bueno de Mesquita. These authors all made extensive use of the instruments of game theory. Although the neo-realism approach in most cases was also built upon a game theoretic foundation (e.g. the Prisoner’s Dilemma model), game theory served here more as a heuristic tool than as a stringent analytical instrument.

During the last decade or so, the issue whether a hegemon is needed to foster cooperation between states has become less pronounced. Although the existence of a strong actor might facilitate cooperation, mainstream analysis has gravitated towards the Humean view, where interstate cooperation is explained rather as the outcome of self-interested strategies, than by coercion exerted by a hegemonic state. The reason to this increased emphasis put on voluntary cooperation is twofold. First, during recent decades, US capability has probably decreased relatively in most areas, with a possible exception for military capability (Keohane 1984). This does not mean that the US is not the most powerful country in a large number of areas, but rather that other actors have narrowed the gap between them and the United States. In relation to, for example, Japan and the European Union, the US share of total economic production has decreased. Even though it is questionable whether the EU can be regarded as a single actor, it is nevertheless true that the combined GNP of Europe is nowadays considerably larger in relation to that of the US compared with the situation a few decades ago.

Since the focus of interest has shifted from being almost exclusively on military capability during the Cold War, to being either a power assessment based on capability in different areas or issue-specific capacity, the concept of hegemonic power cannot capture this increased complexity. Most issues in most regions were to a considerable extent influenced by the conflict between the United States and the Soviet Union. Because the foreign security

---

3 The term voluntary cooperation in this context is analytic rather than empirical. A situation that might be perceived as coercive in ordinary language could still be modeled as a voluntary agreement as long as the “coerced” actor can choose between at least two alternatives, how ever bad these are.
issues were so prominent in international relations during this time, military capability became the natural focal point for analysis of interaction between states. Today, however, issues where the application of military power in most cases is not even considered, such as for example environmental, trade, and migration matters, have taken a higher place on the international agenda. There are issues that might have foreign security aspects attached to them, but these aspects are not of central importance for the understanding of the outcome of negotiations, conflict management, and the like. Somewhat paradoxically then, despite the fact that the US is the only superpower at present, its power in several areas has diminished.

While the first reason for the decreased focus on hegemony was related to changes in the international system, the second was intra-scientific, and caused by the refinement of existing theories in international relations. The first attempts in the rational choice tradition to model issues related to hegemonic stability and institutional change in the international system were based on comparably simple games such as one-shot Prisoner’s Dilemmas (Keohane 1984). As models of international issues as games became more elaborate, however, the focus changed from hegemonic competition and towards models emphasizing assumptions of utility-maximizing behaviour by the states. The issue was now not so much whether a particular state had the necessary capability to influence other states behaviour, but rather whether they wanted to do so or not. The point is although an actor, a state for example, would like to influence some other state to change its behaviour, such influence seldom comes without a cost. This cost is in most cases not a monetary cost, but rather a loss in terms of lower expected gains from future interaction with this and other actors who have observed their interaction. The calculation before attempting to influence another actor must, in other words, be based upon a cost-benefit analysis. The actor can only be assumed to attempt to influence another actor not only if the gain is larger than the cost, but also if the expected net gain is larger than what could be expected from other alternatives.
Defining international institutions

The study of institutions has attracted huge interest among scholars in international relations. The primary issue has been to better understand the pattern of interaction between two or more actors. Why do states accept certain rules and norms guiding behaviour, thereby decreasing their freedom of action? What characterizes the rules and norms that become observable patterns or formalized into conventions or international laws? Are the institutions we observe optimal responses, or merely anything preferable to completely uncoordinated behaviour?

It seems rather clear that despite the considerable efforts to explain the emergence and continued existence of international institutions, no generally accepted explanations have been put forward. In fact, not even a precise definition of what constitutes an international institution has emerged, or rather; various scholars have suggested a plethora of definitions. None of these have taken clear precedence over the others. Other concepts, such as for example international regimes, have furthermore been introduced, although the difference between regimes and international institutions seems unclear (Haas 1980; Ruggie 1982; Krasner 1983; Young 1991; Levy 1995; Feinberg 1999).

A rather broad understanding seems to have emerged within rational choice analysis of institutions, however, of the advantage of distinguishing between organizations and institutions (North 1998). Organizations are physical entities such as, for example, the World Trade Organization (WTO) or the European Bank of Reconstruction and Development (EBRD), while institutions are either structures that restrict the freedom of action of the actors or the outcome of the interaction between actors. Organizations can be actors, but never structures, while institutions can never be actors. When institutions are perceived as restrictions put on the actors, these restrictions serve to delimit the number of available strategies among the actors, thus making it easier to coordinate their behaviour. In other words, the institutional structure increases the information available to the actors. Without an
institution, the formation of mutual expectations about the behaviour of others becomes more difficult (Rogowski 1999).

Institutions are typically beneficial to most of the actors involved. Without them, coordination and the potential of the actors to make reasonably probable predictions about other actors’ choice of strategies would be most difficult. These difficulties would produce losses, since the cost of obtaining additional information about the preferences of other actors would decrease net benefits from any potential concerted action. The fact that an institution is valuable to the actors concerned is not, however, an explanation why it has emerged. Sometimes this functional fallacy has led observers to jump to the conclusion that the materialization of an institution has been explained. But this is not a valid conclusion, since it would not necessarily be rational for the individual to contribute to the creation of an institution, even though he might gain from its existence. The simple explanation is that the same individual would gain even more, if others bore the costs of institution creation, provided that the benefits generated by the institution are collective (Olson 1965).

Several scholars - Hayek is probably the most often cited - have suggested that if we assume for the moment that an institution has emerged, this institution would be functionally optimal (Hayek 1978). The mechanism making it optimal is, in simple terms, evolutionary pressure. In analogy with the Alchian notion of competition among firms in a perfect market, institutions are assumed to be evolutionary pressured into optimality. Those not adapted to their environment will succumb to the competition among institutions. The result would thus be an optimal solution, that is, no other kinds of institutions would be preferable from a societal point of view.

It is not obvious, however, that the analogy between firms in a market and institutions in the international system is valid. In fact, they could be understood as opposite outcomes of a Prisoner’s Dilemma game. In a perfect market, the firms have strong incentives to cooperate with each other, to collude. The most lucrative situation for each of them would be to achieve a monopoly status. However, since consumers are assumed to always choose a given
product at the lowest price, competition among the firms will force them to sell at marginal cost. The perfect market is, in other words, the result of a failure among the firms to cooperate. Flourishing institutions, on the other hand, are the results of successful cooperation between relevant actors. The analogy with the market would thus rather be with a market failure, where the actors are colluding effectively, since the corollary to the creators of the institution is the firms, and not the consumers. If it were true that institutions are not exposed to an evolutionary pressure that over time makes them optimal, we have no reason to believe that social optimality would be an accurate characterization of existing institutions.

Because of the lack of valid micro foundations explaining the behaviour of the individual actors in several of the most common approaches (e.g. regime analysis, epistemic communities, traditional realism), neither the emergence of international institutions, nor the adaptation of them to changing external conditions have been adequately explained. The most important reason for this is that the individuality of the actor’s interests has not been modelled in a consistent way. Outcomes at the macro level have often not been logically deduced from micro assumptions regarding individual actors. This is somewhat odd, since the explanatory power of neoclassical economics, for example, is directly derived from such a separation between analytical levels. The functions of a perfect market are, very simplified, explained by the self-interest of the individual producers. The functions of international institutions have not, however, similarly been based on the incentives facing the individual actor. Consequently, the characteristics of international institutions have not been sufficiently explained.

Recently, however, adequate micro foundations designed to explain macro observations have been suggested (Knight 1992; Sened 1998). This approach takes the assumptions about individual rationality and utility-maximizing behaviour as its point of departure. It is assumed that the individual actors will only contribute to the creation or change of institutions if it is in their private interest to do so. This means that individual actors will in general not contribute to the creation of institutions aimed at creating collective goods,
since they are trapped in collective action dilemmas. They will only do so in so far as (1) private net gain from collaboration is positive and (2) the strategic situation does not make it rational to free ride. The individual actor will thus not intentionally create collective good institutions. Instead, the emergence of institutions is a side effect of rational actors’ seeking private gain (Knight 1992). These institutions would in the characteristic case not be optimal societal solutions, but rather reflections of actors’ preferences and nature’s restrictions.

**Institutions as strategic equilibria**

Game theory is a powerful tool to model interdependent interaction between actors that are assumed to act in accordance with utility-maximizing criteria. Not only can complex single situations be formulated as games, but also sequential and repeated types of interaction can be formalized (Kreps 1990). One of the most fertile areas of research during the last two decades or so has been the application of rational choice theories to the analysis of institutions (Knight 1992; Calvert 1998; Sened 1998). It has been suggested in this field of study, that institutions can be defined as strategic (Nash) equilibria in an iterated sequence of underlying games (Schotter 1981; Calvert 1998).⁴ Firmly in line with traditional rational choice assumptions, the emergence of institutions can thus be explained as the result of various actors’ pursuit of individual self-interest in the setting that nature endows them with. The individual actors choose strategies in accordance with their private, self-interested utility functions. When it is in their private interest to choose strategies that stimulates the creation of institutions they will do so, otherwise they will not.

This will be elucidated by the following example on trade strategies, which is limited to only two actors in order to simplify the line of reasoning. The general conclusions, however, are also valid for larger sets of actors. It

---

⁴ In modern game theory terminology, Nash equilibria have been relabeled to strategic equilibria, since the latter gives clearer connotations as to what is actually meant. Both terms are, however, in current use.
could be noted that precisely this example, where an actor cannot choose trade strategy without taking other actors’ choices of strategy into consideration, was given attention to by one of the founding fathers of modern economics, Antoine Augustin Cournot, as early as in 1838:

The question will no longer be the same if the establishment of a threshold for the benefit of A producers might provoke, by way of retaliation, the establishment of another threshold for the benefit of B producers, against whom the first threshold was raised. The government of A would then have to weigh the advantage resulting from the first measure to the citizens of A against the drawbacks caused by the retaliation. (Cournot 1899:164. Cited in Shubik 1989:121)

The tools of game theory had at that time, however, not yet been developed. The same analytical situation will therefore be presented below, but here it is described and analysed within a game theoretical framework.

Suppose two states, say Sweden (S) and Poland (P), are negotiating a trade agreement. For simplicity, it is assumed that these two states have no history of trade agreements, which in this case means that their behaviour is not influenced by previous interactions. They both know that it would be preferable for each of them if a free trade (FT) agreement was reached, compared with a breakdown in negotiations which would result in no agreement reached, and mutual trade barriers (TB) remaining unchanged. The most preferable outcome to both states, however, would be if the other state adopted a FT policy while keeping TB for itself. The worst outcome for both states would be if a FT policy were adopted, while the counterpart kept its TB policy. This hypothetical situation is an example of a Prisoner’s Dilemma and is depicted in Figure 1 below.
The unique strategic equilibrium in this one-shot game of Prisoner’s Dilemma is when both states choose their TB strategies.\(^5\) The interesting feature of this outcome is its Pareto inferiority, that is, both actors would benefit from choosing their FT policy.\(^6\) Because of both actors’ fear of being caught in the worst situation where the counterpart matches a FT policy with a policy TB, they have no rational alternatives but to choose their TB strategies.

This one-shot Prisoner’s Dilemma, however, is too simplified to serve as a model of reality. It should rather be perceived as a heuristic tool, describing one of the most important underlying dilemmas of cooperation on collective

\(^{5}\) A strategic equilibrium is defined as an outcome where no actor can benefit from unilaterally changing strategy.

\(^{6}\) It is assumed in this normal form game that both actors choose strategies simultaneously. If the same situation were depicted as a sequential game, where one of the actors makes its move first, the outcome, however, would be the same. In this case, the first actor would not dare to choose FT, since this would give the other actor the opportunity to choose TB, thereby creating the worst outcome for the first actor (Hirshleifer 1989).
In reality, most interactions are not occasional and independent of each other. Interaction between states could more realistically be modelled as a sequence of iterated games, for example, of Prisoner’s Dilemmas. Not only would this make each actor’s choice of strategy depend on projections of the repercussions from alternative strategies on future outcomes from other interactions, but it would also serve as a signalling device to other actors. By observing the behaviour of these actors, other states would receive information that would potentially affect their strategies in future interactions with those actors. This has to be taken into consideration when choosing from among available alternatives in the first place.

The change from a single-shot to an iterated Prisoner’s Dilemma fundamentally alters the prediction of outcomes. When the actors have to take into account how present action influences available future strategic alternatives, a more complex estimation of benefits and costs has to be made. An attempt to cheat on the other actor in the first game might, for example, make it harder to achieve mutually beneficial outcomes in future games. Actor reputation thus becomes important. Depending on the type of opponent one is facing, different strategies might be optimal. Against an opponent that, for example, uses a so-called tit-for-tat strategy, the best counterstrategy is often to also use a tit-for-tat strategy (Axelrod 1984). Against another actor, that for example always chooses defection, the optimal strategy is also to always defect, while unconditionally cooperative strategies most successfully are met by the same strategy. In other words, the prime issue facing the actors is

---

7 The type-concept in game theory defines what precise strategy a particular actor plans to use in an iterated game sequence. An actor’s strategy is a complete description of that player's choices in all stages of the game, and towards all possible counter-strategies. Various potential strategies in this way are condensed into a limited number of types. For example, a tit-for-tat player always begins with playing cooperatively, and then responds with the same kind of move as the opponent chose in the previous game. A so-called saint always plays cooperatively, irrespective of how the opponent plays. Other strategies might be to always defect, or to always counter defect with a given number of defections in subsequent games. Even though the possible number of strategies in an infinitely iterated game is infinite, this procedure makes it possible to deduce logical consequences from a number of simplified strategies.
to try to find out what type the opponent is. Only then can an optimal strategy be implemented. In real-life situations, the actors presumably use all available information sources to make as adequate predictions as possible of the opponent’s type. Previous interaction with this actor might be a valuable source of information, but also different kinds of statements about him, estimates of his evaluation of future outcomes (discount factor), and observations of his prior interactions with other actors can be useful.

It has been shown in the so-called *Folk theorem* that in an indefinitely or unknown number of iterated Prisoner’s Dilemmas, cooperation (mutual free trade strategies in the example above) to various degrees can be strategic equilibria, provided that the discount factor is not too large (Kreps 1990). Patterns of cooperation can thus be sustained, and it is these patterns that we call institutions. This does not mean, however, that cooperative outcomes are guaranteed. The absence of cooperation, as well as combinations of cooperative and non-cooperative strategies can all be part of equilibria, depending on the type of opponent. In terms of the example on trade policies discussed above, this means that observations of free trade as well as trade barriers are congruent with the model this far.

Standard game theory does not offer any obvious solutions to the problem of multiple equilibria in an iterated sequence of Prisoner’s Dilemma. If no additional information about the game or the actors is available, the eventual

---

8 It has been empirically observed in a significant number of experiments that individuals also often tend to choose cooperative strategies in a finite and known number of iterated Prisoner’s Dilemmas, despite the fact that mutual defection is the only strategic equilibrium in such a game sequence. This observation can be explained, however, if it is assumed at the initial stage of the game that at least one of the actors assigns a slight probability to the possibility that the opponent might not be fully rational (he might, for example, be a tit-for-tat type). If this assumption is made, it can be shown that cooperation is rational until near the end of the game sequence (Kreps, Milgrom, Roberts & Wilson 1982).

9 The discount factor measures the importance the actors put in future outcomes, the so-called shadow of the future.

10 It should be noted that an additional restriction has to be placed on the Prisoner’s Dilemma game when it is iterated in order to make mutual cooperation an equilibrium. In reference to Figure 1, the inequality $2b > (a+d)$ must hold. Otherwise, both actors would gain from alternating between cheating each other, compared with mutual cooperation.
outcome of the interaction thus cannot be predicted. It should be noted that this inconclusiveness has bearing on the discussion whether institutions should be assumed to be optimal solutions or not. As far as the institutions are perceived as equilibria in an iterated series of Prisoner’s Dilemmas, this assumption cannot be defended à priori. The reason is that many of the possible equilibria are not Pareto-optimal. Mutual defection is, for example, one possible Pareto-inferior equilibrium under standard assumptions in such a supergame (Brams 1985; Mertens 1989).

**Inputs from prospect theory**

The issue how to systematically eliminate unreasonable strategic equilibria has been a major research topic in game theory for several decades. Most of this research has been undertaken within the traditional rational choice framework and has resulted in solution concepts such as subgame perfect, sequential, and Bayes equilibrium (Rasmusen 1994). Sometimes, these solution concepts produce different outcomes. This does not mean that some of them are wrong, and one is right. Rather, they reflect different aspects of a purely theoretical and abstract model that is used to illuminate various empirical observations (Aumann 1989). A single observation can be interpreted in an infinite number of ways.

The solution concepts mentioned here are but a few. Although these solution concepts have produced significant theoretical progress, another path, however, will be followed in this paper; a path focused less on stringent solution concepts, and more on assumptions regarding rational behaviour and on the context of the game situation. The context of the game situation is particularly emphasized, since this might influence the strategy chosen by the actors. From a non-myopic perspective, the actors have to take all relevant factors into consideration. Although only a limited number and highly stylised set of such factors can be integrated in the theoretical framework, it will be argued that their incorporation might increase the empirical validity of the study substantively. These contextual factors might furthermore raise
research issues not usually analysed in traditional rational choice studies.

Proponents of the so-called prospect theory have questioned the standard assumptions on the rationality of the individual in rational choice theory (Kahneman 1979). The assumption that the individual maximizes expected utility is challenged. Different kinds of cognitive limitations make the assumption of strict rationality dubious, according to this school. More importantly here, however, is the significance prospect theory puts on the context of the game.\textsuperscript{11} Mainstream game theory is not concerned with contextual issues. Since all relevant parameters should be made internal to the game model, there is no need to dwell for example on the context of the decision situation:

\begin{quote}
The model should contain all relevant aspects of the situation; in particular, any possibility of (pre)commitment should be explicitly included (Van Damme 1989:139).
\end{quote}

In fact, to do so would make the model less stringent. Nevertheless, there have been significant attempts to bridge the gap between traditional rational choice theory and prospect theory (Geva 1997).

Two of the most significant issues in the debate between rational choice and prospect theorists are concerned with preference formation and intensity on the one hand, and context dependence on the other. Both aspects will be discussed below, and the focus will be put on the generation of tangible research issues and testable hypotheses rather than on theory development. Since the concepts of preference intensity and context dependence primarily have been analysed theoretically and validated data from laboratory settings, this procedure seems sensible. The concepts have been properly defined and theoretical implications have to some extent been drawn. What is to a considerable extent still lacking, however, is validation against empirical data from natural settings, and to achieve this, theoretical implications with clear

\textsuperscript{11} The term framing is sometimes used when discussing the context of a decision situation. The framing concept, however, have been defined in several different ways, depending on the scientific approach. Therefore, the terms context and reference dependence are used exclusively in this paper, in order to enhance clarity.
empirical relevance have to be formulated. The issue in what ways preference intensity matters will be discussed first.

Returning to the Prisoner’s Dilemma game on trade policy discussed above, it is clear that the preferences of the actors are only given on an ordinal scale. Nothing whatsoever is known regarding the cardinal differences between the outcomes. Although the actors are assumed to be choosing strategies according to Von-Neumann-Morgenstern expected utilities, this does not change the fact that the outcome in a single Prisoner’s Dilemma is insensitive to actors’ intensity of preferences (Kreps 1990). Only the ordinal ranking of the outcomes matters.

In an iterated game sequence of Prisoner’s Dilemma games, however, it seems almost obvious that preference intensity matters. If the compounded gain from repeated cooperation is very large, or the loss from defection is considerable, the incentives for choosing cooperative strategies increase, ceteris paribus. Furthermore, preference intensity could differ significantly between the actors. One of the actors might, for example, gain very much from cooperation, while the other’s potential gain is almost negligible. Similarly, one of the actors might lose heavily from non-cooperative strategies, while the other’s risk from failure to cooperate is very small. Assuming that preference intensity is directly correlated with the magnitude of potential gains and losses, it might be reasonably suggested that it will matter to the actors whether there are large discrepancies between them in preference intensity over different outcomes.

When an actor chooses a cooperative strategy (that is, free trade) in such a game as the one described above, he knows that two things could happen. If his counterpart chooses not to cooperate (trade barriers) he will get his worst outcome. If he chooses to cooperate, he will get his second-best outcome. If he instead chooses not to cooperate (trade barriers policy), he could either get his best outcome or his second worst outcome. Now, it might seem to be a reasonable assumption that the more the actor has to gain from a cooperative outcome as compared with a non-cooperative (trade barriers), the more likely he will be to gravitate towards choosing a cooperative strategy. This,
However, this is not necessarily the case. Consider a situation where one of the actors would gain considerably from mutual cooperation, while the other one would only benefit marginally. This case is depicted in the Prisoner’s Dilemma game with unequal gains from cooperation between actor A and actor B in Figure 2 below.

*Figure 2. A Prisoner’s Dilemma game with unequal gains from mutual cooperation*

<table>
<thead>
<tr>
<th></th>
<th>Cooperate</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>100, 51</td>
<td>49, 101</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>49, 101</td>
<td>50, 50</td>
</tr>
</tbody>
</table>

Assuming utility-maximizing behaviour, the actor with the highest degree of leverage would in this situation be the actor with the least to gain from cooperation (B), since this actor could threaten A not to cooperate with higher credibility. Actor B would furthermore and under given circumstances, potentially be able to request considerable side-payments, provided that transferable utility is assumed. In an iterated game sequence, this would mean that the initial game is gradually transformed into a Prisoner’s Dilemma where the gains from cooperation are more equally distributed. The transaction costs in such a transformation, however, could be significant. Aggregated gains from cooperation thus decrease.

It should be noted that the purpose here is not to suggest formal statements
on credible threats, but rather to suggest hypotheses on marginal outcomes, statements that make suggestions about the expected outcome directions, rather than point predictions (Shapiro 1994).\footnote{An accessible discussion on credible threats and similar issues concerning various refinements of strategic equilibria can be found in (Morrow 1994).} When one of the actors has little to gain from cooperative outcomes, he would presumably be more inclined to take risks, and thereby would potentially stall cooperative processes. The actors might obviously have such incentives to “take a chance”, since the potential gains from cooperation in most cases can be distributed between the actors in a large number of ways. One or both actors might then perceive risky negotiation strategies over spoils as optimal (Schelling 1960).

The suggestion here is that when the potential gains from cooperation in an iterated Prisoner’s Dilemma are fairly equally distributed between the actors, the level of cooperative outcomes will correlate positively with the size of these gains. A highly unequal distribution of gains will, although both actors potentially could benefit from cooperation, tend to decrease cooperative behaviour. Although allowing side-payments could potentially increase the expected level of cooperation by levelling out initial inequalities in the distribution of gains, transaction costs would nevertheless, as a rule, cause enough friction to reduce the expected level of cooperation. The size of the transaction costs could furthermore be assumed to correlate positively with the size of the initial difference in gains, which means that the correlation between equality in gains and the level of cooperation is not negated with the assumption of transferable utility. The absolute level of cooperation could be affected by allowing side-payments, but since we are only interested in marginal predictions – whether there is in fact a positive correlation between equality in gains and level of cooperation – and not their magnitude, this issue is outside the scope of this paper.

This prediction furthermore seems to be strengthened by the argument in prospective theory that risk aversion tends to favour cooperation in an iterated Prisoner’s Dilemma (Raub 1997). The more risk averse the actors are
and the higher they perceive the risks involved, the more they will tend to choose cooperative strategies.\textsuperscript{13}

The actors in a particular decision situation are, according to traditional rational choice theory, supposed to be pure utility-maximizers. They are not supposed to be influenced by factors such as reference dependence (actors evaluate outcomes from a particular reference point rather from a strict expected utility perspective), loss aversion (gains and losses are treated differently although changes in expected utility are identical), or instant endowment effects (actors adapt to gains more swiftly than to losses), as suggested by prospect theory (Levy 1997). An understanding among scholars still awaits, whether it is possible to throw away such suggestions in traditional rational choice theory, without of a too much loss in theoretical parsimony. It seems, however, that each of these propositions has gotten considerable support in quite a large number of laboratory settings, which make it difficult to discard them without further consideration, although few of these results have been corroborated in natural settings (Levy 1997).\textsuperscript{14} These propositions seem to have a substantial intuitive appeal, which makes it worthwhile to explore them further. This is especially important to do in close reference to empirical observations, since the main interest hitherto has been regarding purely theoretical issues and laboratory experiments. The focus will here be put on the concept of reference dependence.

The phenomenon of reference dependence has mainly been observed in laboratory settings and refers to the finding that people often base their deci-

\textsuperscript{13} This conclusion might, however, be jeopardized if it is very easy for the actors to switch partners to cooperate with. Risk aversive behavior might lock actors in sub-optimal relations, since they are not willing to change partners when this could have been to their gain (Flache 2001). Regarding trade relations, however, switching trade partners is a rather complex issue, which means that considerable (sunk) costs have been invested in the trade relation, making switches costly. Therefore, risk aversion tends to foster cooperation in this, and other similar comparably rigid, contexts.

\textsuperscript{14} This is quite natural, since most work on prospect theory has been undertaken in cognitive psychology (and to some extent in economics), where laboratory experiments are part of the mainstream discipline to a much higher degree than in contemporary political science. This also means that the number of applications in political science has, so far, been limited.
sion on some prior notion of “reasonable outcomes”, e.g. initial endowment, rather than on a strict calculation of expected utility (Kahneman 1991). The reference dependence axiom furthermore violates traditional assumptions on rational behaviour, since people often seem to overestimate losses compared to gains. In simple words, people tend to avoid potential losses even though expected utility is positive, and to gamble on gains even though expected net benefits are negative (Kahneman 1979). Identical outcomes in terms of expected utility might thus induce different strategies depending on the subjective notion of risk involved. This kind of behaviour is difficult to explain within the traditional rational choice framework and is one example of the contemporary debate between cognitive and rational choice theorists (Levy 1997).

The importance of context

Reference dependence seems to be an important aspect influencing bilateral cooperation. The issue-area where the cooperation takes place, however, would not affect predicted behaviour from a pure rational choice perspective, even though “…it is clear that the context and the presentation of the game matter” (Kreps 1989:171). It would not matter whether the interaction concerned mutual benefits from, say, environmental investments, trade, or increased security, since the actors would maximize their utility in a self-interested manner in all cases. From an intuitive perspective, this does not seem fully congruent with empirical observations. Depending on the particular issue-area, diverging sets of commonly accepted norms for cooperation possibly have emerged. This does not mean that the general assumption of self-interested and rational behaviour as the most general description of actors’ incentives has been abandoned. Rather, the point is that the kind of self-interested and rational behaviour is influenced by the particular context where the interaction takes place. The meaning and content of rational self-interested behaviour is, to some extent, influenced by contextual factors, simply because these factors affect restrictions on behaviour in terms of costs
and benefits. In a discussion on the Nash solution concept, a centrepiece in mainstream game theory, the distinguished game theorist John C. Harsanyi (1989:66) states that,

…predictions are based on the assumption that both sides’ expectations about each other’s behaviour are endogenous, that is, are based on the intrinsic parameters of the bargaining situation (game situation) as such. But in many cases this may not be true. For example, the two parties may live in a society where it is customary for an older person to receive twice the payoff his younger opponent will receive [irrespective of what would be expected from the Nash solution]. If each player expects the other player to follow this rule, it will become rational for him to follow it too.” (Emphasis in the original)

It could be argued that these traditionally exogenous factors could very well be made endogenous to the rational choice model, if preference formations were allowed to be influenced by such factors. Thus far, however, this has not been systematically attempted.

There is almost always a trade-off between theoretical parsimony and empirical realism. By simplifying a decision situation into a tractable model, information is inevitably lost. Where to draw the line between parsimony and realism is not possible to determine in a positivistic way, but is rather a matter of judgment (Aumann 1989). The argument here is that as long as the attempts to integrate contextual factors of the type discussed in this paper have not been proven successful, a productive research strategy might be to analyze how contextual factors influence the actors’ perception of optimal strategy choices in a more tentative way.

Consider, for example, the difference between environmental bilateral cooperation and trade policies. Different kinds of environmental disturbances often have considerable transboundary effects. Inadequately managed farmlands in southern Sweden, for example, contribute considerably to the increasing eutrophication of the Baltic Sea (HELCOM 1997). The effect is disturbed ecosystems, not only in Swedish waters, but in Polish and other countries’ as well. Now, in this issue-area, a mutual understanding emerged during the late 1980’s and early 1990’s, where it was perceived as reasonable by all involved parties that the more affluent Western countries should provide
the former Soviet Union countries with considerable sums of assistance (Stockholm Environment Institute 1994; Berg 1995; Hassler 2000). Although most of these support programs had significant facets of self-interested considerations by the donor countries, these programs where nevertheless initiated in a framework where assistance was perceived as appropriate. The governments of the Western donor countries apparently interpreted the contemporaneous situation as a state of affairs where traditional foreign policies emphasizing state interests were not fully legitimate (Hassler 2000). This has generally not been the case in regard to, for example, trade issues. Although considerable assistance targeted at economic development has been provided, preferential trade agreements have generally not been part of the discourse (Berg 1995). Instead, the free trade norm has been the obvious focal point, and has not been open to discussion to any significant extent. The “support” in this area has thus been various kinds of trade liberalization agreements, which hardly can be seen as part of an assistance discourse.

A typology of some different kinds of bilateral interaction contexts will be outlined in the remaining part of this paper. This typology is intended as an instrument to facilitate comparisons of strategic behaviour between different issue-areas. The most important features of these contexts are their differing implications regarding the distribution of gains (Knight 1992).

**Definition of four general contexts**

Various contexts imply the emergence of different institutional solutions. Therefore, depending on the nature of the situation and the preferences of the actors, each actor will try to mold a particular issue of cooperation into the context most beneficial to him, according to subjectively predicted outcomes. In other words, these contexts are defined in a similar way as institutions were above. They are the result of the interaction between self-interested actors, and are defined as equilibria of an iterated game sequence. This means that they cannot be a priori assumed to be optimal from the perspective of any particular actor. Only in rare cases where one actor has been
extraordinary influential over the outcome, can the observation of a particular outcome be explained by a single actor’s preferences. On the other hand, when a specific empirical context is identified, this observation can serve as a starting point for the further analysis of actors’ interests. From the tentative assumption that the actor gaining most from this observed context is the actor that has been able to influence the interaction process to the largest extent, further tracing of actors’ preferences and power can be performed.

This interpretation of reference dependence is somewhat different from the ordinary use of the concept in prospect theory. In most cases, reference dependence is seen as a component of the human psyche, and is thus not open for strategic action or conscious manipulation. The actors are typically not assumed to take this knowledge into consideration when they formulate their strategies. Whether or not the individual’s preferences are influenced by reference dependence, has been one of the primary research questions in prospect theory. These research efforts have been mainly directed towards questioning traditional rational choice rationality assumptions. Although these studies can be of considerable value, it is rather surprising that so few studies have been undertaken where reference dependence is taken as a point of departure. In other words, the notion of reference dependence is assumed to be valid in similar ways that traditional rationality assumptions are assumed to be so, and the analysis will then show the value of this postulation. This approach opens up possibilities for some quite interesting research issues. When reference dependence has been assumed to be part of individual rationality, the actors must be assumed to take notice of this. They will thus adapt to a situation where all actors are influenced by reference dependence and use this knowledge to further their individual interests. This makes it rational for the actors to try to portray a decision situation in such a way that reference dependence serve their interests. Thereby, the context of a particular decision situation becomes contested ground among the actors. In order to evaluate this approach, the next step is to define some general characteristics of these contested grounds.
At least four general contexts with relevance for bilateral cooperation processes can be identified. We call these liberal, realistic, responsibility, and solidarity contexts. Each of these is defined below, and examples are provided in order to make their specific features more easily discernible. It should be noted, that these contexts are empirical, not analytical or ideological. These contexts are the outcomes of the interaction between the relevant actors.

The liberal context is based on the characteristic that all actors should be obliged to follow the same rules. In analogy with the classical liberal society where all individuals are supposed to be governed by identical rules, this context presupposes that all states should have identical obligations to follow a common set of rules. This is probably the most common context. In the issue-area of international trade, for example, negotiations taking place in a liberal context would have free trade as a natural focal point, or at least an identical set of rules covering different kinds of trade barriers. Such a liberal norm would probably be more attractive to some countries than to others, even though the Ricardian theory of comparative advantage states that all countries would benefit from free trade. Countries with a high degree of dependence on foreign trade, for example, would most likely be differently affected by a liberal trade regime than countries that are more self-sufficient. Similarly, small countries would most likely be more inclined to support such a regime than larger countries, since the latter have a larger potential to make credible commitments regarding the punishment of unwanted behaviour. Hence, the large state is less dependent on a uniform set of rules, since it has a higher potential to impose restrictions on others’ behaviour also in ad-hoc type of interactions. Furthermore, a more economically developed country could possible attain a more beneficial outcome in a liberal context, compared with developing countries. This so-called infant industries argument is, however, only accepted as valid by some scholars (Todaro 1989).

In a realistic context, the norm pattern has a different flavour. Here, the equality before the law is not the main feature. Rather, the guiding principle is that those harmed bear the costs themselves. There is really no moral code
here, which makes this context akin to how the anarchic international system in general is often perceived (Goldmann 1978). An example from the environmental issue-area will make the argument clear. PCB’s and DDT’s have been found to have negative effects on both humans and on animals. These substances have been found to become very widely dispersed over time, by sea currents and transportation through various biological food chains. This means that different countries will be differently affected by continued emissions of PCB and DDT. Not only are these substances unevenly distributed due to a multitude of factors regarding, for example, geographical and meteorological conditions, temperature differences, and type of biota, but also different countries are unequally affected due to the country’s particular socio-economic conditions. One country might rely significantly more on the resources threatened by the PCB’s and DDT’s than some other country.

Now, in the realistic context, the context of the situation is characterized by the perception that no one will cover someone else’s losses. An actor that has experienced considerable damages from PCB’s and DDT’s might try to influence the behaviour of those countries responsible for the emissions, but they will not likely be very successful. Since there are in most cases no enforceable regulations in this context, emitting countries will be reluctant to cover costs for alternative production methods and similar measures, as long as they are not themselves negatively affected. As often is the case, however, mutually beneficial solutions are available in this context as well. If all actors perceive the realistic norm as legitimate, such a solution would most likely entail some kind of side-payments (i.e. transferable utility is assumed) (Bernauer 1996). As long as the negatively affected country benefits from, for example, subsidizing investments in the other country to decrease those negative effects, the transfer of resources would be mutually beneficial. These kinds of transfers of resources have been characterized as an example of the so-called VPP (Victim Pays Principle), which is a direct corollary to the PPP (Polluter Pays Principle) (Mickwitz 1998). The strong incentives of the victimized actor to appeal to other normative contexts, however, should be noted.
The third identified context is that of responsibility. This kind of reference dependence is, in a way, a mirror image of the realistic context. The main feature here is the moral obligation to take responsibility for all costs that one is the cause of. Using the same example as above on DDT and PCB, this would imply that a country emitting such substances would have to pay compensation to all affected countries for all negative effects. This is the international variation of the familiar Polluter Pays Principle (PPP). The difference is that while this principle in most cases can be enforced in the domestic context, this is generally not the case in bilateral relations. It should finally be noted, that there is no analytical difference between realistic and responsible contexts - between VPP and PPP - in terms of efficient outcomes (Coase 1960). If side-payments are allowed and transaction costs are assumed to be zero, the outcome in terms of the reduction of emissions would be identical under both principles. The only difference would be who covers the costs. It would be either the country suffering from the emissions (realistic context) or the emitting country (responsible context).

The last of the four identified general cases of reference dependence is the one we have called solidarity context. This kind of reference dependence is characterized by a perception among key sub-national actors that a moral responsibility to provide support might be warranted. A representative case in this category would be a bilateral interaction between two countries with very unequal capability. This capability could be of very diverging nature, depending on the issue-area analysed. One of the countries in the bilateral relation might, for example, have a very limited military capability. Suppose the other country has a significant capability, and that the relation between these countries is comparably congenial. It could then be argued, in both the strong and the weak state, that the more powerful country has a moral imperative to provide the less fortunate state with some kind of security assurances. A tentative proposition could be, that the larger the difference in potential/actual capability, the larger the imperative for support, all other factors given.
Implications for practical research efforts

The issue whether preference intensity in an iterated Prisoner’s Dilemma matters has been elaborated in this paper. It has been suggested that in situations where the potential gain from mutual cooperation is great and approximately equally distributed between the actors, cooperative outcomes will be more likely.

Four general types of contexts in bilateral relations have furthermore been briefly described: liberal, realistic, responsibility, and solidarity contexts. These contexts, or institutions, are all defined as strategic solutions in a sequence of iterated Prisoner’s Dilemmas. This means that they are not to be seen à priori as optimal from a societal perspective, but rather as the result of the actors’ self-interested individual strategies. Depending on to what extent the actors have been able to promote their particular interests, different institutions will evolve. It should be noted, however, that the iterated game sequence is not assumed à priori to be a traditional supergame, that is, a game sequence where identical games are repeated one after another. Rather, the assumption is that although the game situation starts with a traditional Prisoner’s Dilemma, it might, at later stages, evolve into other kinds of games. Factors such as external shocks, changed information levels, reputation building and the like, might cause the original Prisoner’s Dilemma to transform into, for example, a coordination game. It is furthermore quite possible that the evaluation of different outcomes might change over time, due to similar factors, without changing actors’ preference orderings. This means that changes of intensity of preferences have taken place, although the fundamental incentives remain unaltered in this particular game. Incentives for cooperation in future games might, however, have been affected due to these intensity changes. Irrespective of whether the game has transformed into another kind of game or if only intensity changes have occurred, the actors’ original preferences, their “tastes”, have not, however, changed. What drives the changes here is not endogenous taste transformation, but exogenous changes in information level, structural alterations caused by external shocks, or the like.
It is important to notice that the institutions, or contexts, discussed here are seen as endogenous variables. They are not part of a given structure where the actors interact, but open to strategic manipulation. The actors are assumed to continuously try to mold present contexts into ones more favourable to themselves. At the same time, however, all decision-making is interdependent. At a given time, optimal strategies are determined by actor preferences and ability to influence the strategic situation, estimates of the opponent’s preferences and ability, but also by how the present context is perceived by the actors at that time.

The last factor, concerning the present context, should formally be interpreted as the history of the game up to that point. The actors are assumed to have complete information about all previous moves by all actors. In this research approach, however, this formal interpretation serves as a point of departure, rather than as a complete guide for analysis. This means that the present context should be understood as assumptions of how the actors subjectively perceive the situation. In other words, this provides the analysis with a considerable degree of dynamism. The preferences of the actors are not assumed to change, but as the context continuously transforms in accordance with interaction outcomes and potential external disturbances, optimal strategies might vary over time.

Following the research strategy suggested here, the first step is to estimate the objective potential gains from bilateral interaction in the issue-area being studied. This estimation is objective in the sense that it is inferred from the assumptions concerning actor preferences over outcomes, rather than from their strategic incentives. It only concerns whether a joint gain from interaction is possible or not. Issues of distribution of these potential gains are, for the moment, disregarded.

In terms of trade issues, for example, it can reasonably be assumed, based on the utility functions of the actors, that there is a potential gain from increased trade, or from reduced trade barriers. The basic premise here is the Ricardian notion of comparative advantage. Another example is potential gains from cooperation on environmental issues. The objective estimation of
these gains would be based on actor preferences regarding this issue; to what extent these environmental benefits are collective in their nature, and the cost of measures undertaken.

The second step would be to evaluate the distribution of net benefits on a similar abstract level. The critical issue here is how the distribution of the gains from cooperation varies with the different contexts defined above. In what way, and to what extent, does the institutional solution influence the distribution of gains? What implications would a liberal context, for example, have, as compared with a realistic on the distribution of gains related to this particular issue? If the issue being studied is, for example, environmental cooperation, how would the benefit accrued to each actor vary with type of the context. How would the costs be distributed? If the issue/area is environmental cooperation, say, for example, protection of the Baltic Sea, to what extent does the each actor gain from joint action? How would these gains be related to potential solutions within different contexts?

The third, and final, step would then be to compare empirical data with theoretical predictions. Empirical data would most likely primarily consist of information about action taken by the actors, but additional clues on actors’ motivation should be sought also from other sources, such as, for example, official statements, public reports, parliamentary bills and the like. These data should then be compared with what would had been the best interest for each actor according to what has been stated during set one and two as described above. Here, preference intensity among the actors becomes important, as one factor influences not only strategic choices within a particular context, but also attempts to change the existing context into one more favourable to this actor.

In order to include the dynamic elements mentioned above, data cannot be analysed statically. In addition, the process of changing incentives has to be traced over time; otherwise it will not be possible to understand why actors sometimes choose strategies that seem sub optimal, when the context is seen as given. Through attempts to change contextual factors, this actor might increase net benefits in the future. In this way, what is perceived as structural
factors from a static perspective becomes endogenized from a dynamic point of view. In order to be able to more fully grasp these sometimes very complicated, dynamic aspects, it is necessary to also take the history of the game situation into account. This does not imply, however, that the actors are not assumed to be purely forward-looking utility maximizers.\textsuperscript{15} Taking history into account here means that it could be necessary to consider prior action in order to understand present incentives. The actor might, so to speak, have a “grand plan” covering several iterations of the game, and this might explain observations of sub-optimal choices in some particular single game.

By following these three steps, an understanding of the contextual influence on rational behaviour can be reached. The traditional assumptions regarding utility-maximizing behaviour are kept intact, but by making certain structural factors endogenous, the importance of context can be integrated in the research effort. As has been stated above, the research strategy suggested here sacrifices theoretical rigor to some extent in order to achieve empirical relevance. Considering the huge research efforts that have been invested in rational choice theory development and prospect theory laboratory experiments as compared with empirical testing in natural settings, this sacrifice seems justifiable.

\textsuperscript{15} It has been suggested that the traditional rational choice assumption regarding forward-looking maximization of expected utility could be replaced by an assumption that the actors instead continuously adapt to prior experiences (Kanazawa 1998). In our opinion, however, it does not seem reasonable to assume that the actors only learn from experience and are not able at all to make reasonably accurate predictions of future outcomes when they have no prior experience of such particular situations. Since an integrated approach where the actors are assumed to be both backward- and forward-looking has yet to be formalized, the approach used in this article is the traditional one, where actors only try to estimate future outcomes.
References


Berg, C. 1995. The environmental support to the Baltic States, Poland and western Russia. Stockholm, Swedish Environmental Protection Board.


North, D. C. 1998. *Five propositions about institutional change*. In 'Explaining social institutions'.


