

## Arms Races and Escalation: A Closer Look\*

PAUL F. DIEHL

Correlates of War Project, University of Michigan

The relationship between arms races and war is a critical consideration in both peace research and strategic planning. This study reconsiders the work of Michael Wallace which has postulated that arms races significantly increase the probability of a serious dispute escalating to war. A critique of Wallace's coding procedures and arms race index precedes an attempt to replicate his findings. In the replication, serious disputes, taken from the Correlates of War Project, among major powers during the years 1816-1970 serve as the population to be tested. Adjustments in coding and index construction from the Wallace work are made. It was discovered that only 25% of those disputes preceded by a mutual military buildup escalated to war, while almost 77% of the wars in this population were preceded by periods lacking armaments competition. Controls for inter-century differences and unilateral military buildups failed to alter this apparent lack of a relationship between arms races and dispute escalation. Differences with Wallace's study are analyzed and the implications for peace research discussed.

Conventional wisdom has always presupposed a link between rapid military buildups and war. The old dictum 'if you want peace, prepare for war' offers one perspective on the inter-relationship of military spending and the outbreak of conflict. The spiral model is indicative of a more dangerous connection between increasing weapons and war. Whether the effect is deterrence or provocation, a nation's decision to significantly increase its military capability could be an important factor in the understanding of interstate war.

Despite the central nature of military spending in national security decision-making, empirical researchers have generally ignored its possible effect on the initiation of war. This void in the academic literature noted by Singer (1979) in 1969 remains large today.

Nevertheless, some recent efforts have ex-

tended the pioneering ideas of Richardson (1960), studying the impact of arms races and military spending decisions on the outbreak of war. The most interesting work in this area has been that of Michael Wallace. He used early Correlates of War (COW) Project compilations on major power military expenditures and serious disputes to investigate nation behavior in conflict generated situations. In a widely quoted article, Wallace (1979) concluded that the presence or absence of an arms race between two rivals correctly predicted war/no war outcomes in over 90% of the serious disputes studied. Those results are summarized in Table I.

A later study by the same author (Wallace 1982), using the same data base, served to reinforce this strong association between arms races and war. The general paucity of alternative investigations makes Wallace's studies the most definitive to date.

If Wallace's findings are correct, the implications for policymaking on arms limitation are clear. The START negotiations must proceed with all deliberate speed, lest a clash between the superpowers should escalate to all-out war. However, certain methodological problems cast doubt on the validity of Wallace's conclusions. It is the purpose of this paper to detail these difficulties and retest

\* An earlier version of this paper was presented at the Annual Meeting of the International Studies Association-South, Atlanta, 4-6 November 1982.

The author would like to thank Michael Champion, Miroslav Nincic, Peter Wallenstein, J. David Singer, Bradley Martin and Bruce Russett for their comments and suggestions. In addition, gratitude is expressed to Mary Macknick and Louis Erste for technical assistance in this project.



Table 1  
Wallace's 'arms races and escalation'

	War	No war
Arms race	23	5
No arms race	3	68

$\chi^2 = 62.99$        $\emptyset = .8$        $Q = .98$

the military buildup-war relationship with a modified set of assumptions and indicators.

#### *A closer look*

The *Correlates of War* treats multi-party serious disputes (three or more disputants) as one integrated dispute. Where there was a clear informal/formal partnership among the disputants on one side and where the subject matter of the dispute was the same for each partner, the dispute classification scheme served to reflect the interconnection of events and interests for all those involved. Wallace chose to code each dispute participant dyadically against those nations in opposition. Thus, the original 1914 dispute which escalated to World War I is analyzed as if it were six separate disputes and consequently six individual wars. Overall, 26 distinct wars are created where only 7 or 8 integrated ones occurred.<sup>1</sup> As a result, the strength of the arms race-war relationship stems not from an abundance of distinct cases of dispute escalation, but merely is a function of a coding decision. Wallace's (1980) response to this problem was to re-evaluate his results using only *formal* alliance patterns to combine certain sets of disputes. This only partly solved the difficulty of numerical inflation of disputes and wars. The follow-up analysis failed to consider situations which share similar characteristics to those involving formal alliances, but merely lack a signed instrument between the parties.

A related problem is Wallace's inclusion in his population of serious disputes those cases which were not independent of ongoing wars. In some ways, this explains the fact that the two World Wars account for over 80% of the explanatory capability in his study (Weede 1980). Serious problems are inherent

in studying war-related disputes in this context. Wallace's purpose was to assess whether an arms race affected the probability of a serious dispute escalating to war. However, in cases where one or both disputants are involved in a war, the probability of that war-related dispute escalating is greater than that of a dispute independent of an ongoing war. An inference about the effect of an arms race on a war-related dispute must be considered tentative at best. In Wallace's work, the arms race impact (if any) on dispute escalation is indiscernible from the effects of the ongoing war. Nevertheless, it is these cases of war-related disputes which enhance the confirmation of the hypothesis that arms races lead to war.

Wallace's results indicate the U.S.S.R.-Japan dispute of 1945 to be an instance of an arms race leading to the escalation of a dispute. Yet, it is difficult to believe that this brief war was anything but a result of the hostilities associated with World War II. To suggest that an arms race in the late 1930s exercised any influence on the outbreak of war in this dispute five years later is premature without additional research and runs contrary to accumulated historical opinion on the subject.<sup>2</sup>

Wallace (1979:8) justifies his choice of cases and the dyadic coding method by stating:

'In the case of those wars which involved more than two powers, each dyad is coded separately. Thus, for example, World War II is coded as an initiation of Franco-German and Anglo-German hostilities in 1939, an Anglo-Italian and Franco-Italian outbreak in 1940 and a Russo-German and Japanese-American conflict initiated in 1941. This was done to avoid the practical and conceptual difficulties of aggregating military capabilities of nations entering the conflict at different times.'

This explanation is open to criticism as the disputes involving different actors in different years are coded as separate disputes anyway according to COW criteria. If the disputants entered the dispute in the same year (thereby determining the same data points for military expenditures — the ten years prior to that year in the Wallace study), it is irrelevant if the disputants entered the war emanating from the dispute at different times. Furthermore, Wallace's explanation provides no justification for separating a dispute (such as the 1939 Poland crisis) in which all parties entered the dispute and the war at the same time.

Wallace's method of determining an arms race is not without conceptual problems. An arms race is conventionally described as a process involving competitive interaction, manifested by rapid increases in military spending and weaponry. In Wallace's analysis, it is only determined whether or not the disputants are rapidly arming themselves; there is no determination if this spending is directed against the dispute opponent(s). More properly, the process described by Wallace is a 'mutual military buildup' rather than an 'arms race' per se. A determination of the latter must await the completion of a more sensitive test of military spending decisions.

Beyond the absence of definitional rigor, the mutuality of rapid spending increases cannot necessarily be deduced from Wallace's arms race index. By multiplying the products of each side's cardinal spline estimate of military spending, a unilateral buildup by one side might be defined as an arms race. For example, if country A had a score of 100 (high) and country B had a score of 1 (low), the net index would be 100 and the situation classified as a Wallace arms race. Clearly, this instance is neither a mutual military buildup nor an arms race in any reasonable definition of the two terms.

Wallace also chooses a seemingly arbitrary threshold of a 90.00 index score to distinguish between arms races and their absence. No justification is presented and experimentation with alternative thresholds is not evident. Wallace's conclusions are substantially

weakened if the threshold is lowered to 50.00. Then, ten additional cases would be contrary to the escalation model, while its strength would not be enhanced by even a single case.

While Wallace's striking findings have dominated this topic area over the past few years, it seems that criticisms associated with his work are sufficient to warrant a re-examination of the relationship between rapid military buildups and war.

#### *A re-evaluation*

In retesting Wallace's conclusions, a number of modifications are made. The temporal domain is extended five years, now 1816-1970, to reflect the most complete listing of COW serious disputes. Since Wallace used only a preliminary draft of this same compilation, the dispute populations are not identical across studies. The addition of newly discovered disputes and the correction of coding errors would suggest that the set of cases used by this replication attempt is more accurate than that employed by Wallace.

Each serious dispute is treated as an integrated whole (non-dyadically) as is consistent with the original COW coding scheme. For example, the dispute leading to the outbreak of the Crimean War is coded: Russia v. Great Britain and France. Where it was clear that partnerships existed and the partners were inexorably tied up in the disposition of the dispute, it seems reasonable to consider allies together rather than separately in the dyadic scheme. Surely, the military calculus of the opposing side must consider the joint actions of those partners. Moreover, good scientific practice requires that findings should be reanalyzed under the most conservative, yet reasonable, set of cases. If Wallace's results are conclusive, they should be able to be reproduced under the conditions outlined here.

Certain factors dictated that some of the cases be eliminated from the study. Any dispute which was related to an ongoing major power war was dropped from the population.<sup>3</sup> As discussed above in the critique of Wallace, the escalatory effects of the ongoing

war are indistinguishable from those of arms increases. To include those cases might yield a false indication of the real effects military buildups have on the initiation of war.

It is all but impossible to accurately estimate war-time military expenditures for a nation whose whole economy is devoted to the war effort. Accordingly, COW treats expenditures during the two World Wars as missing data. Due to this data limitation and the nature of the arms race indicator, cases independent of ongoing wars from 1915-1920 and 1940-1947 are necessarily eliminated.<sup>4</sup> Idiosyncrasies in the data set resulted in the elimination of another case.<sup>5</sup> Overall, the analysis here considers 86 separate disputes.<sup>6</sup>

A measure of mutual arms buildup must reflect significant military increases for both sides in a dispute and yet be able to detect instances where only a unilateral buildup is present. In addition, an appropriate threshold point must be chosen to differentiate between incremental spending patterns and those which are abnormally high. These two considerations were judged to be lacking in the original Wallace article.

With this in mind, an index of military growth for each side in a dispute will be constructed from the newly revised COW file on military expenditures (an earlier version was used by Wallace in his work). This index is the mean rate of change in military expenditures (expressed in common currency and controlled for price fluctuations) for each side in the three years prior to the initiation of the dispute.<sup>7</sup>

Only expenditures prior to the initiation of the dispute are analyzed. This precludes consideration of military spending which was reactive to the dispute itself. In this way, the index is able to isolate the effects of an arms buildup from the bias of dispute-

induced spending. Dispute spending patterns tend to reflect significant spending increases whether war results or not. This is not surprising as nations seek to ensure security in a crisis situation.

Furthermore, the index measures only spending trends in the immediate past of the dispute. Military expenditures tend to show greater variation as one moves farther back in a time-series. Too often, studying a time period of five or ten years will cause an overlap into a war period or time frame in which other disputes influenced expenditure patterns. Focusing on the three years prior to the dispute allows consideration of behavior which is more perceptually important than comparable spending decisions ten years before. Rapid changes in military expenditures are warning signals not only for peace researchers, but are perhaps one of the indicators used by foreign policy elites to ascertain a large scale military buildup by an opponent.

In this study, I have chosen to designate any instance of both dispute sides increasing their military expenditures at a rate of 8% or greater for the three years before the dispute as a 'mutual military buildup'. This threshold was chosen because it most perfectly captured the dividing line between incremental increases in military outlays and abnormal spending increases. This coding decision allows the World War I dispute to be classified as a mutual military buildup as is the concerted opinion of most historians. In no case, however, are both sides in such a buildup below a 10% threshold often cited as an indicator of high level military spending increases.<sup>8</sup>

The operational definitions of a 'major power', 'serious dispute', and 'war' developed by Singer & Small (1972) and used by Wallace are retained in this study.

Table II  
Mutual military buildup and escalation

	War	No war
Mutual military buildup	3	9
No mutual military buildup	10	64

$\chi^2 = 1.06$

$\emptyset = .11$

$Q = .36$

### Results

The determination of a mutual military buildup or its absence for each dispute was combined with the outcome of that dispute and the aggregated results are presented in Table II.

No meaningful covariation exists here between mutual military buildups and dispute escalation. Only 25% of the disputes which were preceded by a mutual military buildup escalated to war, while almost 77% of the major power wars constituting this sample population were preceded by periods in which there was no incidence of joint and rapid spending increases by the protagonists.

Of the three disputes which fit the escalation hypothesis, one is World War I and the other two led to the Second World War. However, even these cases raise questions about the impact of mutual arms buildup on the outbreak of war. There were five other disputes prior to World War II which were preceded by this type of buildup and yet did not escalate to war. This suggests that the cases which support the escalation model might only be the product of a spurious association.

Overall, it appears that most serious disputes do not involve previous dual military spending increases and most serious disputes do not escalate to war; but there does not seem to be any connection between these facts. The Yule's Q value is .36 indicating a much weaker positive relationship than Wallace reported. However, the more conservative  $\emptyset$  coefficient is only .11 and the Chi-square value is not significant at any meaningful level.

It is possible that this analysis, aggregated over a two-century period, may hide a relationship that is present in only a portion of

this time period. Other scholars have noted inter-century differences in studying international conflict. Therefore, the results were disaggregated, divided into 19th and 20th groups and the hypothesis retested. The findings were quite similar to the original results. Although the association was stronger in the 20th century, neither relationship was significant at the .10 level.<sup>9</sup> It is also important to consider the effect of an unilateral arms buildup on the outbreak of war. It may be that the absence of mutuality in arms increases causes one side or the other to consider war a more viable means of competition. In testing this possibility, cases in which there was no mutual military buildup are considered (N=74). The proposition that a unilateral buildup (constituting a 10% or more increase in military expenditures by one and only one side using the same index construction as before) affects the probability of a dispute escalating to war is considered in Table III.

There seems to be no basis for concluding that a unilateral military buildup prior to a dispute increases the chances of war. A Yule's Q value of  $-.35$  suggests a possible negative association between unilateral buildups and war but the Chi-square value indicates that the association is not statistically significant.

### *Towards a convergence of findings*

The findings presented here are quite contrary to Wallace (1979). This lends itself to a number of possible explanations. Immediately the differences in arms indices come to mind. However, Wallace's cardinal spline function is heavily weighted toward changes in military expenditures in the three or four years prior to the dispute, much as the index used in this study. In applying the Wallace measure

Table III  
Unilateral buildup and escalation\*

	War	No war
Unilateral Buildup	3	30
No unilateral buildup	7	34

$r^2 = .997$

$\emptyset = .12$

Q =  $-.35$

\* Table includes only those cases which fail to meet the criteria for a mutual military buildup.

to the data set used in this study, similar findings can be reported where coding rules between the studies were not in conflict. Where differences did exist, the variation can be explained by reference to other factors beside index construction. Thus, I conclude that the differences in findings are not attributable to differences in the military spending indices.

Another hypothesis is that differences between the two studies' data sets led to divergent findings. Wallace's list of disputes was only in its early stages of completion when his article first appeared. The population of serious disputes used in this replication attempt represents a more recent and complete version of that list. In comparing the two versions (prior to any coding decisions), the disparities do not seem to be extraordinary, at least not to suggest radically different conclusions. The newer data set includes a few more cases of pre-World War II disputes which were preceded by dual arms increases but did not escalate to war. The Korean War, actually preceded by spending cutbacks, is another instance of a dispute not covered by Wallace. Most of the other changes in the new file are additions or deletions of 'no buildup-no war' disputes. As a whole, the empirical validity of the escalation hypothesis is weakened when tested with the updated file, but the changes alone are insufficient to reject Wallace's conclusions.

Beyond simple changes in the number of cases, the new file contains some corrections. The Russo-Japanese War of 1904 had its dispute beginning in 1903 according to the latest file. This is an update from the Wallace report that the dispute began in 1904. Wallace considers spending increases in this dispute through the actual first year of the dispute (1903). It is not surprising then that he concludes that an arms race took place before the war. However, according to the corrected files, one might infer that spending in 1903 was reactive to the dispute and that the Wallace index would yield a false indication of prior military competition. Looking only at military spending patterns prior to the beginning of the dispute (pre-1903), this study

finds no significant joint arms increases occurring.

Differences in military expenditure figures might account for opposite conclusions in a few cases. Wallace replaced some interpolated data points with his own estimates of military appropriations. This may explain why different results are obtained in the 1866 dispute/war between Germany, Italy and Austria.<sup>10</sup>

While revision of research files is to some degree a continuing process, it is presumed (pending comparison) that the data used in this study are more complete and accurate than previous compilations.

Most of the remaining conflict in the aggregate findings of the two studies can be traced to differences in coding procedures. Ten cases which were not independent of ongoing wars, yet exhibited covariation of spending increases and escalation, were eliminated in this study. In addition, the non-dyadic coding method used here resulted in the collapse of ten cases, which fit the escalation hypothesis, into three integrated disputes. In each case, the two World Wars account for almost all the instances. In effect, the strength of the arms race-war relationship cited by Wallace rests heavily on the two World Wars. The relationship seems absent in any other circumstance and gains statistical significance only through an artificial division of an integrated situation.

### *Conclusions*

This study retested Wallace's (1979) findings that a mutual military buildup between major powers increased the probability of a serious dispute escalating to war. Using a modified set of assumptions and indicators, it was discovered that only one-fourth of the disputes preceded by mutual military buildups resulted in war, while ten of thirteen wars occurred in the absence of joint arms increases by the dispute participants. Therefore, it was concluded that mutual military buildups did not exercise any general impact on the initiation of war under the limited conditions studied. This lack of a relationship between military spending and dispute escalation remained

unchanged when controls were instituted for inter-century differences and when retested to ascertain the influence of a unilateral military buildup.

In considering the differences in findings between this study and Wallace's work, the importance of assumptions, coding decisions and data manipulation techniques in empirical research should be highlighted. Apparently insignificant research choices can collectively influence results in a profound way. We owe it to our colleagues and those in policymaking circles to be explicit and reveal all relevant information that impinges on the conduct of the study. This is not to imply that academics should retreat to merely debating semantics or methodological approaches. Rather, it means we must give greater attention to the research design of a study and consider the study and its utility in light of the validity of that design. The operative message is that creativity and rigor must coexist in research. Neither is a substitute for the other.

This paper does not in any way lay to rest the debate over the danger of rapid military spending increases. What is apparent, however, is that they do not constitute an explanation by themselves for the escalation of disputes. Future research should expand the scope of past studies to consider arms buildups in their contexts of the national attributes of the participants as well as the systemic conditions prevailing at that point in time. There is also a need to study outcomes of military buildups beyond those which end in war. Incidences of compromise and capitulation resulting from arms acquisition may be just as significant as war outcomes. The relative mix of mutual and self-stimulatory processes driving each nation's spending increases may be an important key in predicting those outcomes.

A careful examination of the relationship of the arms races and war might yet provide guidance to foreign policy elites, such that at a minimum, those decision makers can avoid mistakes which could have unintended but disastrous consequences.

## NOTES

1. The number of separate wars in that study is variable, depending on how the analyst would treat the interrelationship of the European and Pacific theatres in World War II.
2. In cases which involved war-time disputes, Wallace studied the disputant's military expenditures in the nearest pre-war year instead of the year before the dispute.
3. The cases dropped are mainly those which occurred after the outbreak of World War I and World War II. Other cases eliminated were those associated with the Crimean War, and the 1866 war involving Germany, Italy and Austria. The Russo-Japanese conflicts in the 1930s are included as both temporally preceded the Polish crisis and were distinct from tensions in Europe and elsewhere in Asia.
4. The most notable absence necessitated by data unavailability is the U.S.-Japan dispute of 1941. Unlike other disputes surrounding the World War II, this dispute could justifiably be considered separate from the European War. There was also missing data for one side in two disputes: U.S.S.R.-Great Britain in 1923 and the dispute over Korea in 1950. However, calculation of a military buildup index for the side for which data was available indicated an absence of a mutual military buildup according to the criteria used in this study. Consequently, these cases remain included in the analysis. In addition, a 1948 dispute involving the Soviet Union and the three Allied Powers was not excluded despite the inclusion of only two (instead of three) years of military expenditure data in the military buildup indicator. Missing data for the war year 1945 were responsible for this alteration.
5. The dispute between France and Italy in 1860 is not considered. Prior to the dispute, Italy's military expenditures were coded as those only of Sardinia. With the advent of Italian unification, expenditure figures were derived from all of Italy. Consequently, an artificial jump in military spending occurs in the data set. This changeover occurs in one of the years under study for this case and to avoid misconceptions, this dispute is eliminated from the sample.
6. Some wars which resulted from a major power serious dispute, but did not involve major power participation on both sides are included in this study. These are: the 1863 war between Denmark on one side and Germany and Austria on the other; Italy's invasion of Ethiopia in 1934; and the Sino-Indian border war of 1962.
7. The formula for the index is:  

$$\% \Delta (M_{t-2} \rightarrow M_{t-1}) + \% \Delta (M_{t-3} \rightarrow M_{t-2})/2$$
 where  $t$  is the year of the dispute and  $M = m_1 + m_2 \dots m_k$ ;

where  $m_i$  represents the military expenditures of a major power on that side of the dispute.

8. Experimentation with other thresholds did not significantly affect the results reported here.
9. The  $\emptyset$  value for the 19th Century war .08 and Chi-square was a paltry .198. The 20th Century yielded a  $\emptyset$  of .18 and a Chi-square value of 1.79.
10. These manipulations are not apparent in any of the Wallace articles. The disclosure of these transformations was graciously made to me in a personal conversation with Professor Wallace. However, analysis of the extent and validity of these changes cannot be assessed without a copy of the Wallace data base. Professor Wallace is attempting to reconstruct that file and results as of this writing are incomplete.

## REFERENCES

- Huntington, S. 1958. 'Arms Races: Prerequisites and Results.' *Public Policy*, 18:41-46.
- Lambelet, J. 1975. 'Do Arms Races Lead to War?' *Journal of Peace Research*, 12, 2:123-28.
- Richardson, L. 1960. *Arms and Insecurity*. Pittsburgh: Boxwood.
- Singer, J.D. 1979. 'The Outcome of Arms Races: A Policy Problem and a Research Approach'. In *The Correlates of War: I*, ed. by J. David Singer. New York: Free Press, 145-54.
- Singer, J. David and Melvin Small 1972 *The Wages of War 1816-1965: A Statistical Handbook*. New York: John Wiley and Co.
- Wallace, M. 1979. 'Arms Races and Escalation: Some New Evidence.' *Journal of Conflict Resolution*, 23,1:3-16.
- Wallace, M. 1980. 'Some Persisting Findings,' *Journal of Conflict Resolution*, 24,2:289-92.
- Wallace, M. 1982. 'Armaments and Escalation.' *International Studies Quarterly*, 26,1:37-56.
- Weede, E. 1980. 'Arms Races and Escalation: Some Persisting Doubts.' *Journal of Conflict Resolution*, 24,2:285-88.