

EMPIRICAL VALIDATION OF RICHARDSON MODELS OF ARMS RACES

Hans Rattinger

Lehrstuhl für
Politikwissenschaft II
Universität Bamberg
D-8600 Bamberg

INTRODUCTION

Lewis Richardson's (1960a, first published 1939) venture to develop a mathematical theory of arms races was initially brought to the attention of a wider audience when Anatol Rapoport (1957) published his review article "Lewis F. Richardson's Mathematical Theory of War". This piece unleashed a veritable avalanche of formal theorizing and empirical research about arms races in subsequent years. The wealth of theoretical and empirical arms race studies has been well-documented in bibliographies (Cioffi-Revilla 1979) and reviews of the literature (Busch 1970, Moll and Luebbert 1980, Intriligator 1976, 1982). In modern textbooks, formal modeling of arms races is distinguished as one of the major scientific approaches to the study of international relations (Zinnes 1976), doctoral theses (e.g. Schrodtt 1977, Smith 1978) and stimulating monographs (Nincic 1982) have emerged from the field.

The purpose of this contribution can be defined negatively in two respects: neither is it an attempt to summarize, classify, or evaluate the body of existing research on mathematical models of arms races, nor will an effort be presented to empirically validate particular models from the Richardsonian tradition in a specific historical situation. Instead, this article will be devoted to a systematic summary of the most important and pressing difficulties and problems encountered in em-

pirical analyses within the framework of formal arms race models. There is no claim, therefore, to be innovative. The current state of the discipline is to be put into perspective by clarifying the key theoretical and methodological obstacles that arms race research has to cope with. Thus, all that is to be achieved is to illuminate what kind of intellectual challenges have emerged in the process of refining and applying the original models set forth by Richardson over the past quarter century. Such an appraisal should lead to a more adequate grasp of what has and has not been achieved so far and what lies in the future, why certain expectations have been disappointed and why certain avenues for further research appear as more profitable than others. One by one, the following issue-areas will be addressed:

- How do we define "arms races"; what cases constitute our universe for empirical research?
- How do arms race models relate to more general models of arms acquisition?
- How do problems of measurement, operational definition, and available data affect empirical arms race research?
- To what extent has empirical arms race research contributed a meaningful separation of external from internal determinants of military efforts?
- How does the fact that arms race parameters vary within and between arms races affect the theoretical status of these models; what could be done about this?
- Are there specific problems of estimation and methodology that haunt the empirical validation of models from the Richardsonian tradition?
- To what extent are arms races and the onset of war related to each other?

WHAT IS AN ARMS RACE?

In a very general sense, arms race models claim to explain observable behavior of national states. Empirical tests constitute attempts to falsify such models. The key question is, for what cases do those models claim to be valid, unless they explicitly aim at describing particular historical phenomena. This, however, would imply forsaking the general nature of theory as fitting curves to particular historical cases is always possible. The generality and organizing power of theories is the higher the less variables they involve and the higher the number of cases to which they apply.

If empirical arms race research is conceived of as a rigorous attempt to systematically establish the empirical validity of rather general hypotheses, a clear notion of the universe of cases to which these models apply has to exist.

This problem of definition is largely neglected in the arms race literature. Typically, models are constructed, refined, and modified with particular historical confrontations in mind, and the empirical evaluation of the goodness of their fit is performed with that particular case. The requirement to test a general model across a multitude of cases, preferably across all available cases, which would be derived from a strict adherence to the principles of the philosophy of science, so far has not been taken very seriously.

Accordingly, no satisfactory historical inventory of arms races is available: when, where, between which actors and how long they have taken place, when they have started and ended. Such an inventory should be comparable to the historical compilations of wars among nations supplied by Richardson (1960b), Wright (1965), and Singer and Small (1972).

Elsewhere (Rattinger 1975a) I have extensively dealt with the issue of operationally defining arms races, which is an essential prerequisite for systematically compiling such a universe of cases along standardized criteria. To arrive at rigorous and comparable empirical research on arms races, the following main difficulties of the operational delimitation of arms races would have to be solved:

First, if there is consensus that at least bilateral more or less rapid increases of armaments have to occur in order to qualify this process as an arms race, an immediate and obvious linkage to the problem of the appropriateness of indicators arises. In what respect do we have to observe increases, in armament expenditures, in military manpower, in the numbers of weapons systems, in military potential, in the lethality of weapons, in fighting power, or in combat readiness? What does it mean if some of these dimensions build up while others stay the same or even go down? Do we talk about arms races if annual increments of arms spending simply reflect increases in the nominal or real income of military personnel?

Second, do perceptions of hostility and mutual directedness of the military effort have to be an integral part of the operational definition of an arms race? How is mutual stimulation to be ascertained, and can this be achieved without referring to data on perceptions and declarations by both sides? Should we talk about arms races if there are no such mutual perceptions and declarations?

Third, the onset and termination of arms races have to be defined unequivocally. We have to consider, whether this should be done via the behavioral and/or the perceptual component. This problem is essential for research on the relation between arms races and wars. If there is no comprehensive collection of well-defined cases including precise information about onset and termination, very little can be meaningfully said on whether arms races do or do not lead to war.

Fourth, should the term "arms race" be restricted to interactions with a clearly hostile component, do both sides have to perceive military action against each other as one possible course of future behavior? Alternatively, should the term be extended to cases where not military threat, but rather military prestige is at stake, as is often said about the armament build-up in Latin America?

Finally, how are arms races with more than two participants to be dealt with conceptually? This problem has two dimensions. The first is the disaggregation of the military behavior of one nation that is racing against more than one opponent: what components of its military effort are due to each of the races it is involved in, what components have to be counted in all those races? The second dimension is the treatment of arms races between alliances: do such processes constitute races of alliance against alliance, or are they to be decomposed into reaction functions for each individual nation, possibly reacting to all of the other alliance or to particular nations out of this other alliance?

These problems can be illustrated by comparing the lists of cases presented by Smith (1978) and by Wallace (1979, 1982). For the period 1860 to 1977 Smith enumerates 32 arms races, for the longer period since 1816 Wallace lists only 28 arms races. In the period where their lists overlap, both authors quote different cases, and sometimes different time-spans of arms races contained in

both compilations. The most extreme example is an arms race listed by Smith, but not by Wallace, between the Federal Republic of Germany and France starting in 1964. Both authors do not supply operational definitions but only conceptual explications of what they mean the term "arms race" to denote. Other historical inventories of arms races than these two are not available so far.

ARMS RACE MODELS VS. ARMS ACQUISITION MODELS

The problem of defining arms races can be seen in a somewhat broader perspective if one conceives of arms race models as a kind of special case of more general models of the arms acquisition behavior of national states. Arms race models normally do not proceed from a systematic overview and classification of the various determinants of national armaments, but assume enemy arms to be the central determinant of own armament. In doing so, many components are neglected that are known to influence military armament, be it in an arms race or not. Figure 1 represents an attempt to compile various types of determinants of military behavior. It demonstrates that arms race models neglect a series of such determinants. Nations in arms races are distinguished from other nations only by the coefficient a_8 in equation (2) that has to be different from zero. Nations that are not involved in arms races also react to the international situation, simply by maintaining armed forces. The international environment of nations can be subdivided in a threefold fashion: There are those components that are longitudinally comparatively stable, e.g. a nation's position in the international system, its alliance involvement, its geographical position, its historical experience etc. Such variables influence the general level of the nation's military effort; they are responsible, e.g., for the fact that Israel devotes such a high share of its GNP to national security while Luxembourg spends so little. Second, there are dimensions of the international environment of a nation that can vary more rapidly but do not characterize a nation's military environment, e.g. international tension levels, the climate of foreign relations with particular foreign countries, or transnational relations. Finally, the military effort and capabilities of surrounding nations and potential adversaries are the militarily most relevant dimension of a nation's international environment. It can be more or less co-linear with variables from the second group. Stable configurations of the international

environment determine gross levels of military preparedness, short-term fluctuations of the international situation can lead to short-term fluctuations of the military effort that are superimposed upon those gross levels. This is what normally happens for all nations. But only if the fine-tuning of military behavior responds to that of other nations do we talk about an arms race.

To evaluate the importance of other states' armaments for determining a given nation's own military effort by means of established arms race models can be rather difficult. As a rule, they are more or less sophisticated variants of equation (5) and restrict the set of explanatory variables to the armament levels of both sides at a previous point in time.

Even if one assumes the international situation to remain roughly constant during an arms race, this implies a major specification error. If equation (5) or similar equations are estimated, the effects of excluded variables will appear in coefficients b_1 and b_2 , and the error term will no longer be random. Especially the autoregressive coefficient will pick up all kinds of relatively stable but excluded effects, such as resource constraints, economic potential, bureaucratic budgeting, internal lobbying, and the impact of the international environment. Quantitatively weighting the importance of the arms race component against other components of armament increases is extremely difficult when such simple models are applied. If reality should look somewhat similar to Figure 1, models of the type of equation (5), that summarizes the basic Richardsonian ideas, will have a rather hard time to make a meaningful contribution towards the explanation of the causes and consequences of military armaments.

The widespread tendency to conceive of arms race research as separate from the general discussion of determinants of armament behavior, be it competitive or not, seems to constitute a major disadvantage. The program for empirical research should not be the validation of Richardson models, but rather the more comprehensive question "Why do nations arm the way they do?" For this broader question identifying reactive components in a nation's military preparations is just a subset from the overall research agenda. Putting arms race research into the framework of this broader ques-

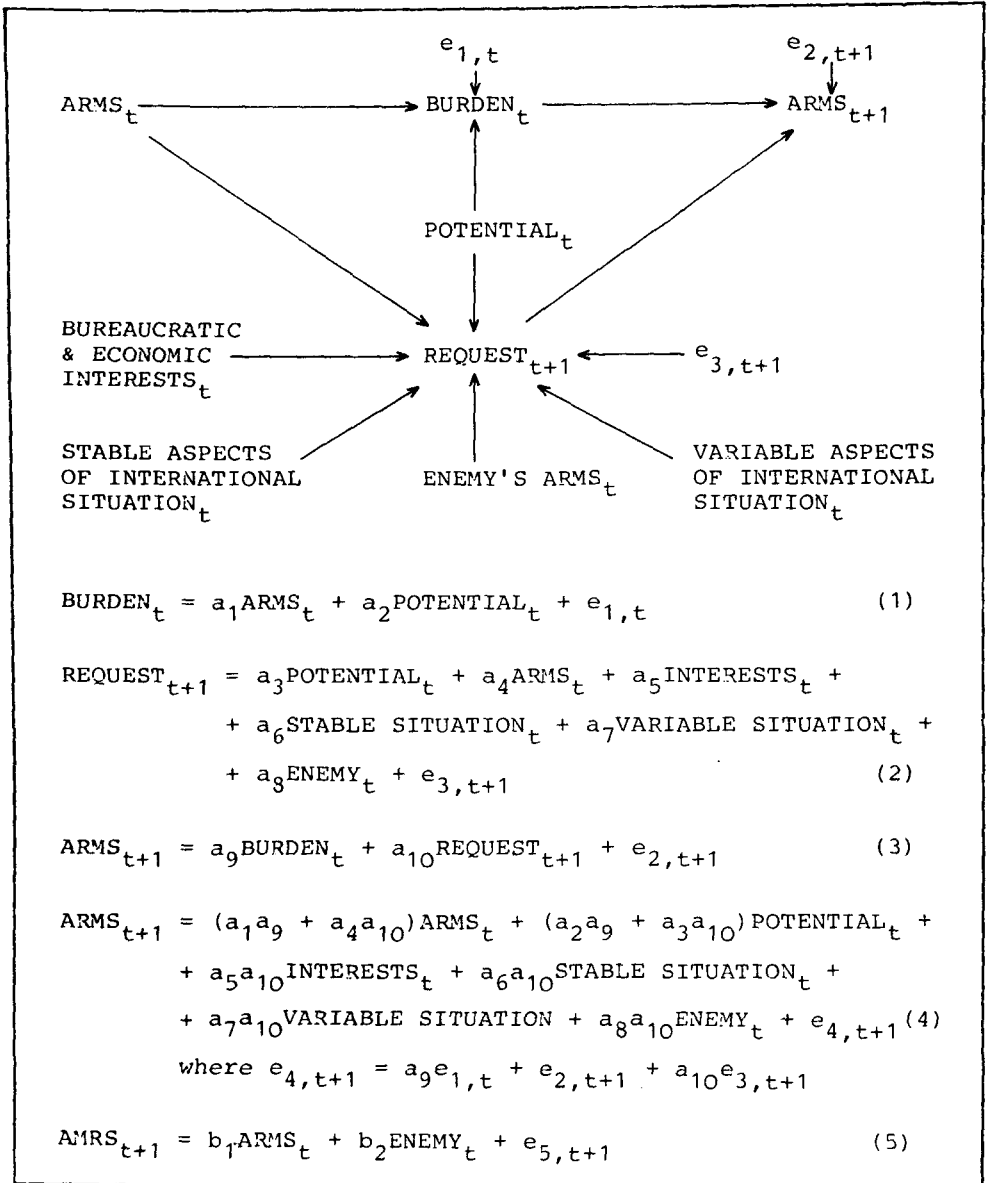


Figure 1: A Model of Armament Behavior

tion also provides us with a much more useful perspective on the issue of how to set arms races apart from the "normal" time-path of military behavior.

MEASUREMENT, DEFINITIONS, DATA

The problem of measuring armaments is important from at least three different aspects. First, it is intimately related to the operational definition of what is an arms race. Second, we do have to have theoretical notions of what we mean by "arms", what we conceive arms races to be about, so we have to be sure that the empirical indicators we use to measure these theoretical concepts are both reasonably valid and reliable. Finally, we have to stand up to the question whether our models and the indicators chosen to connect the realm of theory to the realm of observation are adequate for describing decision-processes going on in the real world. The first aspect has already been discussed above. Whether we believe arms races to be about manpower, military expenditures, numbers of weapon systems, or particular military capabilities (e.g. first-strike capability) has far-reaching implications for what our list of cases is going to look like.

The second aspect, what do we mean by the "arms" or "armaments" of a nation that are to be modelled, how are they to be operationally measured, which data are available and what is the quality of those data, is extremely complex. For a reasonably complete treatment it requires a separate contribution (see the articles by F. Hoerber, P. Stratmann, and R.K. Huber and B. Wobith in this volume). Suffice it therefore, to refer to just a few key concepts here.

The reliability of armament expenditures, as a very popular indicator, is severely reduced by secrecy, conscious distortion on the part of many nations, anything but clear and standard criteria for reporting, etc. The validity of these indicators is questionable for a number of reasons. Armament expenditures confuse stock (salaries, maintenance) and flow (procurement, possibly also R&D) elements. Depending upon the type of recruitment (draft vs. volunteer army) they can measure very different things, longitudinal and cross-country comparisons can be extremely hard to perform due to the necessity to adequately control for inflation, currency conversion, and product-specific purchasing power.

If one chooses to rely on numbers of weapon systems, a series of problems are inevitably encountered. Adequate data have to be found, one has to decide whether numbers are sufficient data or whether some weighting for performance characteristics of weapon systems has to be done (Mihalka 1975). Furthermore, as soon as the common denominator of money is given up, one either has to solve the problem of how to trade off different types of weapon systems against each other or one has to disaggregate arms races into a series of component races (e.g. army, air force, naval races; Rattinger 1976a).

These difficulties get even more intricate if one enters the area of strategic arms races. Do we have to conceive of separate races, one nuclear and one conventional, or is some type of aggregation feasible (Ward 1982)? Within the strategic arms race, the familiar and politically debated distinctions of indicators are relevant for empirical arms race research as well, i.e. launchers vs. warheads vs. throwweight, and static (pre-strike numbers) vs. dynamic (numbers remaining after specific scenarios) comparison.

The third dimension of the measurement problem is the adequacy of the type of reactions and the type of indicators that are being modelled to the real-world decision-processes on armament levels. Normally, empirical arms race research does not at all take this issue into account. One could argue, however, that formal models of arms races and empirical attempts at their validation should be especially useful if they employ those indicators that are of major importance for decision-makers. We have to ask whether the widespread use of highly aggregated military expenditure statistics meets this requirement. Is this type of indicator actually in the foreground of political debate and decision-processes? This is not the place to give a definitive answer, but from this aspect of closeness to actual political decision-processes the use in arms race models of this simple indicator of overall national military spending - that has been criticized frequently and fervently - might even be rehabilitated. Remember, e.g., the heated debate going on in the U.S. since the early Seventies about whether or not the Soviet Union is outspending the U.S. on defense!

On the other hand, the reactions to enemy armament behavior constituting the arms race might rather take

place in singular and discrete identifiable decisions on specific capabilities or procurement programs. Thus, we would have sporadic events of a large-scale response to enemy capabilities (e.g. the SLBM program, MIRV, ABM, or NATO's double-track decision of December 1979), in between these major reactive decisions we would have bureaucratic implementation, improvement, and expansion of those programs. While for arms races over military expenditures, military hardware, or manpower, adequate operational indicators can be designed, at least in principle, for this type of isolated reactive defense decisions, that in political reality might be extremely important, useful empirical indicators are not available. It can be argued, of course, that this type of singular reactive decision-making is not accessible for empirical arms race research, that this is the classical domain of historical case studies which prove superior to macro-quantitative methods of validating Richardson models.

INTERNAL VS. EXTERNAL DETERMINANTS OF MILITARY EFFORTS

As has been mentioned above in the context of Figure 1, weighting internal vs. external determinants of the level of military preparedness should be one of the key challenges for the empirical validation of arms race models. This was seen differently by Richardson himself who regarded the burden of a nation's armaments only as an inhibiting factor against further military build-up. Internal stimulation of armaments is derived from several separate considerations. First, it is argued that threat perceptions and types of behavior resulting from them tend to adopt dynamics of their own, perpetuating increases of armament levels even if the threat from external actors that has initiated them has subsided. Second, military preparations are sometimes regarded as a fiscal instrument, decoupled from specific external threats, to stimulate the economy before elections in order to increase electoral support. Third, internal stimulation of the military build-up is often attributed to the activities of a military-industrial-complex. Finally, internal stimulation can be argued to be important if organizational interests and bureaucratic momentum are taken into account. If one accepts the argument that in the battle for the budget all government agencies attempt to participate "adequately" in the distribution of available or additional resources, there is no reason to expect the behavior of military machineries to be any different.

This modification of Richardson's classical perspective, that military armaments acquired so far can not only impose restrictions upon further military growth but might, instead, stimulate it, for an empirical assessment requires models that are more complex than the ones proposed by Richardson himself, e.g. models of the type that was presented in Figure 1. Available empirical research does not allow clear-cut answers to the question of weighting internal vs. external determinants of armaments. All that can be confidently said today is that military build-ups of nations can be attributed to both groups of explanatory variables without any simple rank-order of importance. This has become very clear in the recent contribution by Cusack and Ward (1981), and has been demonstrated in detail by Russett (1983).

The difficulties encountered when trying to disentangle those factors can briefly be demonstrated by my own work (Rattinger 1975b) on the arms race between the Atlantic Alliance and the Warsaw Pact up to the mid-Seventies (Table 1). At that time, this probably constituted the most complete effort to isolate the effects of those two groups of explanatory variables on the increases of armaments. The analysis essentially consisted of two-step regression models being run. In the first step, military expenditures were regressed upon those of the previous year. In the second step, residuals from these first regressions were run on identically constructed residuals for the opposing alliance. By definition, from the residuals from the first group of regressions all types of trends due to internal stimulation of armament levels have been eliminated. As a comparison of the standardized regression coefficients in the first two columns of Table 1 demonstrates, the auto-regressive component for most of the investigated nations was much stronger than the reactive component measured by regressing a nation's residual defense spending upon residual defense spending of the hostile alliance.

In view of the high multicollinearity between the national time-series of data, however, these findings have to be interpreted with some caution. The attempt to account for as high a share of the variance as possible by a first-order auto-regressive process is statistically conservative, loading the findings in favor of rejecting the hypothesis of the existence of reactive patterns. It can also lead, on the other hand,

Table 1: Standardized Regression Coefficients (β) for Best Fitting Models, 1950-1975, Source: Rattinger (1975b)

Country	Action-Reaction Models		Tension-Armament Models		Combined Models		Residual defense spending of other alliance by Tension	
	Residual defense spending of other alliance		Trend	Tension	Residual defense spending of other alliance			
	Trend				Trend	Tension		
Belgium	.94	.17	-	-	.90	.14	.06	-
Denmark	.96	.09	-	-	.95	-	-	.10
France	.90	.21	-	-	.81	.19	.04	-
West Germany	-	-	-	-	-	-	-	-
Greece	.96	.07	-	-	.95	-	-	.09
Italy	.98	.02	.96	.08	.95	-	-	.09
Luxembourg	.65	.15	-	-	1.00	.56	.53	-
Holland	-	-	-	-	.83	-	-	.21
Norway	.94	.14	.96	.11	.90	-	-	.16
Portugal	-	-	-	-	-	-	-	-
Turkey	-	-	-	-	-	-	-	-
Britain	.36	.44	-	-	.84	-	-	.09
Bulgaria	-	-	.97	.20	.94	.13	.14	-
Czechoslovakia	.96	.18	.92	.16	.91	.26	.15	-
East Germany	-	-	-	-	-	-	-	-
Hungary	-	-	-	-	.94	-	-	.24
Poland	.99	.02	.99	.02	.98	-	-	.03
Rumania	.98	.01	.97	.10	.93	.02	.11	-
USSR	-	-	.97	.12	.94	-	-	.05
European NATO	.95	.10	.95	.10	.92	-	-	.14
WTO	.96	.03	.94	.14	.94	-	-	.11

to an artificial diminution of the importance of enemy arms. The empirical weights contained in Table 1 are the product of a previous theoretical decision on the sequence of entering explanatory variables. If this sequence were reversed, results most likely would look much different. For time-series data with a high degree of multicollinearity - that are typical of arms race situations - the problem of numerically precise quantitative separation of internal from external determinants of armaments by means of formal arms race models has not yet been satisfactorily resolved. We oftentimes can validly state that both dimensions play an important role, but are unable to identify one as stronger or even assign precise numerical weights to them.

UNSTABLE PARAMETERS

Another problem of utmost importance for the empirical validation of arms race models in the Richardsonian tradition is the instability of model parameters within and between arms races. Richardson's classical models allow this to happen as little as the models

presented in Figure 1 do. The survey of the literature by Moll and Luebbert (1980) clearly demonstrates that in a multitude of empirical arms race studies a wide range of parameter estimates is reported. Sometimes reaction coefficients have a negative sign, sometimes "expense and fatigue" coefficients are positive (which, contrary to Richardson's expectations, suggests internal stimulation, of course), sometimes there are drastic changes in parameters if arms races are broken down into separate periods (Rattinger 1975c, 1976a).

To a certain extent the obvious reason is that many authors construct different models for the empirical analysis of different historical cases. However, there is a fundamental difficulty, too, i.e., the causal structures represented by the models are not invariant across space and time. This is a general difficulty of the empirical application of simple formal models in the social sciences. Richardson as a physicist maybe expected to detect some kind of "natural constants", by means of his simple models. In the meantime, we know that even with far more complex models there is no guarantee that causal structures will be constant. The social world, the world of armament behavior, seems to be much more complex than even rather refined systems of equations can grasp.

The direction where a solution to this problem might be found is indicated by Figure 1. Richardson treated nations involved in arms races largely as black boxes. By incorporating the economic system, bureaucratic and social processes, the black box is somewhat opened in Figure 1. But this could go a lot further, e.g. by incorporating some of the conditions for varying reactivity vis-à-vis enemy armaments into the models, by specifying parameters, that notoriously fluctuate, themselves to be functions of other variables that grasp the conditions under which such fluctuations take place. To put it short: In this situation it appears mandatory further to increase the complexity of arms race models, to expand them into open models with exogeneous variables, with variable parameters and interactive terms. The complexity of the models has to be brought somewhat closer to the complexity of the processes that are to be modelled.

Ostrom (1977) and Ward (1982) in their contributions have heeded this necessity to a certain extent. Generally, however, this idea so far has not yet become

very popular, partially so because arms race research has been developed more from the side of pure theory than in response to the results and problems of empirical analysis. For those primarily interested in formal models regardless of empirical validation, abandoning the black-box-approach delivers a severe blow, of course, as a multitude of derivations that have been extensively pursued since Richardson's original work (e.g. on conditions for equilibrium and criteria for stability) are no longer available for complex, non-linear models involving exogenous variables.

My own work on the East-West arms race contains some experiments with open models (Rattinger 1975b). International tension (Goldmann 1974) was introduced into the residual regression analysis briefly described in the previous section as an additive exogenous variable as well as interacting with the armament level of the hostile alliance, serving as a kind of "filter" that would make enemy arms a more or less pressing concern in decisions on defense with growing or receding tension levels. Estimation results for these models are also presented in Table 1 in the form of standardized regression coefficients. This is not the place to repeat the substantive interpretation set forth in the original publication. Suffice it to say, that these ideas to investigate under what conditions various influences on armament behavior might be more or less relevant and to explicitly include these conditions into one's models via exogenous variables might, in fact, be very simple. However, they were not being pursued before, at a time when empirical arms race research was still heavily shaped by considerations of the formal elegance of the models, and less by criteria of empirical validation. Serious pursuit of these and similar simple ideas should prove a profitable strategy for future empirical arms race studies.

METHODOLOGY AND ESTIMATION

So far, this contribution has focussed upon the theoretical, conceptual, and operational problems encountered in the empirical validation of arms race models. Obviously, it is impossible to give a reasonably comprehensive presentation of the methodological issues facing such exercises here. Many of the most important topics have been raised by Luterbacher (1975) and myself (Rattinger 1976b). The discussion of methodological problems will therefore be confined to a

brief enumeration of some of the more important aspects.

First, empirical arms race research is haunted by multicollinearity among model variables. Time-series data of armaments, of economic potential, of economic burden, of requests of funds for the military, etc. are notoriously very highly correlated. This does not facilitate estimation of model parameters.

Second, empirical arms race studies typically are not dealing with sample data, but with population data. The relevance and status of conventional statistical significance criteria therefore has to be critically reviewed and clarified. Even though almost all published work in the field heavily relies upon the basics of inferential statistics, one might argue that considerations of significance should be replaced by those of reproducibility, e.g. whether a given model representing a particular causal structure is confirmed in various instances across time and space with roughly comparable numerical parameter estimates.

Third, the availability of sometimes only very short time-series prevents the application of the same type of more complex models that have been developed for arms races of longer duration. If there are just a few points in time of measurement while we know that particular variables have to be included into models, there is a choice between the devil and the deep blue sea. Either one knowingly commits specification errors by setting up parsimonious models that can be estimated or one sticks to a model that is known to be more adequate but cannot be estimated as too little degrees of freedom are left.

Fourth, empirical application of arms race models often involves estimation of autoregressive processes from rather short time series. From the econometric literature it is known that this is an estimation situation where bias can occur. If lagged endogenous variables are combined with autoregressive error terms, this estimation problem is confounded still. Therefore, everyone empirically estimating arms race models should be aware that research on the properties of estimation procedures for stationary and explosive autoregressive processes with systematically interacting error terms has not yet been developed very far. Thus, employing standard estimation techniques involves quite some confidence that major distortions do not occur.

Finally, the estimation of discrete vs. continuous models has led to some heated debates, for some researchers this seems to be an article of faith. Differential equations, of course, are the original Richardsonian vehicle, but due to the discrete nature of available data, many researchers have found it quite obvious to translate the logic of his arguments into difference equations. This also greatly facilitates empirical estimation. Recently, however, interesting new approaches to estimating differential equations from discrete time-series data have become available (Ward 1982), and it is conceivable that there will be a renaissance of differential equation models of arms races in the future. Naturally, proponents of differential vs. difference equations so far very neatly have divided along the line of theorists vs. empiricists, as the former type of equations lends itself much more easily to formal derivation while the latter type of equations is more handy for translation into empirical research.

ARMS RACES AND WARS

There are numerous exhortations empirically to investigate the relationship between arms races and the onset of war (e.g. Singer 1970), we lack, however, systematic research, but mainly encounter preconceived notions and judgments. These fall into two categories, hypotheses that military preparations make peace more secure (*si vis pacem, para bellum*), and hypotheses that accumulation of arms makes the violent conduct of conflict more likely. The first hypotheses are basic to deterrence theory, the second hypotheses are characteristic of Richardson's work. They reflect the fear of many people that arms that get stockpiled eventually get used.

We face contradictory hypotheses, but the available empirical evidence is rare. The contributions by Wallace (1979, 1982, 1983; see also Weede 1980, Altfeld 1983) and Smith (1978) very well illustrate the two main approaches to the empirical study of the arms race-war connection and their major problems. The key issue in the empirical assessment of this connection is the extent to which arms race modelling and estimation has to be pursued before anything meaningful can be said on whether or not arms races make wars more likely. Does one only have to ascertain the existence of arms races, or does one have to ascertain formal properties

of these processes that can only be evaluated in the framework of formal models?

In the first case, the task is much more simple: One has to identify arms races and to establish whether they have been followed by war. This produces cross-tabulations as Wallace has presented them. In the second case, duration, speed, acceleration, equilibrium and stability of arms races have to be evaluated and to be related to the outcome of the process. This latter research strategy requires the complexities of formal modelling of arms races to be satisfactorily resolved before statements on the relationship between arms races and wars become feasible.

Wallace has chosen the simpler first avenue by identifying arms races along the criteria for "serious disputes", ascertaining whether prior to or during these serious disputes rapid escalations of military efforts have taken place and whether war has followed. Smith has chosen the more difficult research strategy, identifying arms races from articulations of political leaders on reactive components in their nation's armament behavior. She also considerably differentiates explanatory as well as dependent variables, not only recording, e.g. whether or not wars have occurred, but also registering various indicators of magnitude and duration of armed conflict. According to criteria of equilibrium and stability she attempts to identify different patterns of arms races, which presupposes, of course, extensive empirical estimation of formal arms race models.

Describing Wallace's research design as rather simple is not meant to be derogatory. By virtue of the simplicity of the research design he manages to avoid many problems that Smith faces by sticking rather closely to the Richardsonian tradition. She essentially estimates models that are very close to the basic Richardson equations. As has been explained in the third section above, these equations are not overly realistic. Their empirical application commits drastic underspecification and yields unreliable estimates of the reactive component of national armament policies, but they have nice and simple formal properties, especially on equilibrium and stability. If more realistic open models with the most probable exogenous variables were employed, these elegant formal properties would be lost. Smith has decided in favor of the first alternative with the result that she gets a multitude of bizarre estimation

results, to which she nevertheless courageously applies the criteria for stability suggested by Richardson.

We have here, again, said a lot on models of arms races instead of on the arms race-war connection. This has been necessary in order to clarify what choices we face when having to decide on an adequate research design to study this relationship. Existing arms race research seems to permit the conclusion that it is quite straightforward to show that the presence of military armament of a particular gross level is related to the existence of armed forces in other nations, but that it is extremely hard to demonstrate precisely how specific levels of armaments or of military spending of two sides have interacted. There is no convincing approach to this problem available so far that could not be criticized under the criteria presented in this contribution.

This seems to allow the conclusion that it might not be reasonable to construct a sequence of almost insurmountable obstacles, to set the solution of the questions of the quantitative disentanglement of various influences on armaments and of the one general and empirically valid arms race model before any serious empirical research on the relationship between armaments and wars. To make this point more simple: Instead of talking about the arms race-war connection, one could talk about the armaments-war connection. This is an extremely important distinction as all of the connotations of "arms races" have vanished from such a research program. Such a program could be less pretentious, but more realistic, if one is not so much interested in formal models but in the development of an early warning system for armed conflict.

For this reason, the simple approach of Wallace can claim considerable sympathy, even though undoubtedly there are shortcomings. Identifying arms races via serious disputes is as questionable as the criteria he adduces for ascertaining mutual stimulation of armament levels. An alternative to his approach would be further to develop the notion of "normal" military effort or of "normal" expansion of particular indicators of military strength. Maybe there are patterns of the development of military machineries that characterize situations of relatively low conflict, low mutual threat, and low probability of an escalation to war, e.g. constant ratio of armed forces to population, constant or decreasing share of GNP devoted to the military, roughly constant nominal growth of arms expenditures, stagnation

of military expenditures in real terms, etc. (see Rattinger 1974). Maybe deviations from such "normal" patterns are associated with military conflict later on. This research program might even be easier to tackle than that of Wallace, because individual nations - instead of dyads - are the unit of analysis. In fact, this is a very simple research program: Is the probability of a nation getting involved in military conflict further down the road systematically associated with previous deviations of this nation from its regular and "normal" style of military preparation?

This does not exclude the possibility to identify dyads of nations for whom this type of behavior can be observed in a parallel fashion and that later on have either fought each other or resolved their dispute peacefully. The most important point is that this approach would not require to postpone meaningful results from such a miniature version of an early warning system until final solutions of the discussions about the adequacy of particular arms race models and of particular estimation procedures have been established. Maybe the argument can be convincingly made that this is the direction into which - considering the present state of the empirical validation of arms race models in the Richardsonian tradition - the study of the relationship between armaments and wars could be most profitably developed.

CONCLUSION

This contribution has not presented a summary of empirical results on arms race models but has dwelt on the problems of modelling, of concept formation and of operationalization, as this appeared as more important and meaningful considering the present state of the discipline. What general conclusions do our discussions allow?

In the Richardsonian tradition, theorizing and empirical study have moved apart to a certain extent. There are numerous examples of the purely immanent refinement of Richardson's classical models. Empirical research has followed somewhat different paths, sacrificing much of the analytic elegance of the classical models. The difficulties of immanent arms race theorizing appear as minimal and - at least to this present author - as less important compared to those of em-

pirical research that has developed into the direction of more comprehensive arms building models, thus integrating arms race study with other behavioral approaches. The isolation of singular explanatory factors, that Richardson did pursue by his exclusive focus on armaments, has proved to be untenable for empirical analysis. It now addresses the really crucial issues, and in doing so has discovered the really crucial difficulties: What factors determine armament behavior, how are those many various influencing factors to be separated empirically, what are appropriate measurements, what is the quality of our sources of data, and how do reactive military processes work in reality? For empirical research, Richardson still supplies numerous valuable ideas what factors could be related to each other in what form, stimulating the specification of innovative models with empirical content. But his classical analyses of the formal properties of arms race models are less relevant for current empirical research. It appears as much more important to ascertain why particular models sometimes fit very well, sometimes not at all, why model parameters change, etc.

Opening arms race models to include exogenous variables and giving up the purely mechanistic perspective has integrated arms race studies into general research on international relations, international conflict, and foreign policy decisions. If one is interested in formal models of international politics not for their own sake but in order to understand why and how particular forces drive processes within the real world, then this shift of focus is to be much welcomed.

REFERENCES

- Altfeld, Michael F., "Arms Races? - and Escalation?", International Studies Quarterly 27 (1983), 225-231.
- Busch, Peter A., "Appendix: Mathematical Models of Arms Races" in: Bruce M. Russett, *What Price Vigilance?*, New Haven, 1970.
- Cioffi-Revilla, Claudio A., *Mathematical Models in International Relations: A Bibliography* (Institute for Research in Social Science, IRSS Technical Paper No. 4), Chapel Hill, 1979.
- Cusack, Thomas R. and Michael D. Ward, "Military Spending in the United States, Soviet Union, and the People's Republic of China", Journal of Conflict Resolution 25 (1981), 429-469.

- Goldmann, Kjell, *Tension and Détente in Bipolar Europe*, Stockholm, 1974.
- Intriligator, Michael D., "Formal Models of Arms Races", Journal of Peace Science 2 (1976), 77-88.
- Intriligator, Michael D., "Research on Conflict Theory", Journal of Conflict Resolution 26 (1982), 307-328.
- Luterbacher, Urs, "Arms Race Models: Where Do We Stand?", European Journal of Political Research 3 (1975), 199-217.
- Mihalka, Michael, *Arms Accumulation in the Middle East* (Center for International Studies Paper, Massachusetts Institute of Technology), Cambridge, Mass., 1975.
- Moll, Kendall D. and Gregory M. Luebbert, "Arms Race and Military Expenditure Models", Journal of Conflict Resolution 24 (1980), 153-185.
- Nincic, Miroslav, *The Arms Race*, New York, 1982.
- Ostrom, Charles W., "Evaluating Alternative Foreign Policy Decision-Making Models", Journal of Conflict Resolution 21 (1977), 235-266.
- Rapoport, Anatol, "Lewis F. Richardson's Mathematical Theory of War", Journal of Conflict Resolution 1 (1957), 249-299.
- Rattinger, Hans, "Gleichgewicht von Rüstungswettläufen und seine Stabilitätseigenschaften als Erklärungsvariablen für den Ausgang von Rüstungswettläufen", Politische Vierteljahresschrift 15 (1974), 471-523.
- Rattinger, Hans, *Rüstungsdynamik im Internationalen System*, München, 1975a.
- Rattinger, Hans, "Armaments, Détente, and Bureaucracy", Journal of Conflict Resolution 19 (1975b), 571-595.
- Rattinger, Hans, "Rüstung in Europa: Aufrüstung, Wett-rüsten und andere Erklärungen", Österreichische Zeitschrift für Politikwissenschaft 4 (1975c), 231-250.
- Rattinger, Hans, "From War to War to War: Arms Races in the Middle East", International Studies Quarterly 20 (1976a), 501-531.
- Rattinger, Hans, "Econometrics and Arms Races", European Journal of Political Research 4 (1976b), 421-439.
- Richardson, Lewis F., *Arms and Insecurity*, Chicago, 1960a (initially published as *Generalized Foreign Politics*, London, 1939).
- Richardson, Lewis F., *Statistics of Deadly Quarrels*, Chicago, 1960b.
- Russett, Bruce M., "International Interactions and Processes", in: Ada Finifter, ed., *Political Science: The State of the Discipline*, Washington, 1983, 541-568.

- Schrodt, Philip A., *The Effects of a Richardson Arms Race on the Distribution of Power in an International System* (Ph.D.Dissertation, Indiana University), 1977.
- Singer, J. David, "The Outcome of Arms Races", *Proceedings of the International Peace Research Association, Third General Conference, vol. II: The International System, Assen, 1970*, 135-146.
- Singer, J. David and Melvin Small, *The Wages of War, 1816-1965*, New York, 1972.
- Smith, Theresa C., *The Arms Race/War Connection* (Ph. D. Dissertation, University of Minnesota), 1978.
- Wallace, Michael D., "Arms Races and Escalation", *Journal of Conflict Resolution* 23 (1979), 3-16.
- Wallace, Michael D., "Armaments and Escalation", *International Studies Quarterly* 26 (1982), 37-56.
- Wallace, Michael D., "Arms Races and Escalation", *International Studies Quarterly* 27 (1983), 233-235.
- Ward, Michael D., *The Differential Paths to Parity* (Paper, Science Center Berlin), 1982.
- Weede, Erich, "Arms Races and Escalation", *Journal of Conflict Resolution* 24 (1980), 285-287.
- Wright, Quincy, *A Study of War*, Chicago, ²1963.
- Zinnes, Dina A., *Contemporary Research in International Relations*, New York, 1976.

DISCUSSION

QUESTION: Does the research by Smith confirm Richardson's predictions on the arms race-war connection?

ANSWER: There is no clear-cut confirmation of Richardson's assertion that arms races described by formal models without a stable equilibrium point necessarily should escalate into war at some time. In Smith's study the proportions of arms races without stable equilibrium and with a stable equilibrium that were followed by war were roughly equal. One has to admit, however, that the simplest version of Richardson models was used in this study.

QUESTION: What predictive capabilities are associated with the shift in focus of empirical arms race research?

ANSWER: There has to be a distinction between prediction of war from previous armament behavior and prediction of arms levels. Richardson himself, who was less interested in systematically analyzing all the various factors influencing national armament policies, predominantly focussed on the prediction of war from the

formal properties of the models describing competitive armament processes. One can argue that this approach, that has been closely imitated by Smith, at present does not appear as very promising because of the many problems associated with the empirical validation of Richardson-type models that have been described. Therefore, one could resort to the simpler attempt to find criteria for dynamic equilibrium or "normal" patterns of armament behavior. This could eventually be turned into an early warning system that would attempt to predict the onset of war from "abnormal" military efforts of one nation or of pairs of nations. The second predictive goal of comprehensive armament building models, i.e. prediction of armament levels, can be achieved only to a limited extent. As soon as exogenous variables are included, which in many cases could prove mandatory, predictions either have to be very short-range or they are valid only under comprehensive assumptions of ceteris paribus.

QUESTION: As to the relevance of indicators of armament levels for decision-processes, should one distinguish between indicators that actually are used as the basis for decision and others that are employed to publicly support or defend such decisions?

ANSWER: Even though the details of actual decision-processes on military budgets and arms programs generally are not publicly accessible, there is ample reason to suppose that, at least at the very high political level, the arguments and criteria used in actual decision-making are not very much different from those that prevail in public political discussions. These top-level political decisions are what finally matters. That the sizing and costing of military forces is derived from objective threat assessment via an analytic process given security requirements and own and enemy capabilities is a myth. It is very likely that in politics the simple question of "how much is enough?" is actually treated as a simple question involving crude numbers of weapon systems and billions spent, even though for the military operations researcher it might be an optimization problem involving several hundreds of variables.

QUESTION: Is there a lack of communication between theorists and empiricists in arms race research?

ANSWER: What David Singer once termed the "two culture problem" pervades arms race research as much as most other substantively delimited fields of social science. This is not exclusively due to ill will, incompetence,

or ignorance, but also to different backgrounds of training and professional experience. This state is not quite satisfactory, however. To be a good empirical researcher one has to listen what other people have to say on what factors could or should be related in what way. To be a good theorist one should occasionally listen to what can be said about the empirical validity of one's deductions and assumptions. Maybe this is a purely personal perspective, but it appears that empirical researchers so far have listened closer and have received more stimulation from analytical modelling than vice versa.

QUESTION: Is the distinction between pure arms race modelling and applied arms race research comparable to the distinction between normative and descriptive models in decision theory?

ANSWER: This seems to be a very close analogy. Both approaches, when pursued in isolation, appear as scientifically incomplete. In any discipline that refers to the real world, i.e. every discipline except mathematics, the goal should be to draw on theoretical insight as much as possible to specify models with empirical content, then to attempt to validate them empirically, subsequently refining theory, etc.

QUESTION: Are there compilations of data available that are appropriate for the empirical validation of arms race models?

ANSWER: General sources for data on national armaments, particularly defense budgets and numbers of weapons systems, are widely known. When it comes to refining those highly aggregated and rather primitive indicators, however, e.g. by deriving indices of combat power and effectiveness from technical characteristics of weapon systems or by estimating military trade-offs between different types of forces, collection of data can be quite difficult and burdensome, sometimes virtually impossible.

QUESTION: Can the assertion by the peace movement that armaments lead to war be claimed as empirically validated?

ANSWER: In line what has been said on the lack of stability of causal structures, any answer probably will have to be differentiated for different historical periods, if not for different regions of this world. According to Wallace, the argument brought forth by the peace movement was essentially correct up to the Second

World War. Since then, deterrence seems not to have been confined to the nuclear confrontation between the super-powers.

QUESTION: Is there any spectacular example in history where the predictions of the Richardson models on the onset of war have come true?

ANSWER: Richardson has provided this spectacular example himself by investigating the arms race prior to the First World War. His methods were not very sophisticated, however, and his application of his basic equations was rather arbitrary, e.g. by aggregating the nations of the two alliances that later were to fight each other, but excluding some of these nations without convincing arguments for doing so. Moreover, in his great book Richardson empirically tested just one of at least two formal deductions on what the data should look like if the armament process were to correspond to his equations. The second deduction, which he did not test, when applied to his own data in an analogous fashion in fact flatly disconfirms the model's fit. Still, Richardson maintained that his equations were fitting the data beautifully, that there had been no stable equilibrium, and that this had something to do with the onset of war later on. Most of the reviewers of his work showed themselves to be very much impressed.