Public Choice **106:** 365–388, 2001.

Between appeasement and belligerent moralism: The evolution of moral conduct in international politics *

ULRICH WITT
Max-Planck-Institute for Research into Economic Systems, Evolutionary Economics Unit, D-07745 Jena, Germany; e-mail: witt@mpiew-jena.mpg.de

Accepted 17 May 1999

Abstract. What are the conditions for "moral" conduct in international politics to be viable? This question is explored by modeling regional conflicts as a two-stage game in which imitation of other countries' strategies and public opinion formation processes are assumed to have an impact on a country's strategic choice. The results derived by using the notion of an evolutionary stable strategy point to some very special conditions for moral conduct to emerge and survive, a fact that may explain the historical finding that there have been rather rare periods during which peaceableness did indeed prevail in international politics, at least at a regional level.

1. Introduction

Since the growth to world powers of stable democracies in North-America and Europe the question of "moral", or fair, conduct in international politics has become a major issue. The reason is that in these countries public opinion is politically influential and, under the influence of Western political philosophy, public opinion has increasingly rejected Machiavellian political practices as "immoral". What is not entirely clear is whether moral, or fair, conduct can indeed hold out if, in conflicts with other countries, the opponents carry out Machiavellian policies. What are the conditions that must be satisfied if "morality" – adequately defined – is to succeed in international politics? Concern with these questions is clearly expressed in the recent debate on the revision of the political agenda of the North Atlantic Treaty Organization at its fiftieth anniversary. Since the end of the Cold War, the challenge to moral, or fair, conduct in international politics has mainly occurred in relation to regional conflicts in which local aggressors threaten one or a few (more or less) peaceful neighbors for ethnic, religious, or territorial reasons. The questions therefore arises of what kind of political conduct is conducive to defending standards of morality in such regional conflicts so

* The author is grateful to J. Irving-Lessmann, G. Modelski, R. Joosten, M. Reksulak, and especially an anonymous referee of this journal for helpful comments.

that Machiavellian attitudes are prevented from spreading out in the region (Gaddis, 1992/93). This question indeed seems to have informed-political reactions in such conflicts as those caused by the invasion of the Falkland Islands by Argentina in 1982, the invasion of Kuwait by Iraq in 1991, or the now decade long struggle of several countries against Greater-Serbian nationalistic aggression in the Balkans.

In earlier work on international politics (see, e.g., Schelling, 1960; Bernholz, 1985; Brams, 1985; Niou, Ordeshook, and Rose, 1989; Wolfson, Farell, Gill, and Shabahang, 1992) a game-theoretic framework was introduced to capture the logical structure of international conflicts. In such an approach, a simple, frequently chosen, way of modeling conflicts in international politics is to assume a bilateral confrontation in the form of a one-shot 2×2 game. For a discussion of what "moral" conduct in international politics could mean and, in particular, how moral conduct can evolve, it seems necessary, however, to extend the game-theoretic framework to account for some facts which have been widely neglected in the analysis:

- conflicts in international politics usually develop over time in a way not satisfactorily represented as a one-shot encounter; aggression may turn up in a surprising, initially disguised, fashion, and elicits domestic reactions as well as international ones;
- in democratic countries, public opinion formation plays a key role in determining both initial conduct and the policy chosen in response to an international conflict; and public opinion formation is likely to differ from rational political decision making;
- the geographical patterns of regional conflicts require the analysis to account for neighborhood effects over time, i.e., potential "contagion" and dissemination phenomena.

Accordingly, an attempt will be made in this paper to suitably extend the game-theoretic analysis to address these facts, at least at an elementary level. The paper proceeds as follows.

Section 2 briefly summarizes the basic game-theoretic approach to international politics and connects it to the debate on what "moral" conduct can mean in the abstract terms of strategies in a game. Section 3 then turns to the first of the extensions just mentioned and argues that the political reactions, which are elicited by an aggression in a regional conflict and in which public opinion formation plays a role, cannot be represented by a one-shot game. To avoid the complexities of a dynamic analysis, the minimal concession that must be made to accommodate those reactions is to allow for two stages in the game: one in which the conflict builds up through the simultaneous

actions of the international players, and one in which it is settled through their respective reactions. With respect to the second extension, the impact of morality considerations in public opinion formation processes on decision making in international politics, a behavioral assumption is introduced which seems to be supported by the historical record. In contrast to rational choice approaches (e.g., in Bueno de Mesquita and Lalman, 1992) it suggests that best-reply type policies may not always be what democratic governments tend to choose in an international conflict.

Section 4 deals with the third extension. In a strongly simplified version, the geographical patterns of conflicts and the neighborhood effects they imply are integrated into the game-theoretic setting. Section 5 argues that the strategy pursued in international politics by a country may change as a result not only of own experience, but also of observing the outcome obtained by other countries with their strategies in international conflicts, i.e., as a result of learning and imitation. This suggests an evolutionary approach in which, moreover, the consequences of a locally emerging strategy innovation and the conditions for its dissemination through imitation can be explored. The implications of the model are derived in Section 6 using the logic which underlies the notion of an evolutionary stable strategy. The results allow some light to be shed on some major contingencies which appear to play a role for moral conduct and peaceableness to survive and disseminate in international politics, contingencies which indeed seem to be present in the historical record. Section 7 offers some conclusions.

2. Conflicts in international politics: The game-theoretic perspective

International conflicts occurring at a certain time and place are historical individualities and, accordingly, the particular policies pursued by the countries involved are unique combinations of contingent measures and moves. However, in an explanatory approach like that of game theory, the complexity of the historical case must be reduced to a level of abstraction at which some generic features of international conflicts can be grasped (for a discussion of the methodological problems see Snidal, 1985). In the present context this means looking for features which historical policies have in common in some respect and, at the same time, abstracting from the multitudinous differences in other respects. In this way, a policy type, i.e., a strategy in a prototypic game, can be identified. Using this classification allows the indefinite number of unique, past and future, historical policies to be broken down to a limited, and invariant, set of distinguishable variants. Once this step has been taken, the situation of two countries involved in a bilateral conflict can be described by a game-theoretic model.

Indeed, the great analytical achievement of the seminal game-theoretic literature on international politics referred to in the introduction has been to demonstrate that there is a generic logic underlying international conflicts which can often be expressed by a simple, symmetric[1] one-shot game of the following kind. Let there be just two strategies c and d:

c: try a solution by negotiating a compromise without resorting to military threat (cooperation) or
d: mobilize and use military force to extort a solution to the own advantage from the opponent by aggression (defection).

Given these two strategic options, it is not difficult to identify the well-known prisoners' dilemma in this game.[2] If both countries choose the cooperative strategy, c, both can realize a pay-off R (reward) which is higher than the pay-off P (punishment) obtained in the armed conflict which results when both choose the defective strategy d. If, however, only one country is cooperative while the other defects and prepares for a military solution, the country relying on c becomes vulnerable to blackmailing politics or even to open aggression and faces the worst possible outcome S (sucker's pay-off). The other country, having chosen d with the opponent not mobilized, gains a position of strength which it can use to squeeze out the highest possible pay-off T (temptation) from its opponent. Thus

$$T > R > P > S.^{3} \tag{1}$$

As is well known, in the prisoners' dilemma game, defection is the dominant strategy. Rationality dictates that both countries choose this strategy so that conflicts, whenever they occur, would be settled in one, and only one way, namely by military confrontation. Even though each country would actually prefer conflicts to be solved peacefully by the opponents, international relations would always remain in a state of anarchy.

The disturbing implication of the prisoners' dilemma that turns up here has been worrying social philosophers for a long time (see, e.g., Sen, 1987) and has induced moral philosophy to take a stake in the issue from a normative viewpoint (Gauthier, 1985; McClennen, 1989), identifying the peaceful strategy c with the choice that morality would dictate. In fact, this seems to correspond to what, in the public opinion of Western societies, is considered to be the standard of moral conduct applicable to international politics. Moreover, while it is true that war is a frequent concomitant of international conflicts, the recent historical experience in Western Europe, for example, shows that war is far from being the only way of settling conflicts. In the last decades there have now been many cases in which compromises were

achieved in peaceful negotiations, perhaps because of a moral ban imposed on military measures by the countries' public opinion. The question therefore arises how to explain observable cooperative (or moral) conduct in international politics and the conditions under which it can prevail, given the fact that there is always the temptation to defect – the dominant strategy in the just discussed prisoners' dilemma. This question has become particularly significant in the many remaining regional conflicts all over the world, with the end of the Cold War and its exceptional strategic constellation.

3. A simple two-stage game with public opinion formation

An elementary way of extending the model of international conflicts briefly reviewed in the previous section so as to allow opinion formation processes to influence a country's policy is to assume a two-stage game. In the opening stage the strategies c and d of the one-shot game reappear as possible (simultaneous) moves. In a second stage, the opponents involved may or may not react to the outcome of the first stage in a way that terminates the game. Thus, the (again simultaneous) moves in that stage of the game are supposed to reflect the simultaneous responses to information not available when the decisions in the opening stage had to be taken. For example, in countries which, having chosen move c, have fallen victim to blackmailing or even open aggression by an opponent, public opinion often expresses a strong desire for punishing the aggressor, i.e., for retaliation. In a two-stage game, by assumption, the conflict is over after the pay-offs for the second stage have been obtained. As a consequence, in modeling the second stage, specific closing moves have to be assumed as being available for terminating the particular conflict.[4]

To keep things as simple as possible let there be just two closing moves:

a: accepting the first round outcome without further action or
r: seeking revenge.

With the additional moves the strategic situation changes as follows. For each of the players there are now two choices in the opening stage and four *contingent* choices in the closing stage, hence, eight contingent strategies per player and sixteen outcomes or combinations of moves of both players altogether (cf. Figure 1). However, not all of them are relevant or make sense. In view of the public opinion formation process, which is supposed to determine at least a democratic country's reaction in the second round, there is no reason to assume retaliatory action in the second stage after experiencing mutual cooperation in the first stage. The motivation for a

country's public opinion calling for retaliatory action hinges on an event which is perceived in public as necessitating or deserving vengeance. But, even for a country which defects in the first stage, seeking vengeance in the second stage does not seem plausible, if it has been faced with cooperation by its opponent in the first stage. One may therefore reasonably presume:

Assumption 1. If the opponent uses move c in the first stage an international player has no reason to choose move r in the second stage, independent of what her/his own move in the first stage has been.

Assumption 1 reduces the number of contingent strategies that can actually occur to six and the number of outcomes to nine (cf. Figure 1).[5]

Let us turn to the contingent strategies first. Since a symmetric game has been assumed it suffices to inspect those of one player. By assumption, moves are chosen simultaneously. Thus, the opponent's move in the closing stage is not yet known when the own move at that stage has to be made. The latter is therefore contingent only on the opponent's move in the opening stage. (It may also be influenced by what the opponent is expected to do, but this will be ignored for the moment.) Writing down the move in the first stage before the semicolon and the second-stage move, which is contingent on the observed first-stage move of the opponent, after the semicolon, the eight contingent strategies can be depicted and classified as follows. (A summary is given in Table 1.)

- There are two rather trivial strategies with mutual cooperation in the opening stage:
  - {c; a|c} is a "*peaceful coexistence*" strategy (PC): both countries have cooperated in the first stage and, by assumption, accept the outcome;
  - {c; r|c} is excluded by Assumption 1.
- There are four strategies which begin with defection; of these
  - {d; r|c} is also excluded by Assumption 1;
  - {d; a|c} is, thus, the only admissible reply in the closing stage when the aggressor has been faced with peaceable cooperation in the opening stage; the strategy may be labeled "*limited aggression*" (LA);
  - {d; a|d} is a strategy of "*damage confinement*" (DC) after mutual aggression in the first stage;
  - {d; r|d} reflects the destructive attitude of retaliating after mutual aggression; alluding to this the strategy may be labeled "*total warfare*" (TW).

Table 1. Contingent strategies in the two-stage regional conflict game.

Move in first stage of game		Contingent move in second stage	Name of contingent strategy
Country A	Country B	Country A	
Cooperate	Cooperate	Accept	Peaceful coexistence (PC)
Defect	Cooperate	Accept	Limited aggression (LA)
Defect	Defect	Accept	Damage confinement (DC)
Defect	Defect	Retaliate	Total warfare (TW)
Cooperate	Defect	Accept	Appeasement policy (AP)
Cooperate	Defect	Retaliate	Belligerent moralism (BM)

- There are two particularly interesting strategies where a country cooperating in the first stage is driven into conflict by being confronted with a defective, militarily offensive, move by its opponent:
  - {c; a|d} points to a defeatist attitude of a country's public; a strategy which can be labeled "*appeasement policy*" (AP) and which is not uncommon in international conflicts;
  - {c; r|d} a strategy indicating a pugnacious attitude in public opinion formation after experiencing aggression that will be called "*belligerent moralism*" (BM) here.

The pay-off accruing to the players from the various contingent strategies are given by the first move values R, T, P, S minus a value ≥ 0 resulting from the alternative moves in the second stage. While move a does not seem to cause additional costs, the problem with move r is that, as a rule, retaliation incurs heavy costs on both sides, i.e., on those seeking revenge as well as on those being punished. To push the argument to its limits let the costs V_r which a retaliating country incurs be larger than the cost V_p which the punished opponent has to incur. Assume further that revenge causes such heavy costs that in addition to the order relation (1) the following holds:

$$P > T - V_p \text{ and } S > R - V_r, \text{ where } V_r > V_p. \tag{2}$$

The nine outcomes to player (country i) which can result under Assumption 1 have been depicted in Figure 1. The order relations (1) and (2) a rank order of the pay-offs accruing to the nine outcomes in which the three highest and the three lowest ranks are uniquely determined as $T > R > P > \ldots > P - V_r > S - V_r > P - V_f - V_r$. Thus, the highest pay-off resulting from

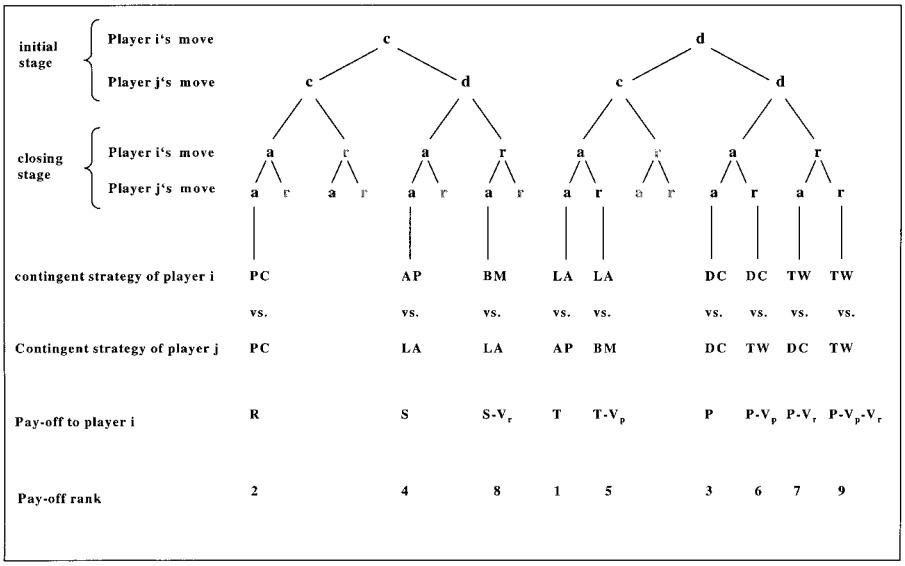

Figure 1.

playing LA against AP, the lowest from playing TW against TW. While two extreme pay-offs result from entering the game with the aggressive move d, the second-best (PC against PC) and the second-worst (BM against LA) can result from choosing the peaceful move c in the initial stage. (For expository purpose, a pay-off ranking is indicated for the case $V_f > T - S$.)

On the basis of rational choice criteria the dilemma diagnosed in the previous section seems to reappear in the two-stage game as given in Figure 1. In terms of the pure, contingent strategies the move sequence {d; a} is a dominant strategy, i.e., playing DC against DC. However, in a country's choice of strategies in international politics, rational choice criteria may be conditioned through the impact of public opinion formation processes. For this to happen, two conditions must be satisfied: first, public opinion must be sufficiently influential; second, the moral value judgements informing it must have implications which suggest, at least partly or temporarily, diverging from what rational choice dictates. The first condition is met in democracies which usually have emerged from socially widely shared value judgements emphasizing equal rights and moral and just conduct. If it is met at all, the second conditions is also most likely to be met in democratic countries because of their very basis in such value judgements. If not overthrown by, e.g., ethnic or religious chauvinism, the morality consideration may be extended to their conduct in international politics.

Indeed, the historical record supports the view that democratic countries are much less aggressive or defective in international conflicts than authoritarian or totalitarian regimes with their strongly manipulated public opinion formation (Bernholz, 1995). The historical record may thus be read to suggest the following behavioral assumption:

Assumption 2. Morality considerations which play a role in the public opinion formation processes of a democratic country tend to prevent their government from choosing the offensive move d in the first stage, if there is a chance of settling international conflicts peacefully.

According to Assumption 2, governments in democracies feel obliged to at least try a peaceful policy in bilateral conflicts. As long as other countries also remain peaceful – which is playing PC against PC – such a policy is rewarded by the second-best outcome R for all players, a kind of "peace-dividend".

Yet, in non-democratic countries the situation is different. The suppression or non-existence of morality considerations in the public opinion formation processes in their countries allows dictators or ruling oligarchies to choose, in a Machiavellian fashion, whatever strategy appears advantageous to them. By hitting a peaceable country with an aggressive initial move d they may hope to draw an advantage if that country follows an appeasement policy (i.e., to play LA against AP with the best possible pay-off T for the aggressor).[7] The true challenge to moral conduct in international politics therefore occurs if the peaceableness of a democratic country is confronted with unforeseen aggression. The public in the democratic country then has to make up its mind as to whether it endorses a permissive (AP) or belligerent attitude (BM). The latter imposes the heavy burden of additional retaliation costs on the country. On the other hand, belligerent moralism may deter opportunism on the part of non-democratic governments.

It may appear that a peaceable country fares best by credibly committing itself to *always* playing BM should it be drawn into conflict by an aggressive opponent.[8] (If credible, such a commitment would induce even a Machiavellian despot to favor peaceful coexistence with the country, provided (s)he acts rationally and has no domestic reason to attack as, e.g., to detract attention from an ongoing struggle for power.) However, a commitment like this is likely to be tested, and it would only pay if it is not too often tested. In fact, the worst case for belligerent morality is to be challenged excessively. The public might then quickly find out what different forms of moral political conduct cost and whether belligerent morality is simply too costly to be sus-

tainable. Needless to say, this peculiar contingency undermines the credibility of belligerent moralism.

How the public opinion will form in such circumstances, and what move in the second stage of the conflict game this is likely to induce, is difficult to decide. Both the permissive and the belligerent strategies are identified by some parts of the public opinion in the modern democracies with moral conduct in international affairs. As mentioned in the previous section, a moral attitude can be attributed to both strategies, because they imply offering co-operation in the first stage where defection is tempting. But, what precisely does "moral" conduct mean with respect to the choices in the second stage, after peaceful cooperation has been disappointed or exploited – appeasement or belligerent moralism? In the military involvements of the West in the Gulf War and the Kosovo crisis, the peace movements in the Western, particularly the European, democracies have shown that society is deeply split about the right answer to this question. Answering it is, of course, a matter of normative judgement. A positive analysis, like that offered here, can only improve the understanding of what the implications of the two alternative policy types are and it, thus, cannot decide the moral question.

4. A probabilistic model of regional conflicts in international politics

To enable the analysis to come to terms with the diverging implications of appeasement vs. belligerence, a broader perspective has to be taken. As a matter of fact, conflicts in international politics are quite frequent. In game-theoretic language, they are played repeatedly by different countries out of a large "population" of international players. In such a perspective, learning and the sustainability (or survival) of certain forms of conduct become meaningful, and a different, evolutionary, philosophy of analyzing games suggests itself – and requires some further extensions of the basic model. One of these relates to the actual, geographic structure of international interactions. If international conflicts occur in sequence at different times and places, and if regional conflicts become increasingly relevant after the end of the Cold War, geographical proximity seems to become a crucial variable. The question therefore arises how the geographical dimension usually neglected in game-theoretic approaches to international politics can be suitably represented within the framework developed so far.

Imagine the countries on the globe. Each country has at least one, but usually several, neighbors. Since two countries may also be connected by regions like the sea which nobody owns, even fairly distant countries may be considered "neighbors". The relevant dimension here is the two dimensional space. However, to simplify matters, its crucial feature – geographical prox-

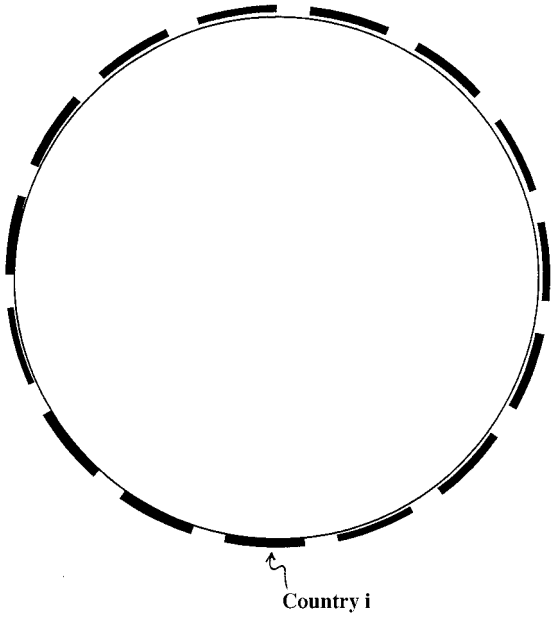

Figure 2.

imity – can even be represented in an abstract way in a one-dimensional space by taking a slice through the globe, rather than the globe itself, to represent the distribution of countries. Thus, in an idealized representation of proximity and neighborhood that abstracts from all other geographical information including differences in the size of the countries' territory, there are n countries scattered along a closed circle with n sections as in Figure 2. For expository convenience let $n = 2s + 1$ be an odd number so that for any given country $i = 1, \ldots, n$ the set of (more or less distant) "neighbors" can be divided into two groups of equal size s to the left and the right of i. With respect to country i, the geographical proximity of country j can be expressed by counting the number of sections on the circle to the left *or* to the right whichever is smaller.

As the countries interact with each other there is always some probability that a regional political conflict will emerge between them. If so, imagine, as before, that it is a bilateral conflict. In reality, complex geographic, economic, and ethnic conditions usually influence the likelihood of a bilateral conflict. However, since an analysis abstracting from the specific details of the singular historical case lacks information about these determinants, only one factor will be treated here as systematically affecting the probability of a conflict occurring, namely geographic proximity. More specifically, a bilateral conflict will be assumed to be the less likely the greater the distance between two countries, i.e., the larger the (smallest) number of sections between them.

Furthermore, the occurrence of a conflict in which country i is involved is supposed to be symmetrically distributed over country i's left and the right hand range on the circle – which means an equal chance of $1/2$ of the next conflict occurring on either side.

Let $p_{i,j}$ be a probability measure for a conflict occurring between country i and any country j on *one* side of i so that the likelihood-decreasing-with-distance assumption and the constraint $\Sigma_j p_{i,j} = 1/2$ are satisfied. Such a measure is

$$p_{i,j} = (s - j + 1)/s(s + 1), \quad 1 \leq j \leq s. \tag{3}$$

Now define a neighborhood N, adjacent to country i on one side, by a closed set (a "segment") of k sections on the circle. By summing over (3), the probability of a conflict occurring between country i and some country in the defined neighborhood N follows as

$$\Phi_k = \sum_{j=1}^{k} p_{i,j} = (2sk - k^2 + k)/(2s(s + 1)) \quad \text{for } 1 \leq k \leq s, \tag{4}$$

where

$$o < \Phi_k < 1/2 \text{ for } 1 \leq k < s$$

and

$$\Phi_k = 1/2 \text{ for } k = s.$$

This probabilistic model for the occurrence of regional conflicts provides a basis for investigating the initial question more thoroughly in the next section of how, and under what conditions, "moral" conduct can survive and spread out in international politics.

5. Innovation and imitation in international politics

With the extended perspective of the previous section we may now return to the problem posed in Section 3 and try to assess the chances of moral conduct occurring and prevailing in regional conflicts in international politics, both with respect to the first move and in encountering the dilemma between "cheap" appeasement and "costly" belligerent moralism in the second move. The relative costs or benefits a country experiences with a certain policy depend on the likelihood of being involved in a conflict with an opponent following some particular kind of strategy. Information about the latter factor, i.e., moves and outcomes created in regional conflicts, is usually communicated worldwide and becomes common knowledge among all countries. With

respect to Assumption 2, this means that the public in a country is likely to be able to assess success and failure, or costs and benefits, of a particular policy type that accords or conflicts with its standards of morality even though the country may not have been involved in any conflict itself. In the light of the experience of other countries, public opinion may then press the government to adopt a certain policy type when the country itself becomes involved in a conflict in the future. This is particularly true with respect to new policy types – moral and immoral ones – which may be introduced somewhere in the world and tried out in a regional conflict against policies already existing in the international arena.[9] When they fail to live up to the expectations of the inventor they are likely to be abandoned. When successful they are likely to be maintained. Such an experiment has a vicarious character for all other countries, and it is the dissemination through imitation of successful innovations that may threaten and erode established standards of conduct and pave the way for evolutionary change in international politics. In an evolutionary analysis, strategic knowledge thus changes as a result of discovery, experimentation, and observational learning or, to put it the other way round, the set of international political strategies *known* to the international players and their prevailing public opinion at a certain point in time is not invariably given.

In the most elementary case, to be used for the present expository purpose, there is just one policy type initially. Accordingly, the analysis starts from a situation in which, in all regional conflicts, one and only one sequence of moves in the above two-stage game is being played. Now let a country which just engages in a newly emerging bilateral conflict somewhere in the world come up with a policy innovation, i.e., an alternative sequence of moves. The outcome of the innovation hinges on what the prevailing policy type – which, by assumption, will be pursued by the opponent in the game – dictates as a response. This is a modified form of an important contingency called "occupancy effect" in evolutionary theory (Witt, 1992). Once the outcome is revealed, it informs those international players who happen to be involved in the next regional conflicts, i.e., play the two-stage game the next times. Where public opinion formation has an influence this will be reflected in the conclusions and, accordingly, the chosen policy.

To be more specific about these conclusions, two cases have to be distinguished, that of the innovating country and that of the other countries. In line with Assumption 2, a democratic country is unlikely to innovate with a defective move in the initial stage of the game. If it innovates, it will therefore introduce move c. Similarly, including the closing stage, the only innovation can be to switch from AP to BM or vice versa. For a non-democratic country, by contrast, these constraints are not binding. Independent of the motives

for, and constraints on, the innovative political move entered, it can be hypothesized that, should the country again be involved in an international conflict, the question of whether to reuse or abandon the innovative move will be decided according to the pay-offs realized. (In the case of a "moral" political innovation this includes an implicit answer to the sustainability question.) We thus introduce:

Assumption 3 (innovator's behavior). Independent of the motives for trying an innovation, the innovating player reuses the new strategy when involved in another international conflict in the future, provided that the new strategy does not result in a lower pay off than the one previously played. Otherwise the former strategy is played again after trying out the innovative strategy in a small number of conflicts (persistence interval).

As far as the conclusions are concerned which the non-innovating countries draw from what they observe the difference between democratic and non-democratic countries seems less relevant. The public opinion in democratic countries may press for upholding "moral" conduct or for adopting it when it has been exercised elsewhere. Yet, if the costs of doing so in an otherwise immoral world are significant, or even excessive, this seems to be an argument that undermines the belief in morality being sustainable. Accordingly, one may hypothesize:

Assumption 4 (non-innovator's behavior). Players observing a new strategy switch, when they themselves are involved in an international conflict in the future, to the new strategy with a probability which depends on the relative size of the difference between the pay-off they can realize with the new strategy and the pay-off of the old strategy. The probability

- is zero if the innovator's strategy yields a pay-off no higher than the prevailing strategy;
- is positive and varies monotonously with the (positive) difference between the pay-offs of the innovative and the prevailing strategy.[10]

Taken together Assumptions 3 and 4 induce a selective replication device which, in turn, allows to determine whether some particular strategy is evolutionarily stable.[11] This is a well-established concept in evolutionary analysis for measuring the success or, perhaps better, persistence of a particular strategy in a game played by the members of a population. Indeed, this concept will be used in the next section to address the initial question of

whether, and under what conditions, "moral" conduct can indeed hold out in international politics.

6. Moral conduct and the occupancy effect: An evolutionary analysis

The model developed over the previous sections allows the chances of different forms of conduct in international politics to be assessed, contingent on the particular composition of countries and policy types prevailing in the "population" of international players, and depending on the geographically determined likelihood of an interaction. Under the simplifying assumptions that have been made, a strategy innovations enters the scene at a certain point of time. Up to that point, international politics has been characterized by one exclusively played strategy. Depending on the realized pay-offs, the innovation may, or may not, be maintained by the innovating country in its future interactions (survival). If it survives, the innovation may be imitated by other countries so that it spreads out (dissemination).

In principle, two situations can be distinguished and the distinction is crucial for the present problem. First, the innovation may be introduced by a single and, hence, isolated country. Thus, one country may either introduce defection (playing LA or DC depending on whether the opponent chooses c or d in the first stage of the game) to an otherwise cooperative international environment, or it may introduce cooperation to an otherwise defective environment. By Assumption 2, in the former case the innovating country is likely to have a non-democratic government and/or a public opinion formation process manipulated by radical values, such as ethnic or religious chauvinism. In the latter case, the innovator is likely to be a democratic country. What happens under Assumptions 1, 3 and 4 once the innovation has been entered can be stated as

Proposition 1 (solution for the single innovating country case).[12] The fate of the respective innovations when confronting a certain prevailing policy type are those given in Table 2.

The results with respect to moral conduct in international politics show a remarkable, but not surprising, asymmetry. Neither of the moral variants can gain a foothold in international politics in a state of global defectionism, or anarchy, when introduced by one isolated democracy. The reasons for this differ (see the Appendix). Conversely, defectionism as innovation has much better chances of overthrowing a regime of peaceful coexistence in international politics than vice versa, though there are differences here as well: belligerent moralism is much less vulnerable than is the permissive appeasement variant. In fact, permissive moralism is doomed to be driven to

Table 2. Result of invasion of a prevailing strategy by an innovative strategy

	Innovative strategy		
Prevailing strategy	Defection (LA or DC)	Permissive moralism (AP)	Belligerent moralism (BM)
Defection (LA or DC)	–	No survival	No survival
Permissive moralism (AP)	Sure dissemination	–	Survival without dissemination
Belligerent moralism (BM)	Dissemination possible	Survival without dissemination	–

extinction.[13] The breakdown of AP may happen in different ways: a country which played AP in the past may be a non-democratic one with a government that, in a Machiavellian fashion, turns around once it observes that the peaceableness of other countries can be exploited; or, in a democratic country which played AP, the public fears the costs of the heavy burden of retaliatory action; or, because a defeatist public opinion rejects both retaliation and a defective mode of behavior, democracy may eventually be overthrown or the country be conquered.

A different situation emerges where the innovation is not introduced by a single, isolated country, but where innovators appear "in clusters", i.e., several neighboring countries make an innovation in international politics at the same time. In this case, spatial proximity as discussed in the previous section matters crucially, since a cluster is defined as a neighborhood N, i.e., a segment (or closed set) of k sections on the circle. An innovative policy type simultaneously adopted by such a cluster of neighboring countries is certainly not an entirely unrealistic event in international politics, particular if the innovation is oriented towards moral conduct. Because of their cultural similarity (correlated with spatial proximity) such neighboring countries may have similar democratic backgrounds and similar values in public opinion formation. Diplomacy may help to coordinate the simultaneous transition to a new policy type. If these conditions are met, the chances for moral conduct surviving and/or disseminating are much better, provided a critical mass of countries can be induced to join.

Proposition 2 (solution for innovators "in clusters" case).[14] If defection

is the prevailing policy type, belligerent moralism (BM) can gain a foothold as an innovation and can disseminate, if adopted simultaneously by a critical number m^* of countries forming a cluster, where $(n-1)/2 < m^* \leq n-1$. No such number can be derived for the same to hold true for AP.

The result offers a more optimistic outlook on the prospects for establishing moral conduct in international politics. Even more important is, however, that it may help to explain the historical contingencies of the rather rare periods during which peaceableness did indeed prevail in international politics, at least in certain regions in the world. In the light of the present analysis, the long period of peaceful coexistence of the Western countries since World War II – unprecedented in the more recent history at least of Western Europe – may, for example, be traced back to a double contingency: the rise of stable democracies in the Western countries *and* their collective, cluster-like, adoption of the cooperative behavior in international politics, first under the guidance of the North Atlantic Treaty Organization and later under that of the European Union.

As the analysis suggests, a crucial prerequisite for moral conduct to prevail in international politics is belligerence. Organizing the support for a collective transition to a new policy type by a number of countries which exceed the critical mass is one thing. Getting all those countries to stick to the collectively agreed change when the costly, and sometimes violent, punishment of offenders indeed becomes necessary is another thing. A problem with Proposition 2 is that, in order to be valid, each country in the cluster must indeed be irrevocably committed to keeping to the innovative policy type BM. Although this is beyond the framework of the present model, international assistance and defense treaties, like the NATO, may be conjectured to be decisive institutional safeguard in this respect. Nonetheless, historical experience –not least that of the regional conflicts after the end of the Cold War – shows that public opinion formation processes in the Western democracies are not always sufficiently aware of the problem and are tempted by the moral appeal of an appeasement policy. However, paradoxical as it may intuitively appear, within the framework discussed here, agitation in favor of appeasement as a means of resolving conflicts can unintentionally cause moral conduct to crumble in international politics.

7. Conclusions

An attempt has been made in this paper to address the tensions which can often be observed to turn up between notions of moral conduct in international politics and the rational pursuit of national interest in regional conflicts. In

order to make the problem accessible to a more rigorous analysis, a two-stage game has been developed as an abstract model of such conflicts. The investigation of the model has shown more clearly where those tensions come from and, using the analytical tools of an evolutionary approach, some conjectures could be derived as to how, and under what conditions, moral standards can emerge and prevail in international politics. What these conjectures suggest is that moral conduct does have a chance of persisting once it prevails. For this chance to be used it is crucial, at least within the simplified framework analyzed here, for moralism to be paired with belligerence. Immoral political conduct must be faced with discouraging consequences. The intuitive logic of this result may explain the frequent call in public opinion formation processes for "setting a warning example" against offenders though this is usually costly and, indeed, is itself often subject to moral objections. Belligerent moralism is therefore vulnerable to being overly challenged. For this reason, it is more likely to survive under conducive conditions, i.e., under conditions which are the same as those that have allowed moral conduct in international politics to emerge in the first place.

Indeed, as it has turned out, the true puzzle for the analysis is to explain how moral conduct could ever come into being in a world of Machiavellian attitudes in international politics. This appears to require some very special circumstances, a fact which may explain the historically rather rare, and rather late, instances of moral conduct in international politics. Following the behavioral assumption in this paper about the governments' choices of strategies in the conflict game, standards of moral conduct can only be expected to be influential where two things come together (as in the more recent history in the Western democracies): moral value judgements do indeed have to inform public opinion formation processes and these processes have to have an impact on government behavior. Even then, however, an additional condition must be met. Since a single democracy following a belligerent moralism would quickly be overly challenged, it seems necessary for democracies to appear in regional "clusters". Only then is the probability of interacting in peaceful coexistence rather than being forced to exercise retaliatory action against Machiavellian aggressors sufficiently high for moral conduct in international politics to be viable. With respect to political history, these results may be taken to explain why, with the rise of stable democracies in Western Europe after World War II, this part of Europe has had an unprecedentedly long period of peaceful coexistence. The results may also be taken to reflect on the conditions which will have to be met in the future if the "peace dividend" can continue to be earned.

Notes

1. This means assuming that no country has a comparative advantage in a possible military combat.
2. International conflicts may often be described alternatively as a game of chicken (see, e.g., Brams, 1985; DeNardo, 1995). The core issue in such a game is to establish a dominance-subordination relationship. Military threat, often taking the form of an arms race, is used, but an armed conflict does not necessarily take place. The reason is that, unlike the prisoners' dilemma game, in a game of chicken it is rational for one side to give in to the other before the escalation reaches the level of warfare.
3. Relation (1) is assumed to hold for each country.
4. This contrasts with the situation in a repeated game where the unaltered pair of one-period strategies, c, d is played sequentially a large number of times or even indefinitely. In such a setting, retaliation would, for instance, have to be expressed by a specific sequence of one-period strategies c and/or d. The well-known tit-for-tat strategy in the repeated prisoners' dilemma (Axelrod, 1984) may be considered an expression of such a kind of retaliation, though its applicability to international politics may be doubtful (Hirshleifer, 1998). Indeed, it seems open to debate whether or not a single regional conflict between two countries is more suitably modeled as a two-stage game or an (almost) infinitely repeated game. Presumably, both ways of doing it involve serious, though differing, simplifications.
5. The assumption is a logical consequence of the semantic meaning of "retaliation". However, as simple calculations on the basis of the order relations (1) and (2) below show, move r is also always dominated by move a for both players in the closing stage.
6. The order relation (2) implies that vengeance is effective in so far as the country that has exploited the peaceableness of another country does not get away with the tempting pay-off T. However, the country which effects this through retaliatory action is made even worse off.
7. If both countries in a conflict defect in the initial move, the outcomes in Figure 1 suggest refraining from move r in the closing stage, because TW is dominated by DC against DC as well as TW. For this reason, the contingent strategy TW can be neglected in the remainder.
8. It may be argued, for example, that the policy pursued by Israel in the Near East has for a long time been inspired by such a commitment.
9. In the individual historical case the policies pursued in a regional conflict are characterized by a multitude of aspects. Even though it is novel in many aspects, an "innovation" of this kind may nonetheless, in abstract terms, be classified into one of the policy types (contingent strategies) of the previous section.
10. Assumption 4 does not imply that the adoption probability is identical for all countries. The adoption probability may instead be assumed to vary with the likelihood of being involved in a conflict with a country pursuing the new policy type. This, in turn, hinges on the geographical proximity of the two countries as explained in the previous section. The gradually increasing adoption probability – rather than a zero-one jump – accounts for the imponderables in public opinion formation and political decision making. Note that in the probability model used here the expected waiting time (the number of conflicts a country is involved in) before a switch is made is inversely related to the probability of making the switch. (If the adoption probability of country i is q_i, and if, *ceteris paribus*, it is treated as a constant, the waiting time is a random variable with a geometrical distribution and an

expected value of $1/q_i$.) Assumption 4 thus implies that the expected waiting time is the shorter, the greater the relative disadvantage of the old policy type.

11. An evolutionarily stable strategy was originally defined (see Maynard Smith, 1982: 10–20) as a genetically fixed strategy which, if prevailing in a population, makes it impossible for a mutant strategy m to invade the population in the presence of natural selection. More precisely, let e, m be two strategies, and let $E(\pi(i,j)), i, j \in \{e, m\}$ denote the expected value of strategy i against strategy j. Then, e is evolutionarily stable if and only if either $E(\pi(e, e)) > E(\pi(m, e))$ or $E(\pi(e, e)) = E(\pi(m, e))$ and $E(\pi(e, m)) > E(\pi(m, m)$.
12. For a sketch of the proof of the proposition see the Appendix.
13. When relaxing the assumption that only one of the "moral" policy types prevails exclusively when defection comes up as an innovation, it is easy to imagine that the results change only in degree and not in principle. The greater the share of AP in the population of international players, the less effective (and credible) is the belligerent attitude of the countries following BM in deterring defection and the more likely it is that defection will disseminate.
14. For a sketch of the proof see the Appendix.

References

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

Bernholz, P. (1985). *The international game of power*. Berlin: Mouton Publ.

Bernholz, P. (1995). Efficiency, political-economic organization and competition between states. In G. Radnitzky and H. Bouillon (Eds.), *Values and social order*, Vol. 2, 157–198. Aldershot: Edward Elgar.

Brams, S. (1985). *Superpower games: Applying game theory to superpower conflict*. New Haven: Yale University Press.

Bueno de Mesquita, B. and Lalman, D. (1992). *War and reason: Domestic and international imperatives*. New Haven: Yale University Press.

DeNardo, J. (1995). *The amateur strategist: Intuitive deterrence theories and the politics of the nuclear arms race*. Cambridge: Cambridge University Press.

Gaddis, J. (1992/93). International relations theory and the end of the Cold War. *International Security* 17: 5–58.

Gauthier, D. (1985). *Morals by agreement*. Oxford: Oxford University Press.

Hirshleifer, J. (1999). There are many evolutionary pathways to cooperation. *Journal of Bioeconomics* 1: 73–93.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

McClennen, E. (1989). *Rationality and dynamic choice: Foundational explorations*. Cambridge: Cambridge University Press.

Niou, E., Ordeshook, P. and Rose, G. (1989). *The balance of power*. Cambridge: Cambridge University Press.

Schelling, T.C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Sen, A. (1987). *On ethics and economics*. New York: Basil Blackwell.

Snidal, D. (1985). The game theory of international politics. *Journal of World Politics* 38: 25–57.

Witt, U. (1992). Evolutionary concepts in economics. *Eastern Economic Journal* 18: 405–419.

Wolfson, M., Farell, J., Gill, A. and Shabahang, H. (1992). *Essays on the Cold War*. London: Macmillan.

Appendix

Sketch of proof of proposition 1

Consider the policy types entering as an innovation into an environment with prevailing policy type as given in Table 2 in the order of the columns from left to right and in each column from top to bottom. Thus

- *defection invading permissive moralism (AP)*: initially, the innovator plays LA against AP and realizes a pay-off $T > R$ with certainty. The more the defection as a behavior disseminates, however, the more often another defecting country is encountered and DC played against DC by assumption 1. The innovator's expected pay-off then gradually converges to $P < R$. Since returning to AP would entail R when playing against AP-countries and S when playing against aggressors, by assumption 3, the innovator never switches back to AP. Countries playing AP obtain a pay-off R among themselves while, as the innovator's opponent, they get S. If they switch to defection, they realize $T > R$ against AP-countries and $P > S$ against the innovator (and, later, other aggressors). This means that AP is an inferior strategy under all circumstances. Hence, by assumption 3, there is a finite waiting time for all countries to switch to defection. Defection disseminates and drives AP to extinction.
- *defection invading belligerent moralism (BM)*: initially, the innovator realizes with certainty a pay-off $T - V_p$ which, by the other relations (1) and (2), is smaller than R. Should the defection disseminate as a behavior, the innovator's expected pay-off gradually increases to $P > T - V_p$. Since returning to BM would entail R when playing BM-countries and $S - V_r$ when playing against defecting countries, the innovator's future behavior according to assumption 2 crucially hinges on whether defectionists can ever gain a significant share among the countries. Consider therefore the countries playing BM and obtaining a pay-off R among themselves while, as the innovator's opponent, they get $S - V_r$. If they switch to a defection, they realize $T - V_p$ against BM countries and $P > T - V_p$ against the innovator (and, later, other aggressors). If any country is likely to switch, it will be the innovator's neighbor. Let $E(\pi_{BM})$ and $E(\pi_d)$ denote the expected pay-off of a country playing BM and the initial move d (i.e., either LA or DC) respectively. According to the geographical structure of conflicts as modeled in Section 4, we then have for the immediate neighbor country of the innovating country

$$E(\pi_{BM}) = \tfrac{1}{2}R + \Phi_1(S - V_r) + (\tfrac{1}{2} - \Phi_1)R.$$

In case of switching the neighbor gets

$$E(\pi_d) = \tfrac{1}{2}(T - V_p) + \Phi_1 P + (\tfrac{1}{2} - \Phi_1)(T - V_p).$$

Equating both expected values and solving yields

$$\Phi_1^* = [T - V_p - R]/[T - V_p - R + S - V_r - P] < \tfrac{1}{2},$$

because $0 > T - V_p - R > S - V_r - P$. Hence, it cannot be excluded that the conflict probability Φ_1 happens to satisfy the condition $\Phi_1^* < \Phi_1 \leq \tfrac{1}{2}$. (Note, however, that since $\Phi_1 = 2/(n+1)$, the condition is the less likely to be satisfied, the larger n.) If this is the case, there is a finite waiting time for the neighbor country to switch to defection according to assumption 4. If the innovator maintains the new strategy for long enough, incurring the opportunity loss $R - (T - V_p)$, the expected pay-off may grow to P in the best case (through the neighbors' switching). P is better than $T - V_p$ but still inferior to R so that the innovator may eventually be inclined to return to playing the old strategy according to Assumption 3. However, if a neighbor has already switched such a move would amount simply to changing places with the neighbor. Thus, while a dissemination of defection as a behavior cannot be excluded if it happens at all, it may be a matter of cyclical convergence, probably even infinite cycling. (In case $0 < \Phi_1 \leq \Phi_1^*$, by contrast, switching is excluded so that for the corresponding parameter valued defection as an innovation is certainly unable to survive.

- *AP invading defection*: here the order relation (1) is decisive; initially, the innovator is certain of realizing a pay-off $S < P$. However, should AP disseminate, the innovator's expected pay-off gradually increases to $R > P$. Since returning to defection would entail P when playing against defecting countries and T when playing against AP countries, by Assumption 3 the innovator is to return to defection after playing a number of times AP. As far as the defecting countries are concerned, they obtain a pay-off P in playing DC among themselves while, as the innovator's opponent, they get T playing LA. If they switch to AP, they realize S against defecting countries and R against the innovator. Since $T > R$ and $P > S$ no switching will occur by Assumption 4 and AP cannot survive.
- *AP invading belligerent moralism*: when playing against BM, an AP-innovator cannot be distinguished from BM. The innovator may maintain its "innovative" policy type forever (without any practical relevance), but the policy will not disseminate by being imitated.
- *BM invading defection*: initially, the innovator can be sure of realizing a pay-off $S - V_r < P$. However, should BM disseminate, the innovator's expected pay-off gradually increases to $R > P$. Since returning to defection would entail P when playing against defecting countries and $T - V_p$ when playing against BM-countries, the innovator's future behavior according to Assumption 3 crucially hinges on whether BM can ever gain a significant share among the countries. Consider therefore the countries playing DC and obtaining a pay-off P among

themselves while, playing LA as the innovator's opponent, they get $T - V_p$. If they switch to BM, they realize $S - V_r$ against defecting countries and $R > T - V_p$ against the innovator. By reasoning analogously to the case where defection invades BM a critical value

$$\Phi_1^{**} = [S - V_r - P]/[S - V_r - P + T - V_p - R] > 1/2$$

can now be derived. Hence, there is no value of Φ_1 in the admissible interval $(0,1/2]$ such that a switch may occur according to Assumption 4. Since the innovator is to return to playing defection according to Assumption 3, BM cannot survive.

– *BM invading AP*: when playing against AP, a BM-innovator cannot be distinguished from AP. The innovator may maintain its "innovative" policy type forever, but the policy will not disseminate by being imitated.

Sketch of proof of proposition 2

Within a cluster of BM-innovators, i.e., a closed segment N on the circle on which defection otherwise prevails, the innovators' pay-offs differ significantly between inner and outer members. The bordering members profit least from the cooperation gain and are therefore most likely to switch back to defection. The expected pay-off of such a bordering member of a cluster of k innovators is

$$E(\pi_{BM}) = 1/2(S - V_r) + \Phi_k R + (1/2 - \Phi_k)(S - V_r).$$

Should that member switch back to defection the expected pay-off would be

$$E(\pi_d) = 1/2P + \Phi_k(T - V_p) + (1/2 - \Phi_k)P.$$

Equating both values and solving yields

$$\Phi_k^{**} = [S - V_r - P]/[S - V_r - P + T - V_p - R] > 1/2.$$

Hence, by Equation (3) no number of cluster members $k \leq s$ exists so that $E(\pi_d) \leq E(\pi_{BM})$. By Assumption 3, the bordering innovating member is to switch back in finite time to defection, while, by Assumption 4, no neighbor of the border member outside the cluster (and *a fortiori* no other country outside the cluster) is likely to imitate the innovation. The situation changes, however, if the cluster is allowed to have more than s members. In that case, the respective expected pay-offs are

$$\begin{aligned} E(\pi_{BM}) &= 1/2R + \Phi_k(S - V_r) + (1/2 - \Phi_k)R \text{ and} \\ E(\pi_d) &= 1/2(T - V_p) + \Phi_k P + (1/2 - \Phi_k)(T - V_p). \end{aligned}$$

Equating both values and solving yields

$$\Phi_k^{*} = [T - V_p - R]/[T - V_p - R + S - V_r - P] < 1/2.$$

In view of Equation (4), this means that a critical number of cluster members m^* exists such that, once m^* has been exceeded, all countries will switch to BM and defection will be driven to extinction. Depending on the size of the variables T,R,P,S,V_r, and V_p, m^* varies in the interval $((n-1)/2,\ n-1]$. No analogous argumentation applies when AP occurs as an innovation in clusters, since defection dominates AP under all circumstances.