

© Journal of Peace Research, ISSN: 0022-3433 vol. 34, no. 1, 1997, pp. 7–22

# Arms Races and Dispute Escalation: Resolving the Debate*

SUSAN G. SAMPLE
Peace Research Centre, Australian National University

The article seeks to resolve the debate in the empirical literature over the effect of arms races on dispute escalation. Until now, tests of this link have remained inconclusive due to the conduct of largely parallel, rather than integrative, research. Michael Wallace's 1979 study finding a strong link between arming and escalation suffered substantial criticism on grounds that both his index and his dispute sample were flawed. Utilizing the arms races measures offered by Diehl and Horn, this article delves into the conceptual, measurement, and sampling controversies that have made resolution before this impossible. The results of resetting the proposition indicate that regardless of the arms race measure or dispute set used, the relationship between arming and the escalation of disputes to war is positive and significant. Removal of controversial dyads from the World Wars does not alter the findings substantially (or in the assumed direction). By introducing a time lag, it is also shown that many of the disputes that both Wallace and Diehl designated as non-escalating arms races were between countries that were at war with each other within five years. Consequently, the conclusion is that arms races are strongly associated with the escalation of disputes to war.

## 1. *The Arms Race Debate*

The vast majority of militarized interstate disputes (MIDs) do not escalate to war. For whatever reasons, after a threat is issued, or violence short of war is used, the states involved do not continue to 'raise the ante', but rather find some other means of resolving the conflict. Of the major state dispute dyads employed in this study, approximately 12% escalate to war. In order to understand the dynamics that bring on war, it is important for us to consider how crises that escalate and crises that do not escalate differ. Are there patterns in these disputes that might let us discern when some disputes between states have a higher probability of escalating to war? If such patterns are discernible, we have more hope of elaborating politics that will prevent the disputes, or intervening before the disputes escalate, and so move a step closer to peace.

Do arms races increase the probability that two disputing states will escalate to war? This question represents one of the most significant controversies in the study of war and peace. A vast theoretical and practical gulf lies between those who argue that war can often be averted by much preparation and those who argue that preparing for war makes it virtually inevitable. Arms races and war are theoretically linked by the security dilemma. For students and political leaders alike, one of the primary lessons drawn from World War I was that war could be the inevitable outcome of arms races that were motivated, innocently enough, by the desire for self-preservation. It is argued that by feeding the security dilemma, arms races generate and reflect enmity, creating a hostile spiral in which violent confrontation begins to be seen as inevitable. Once war is considered inevitable, it is only a matter of time before a small fire will turn into a massive conflagration.

In order to determine if this argument is empirically valid, I have directed my study toward a critical portion of the quantitative literature that seeks a link between the existence of an ongoing arms race between states and the escalation of militarized disputes between them to war. Employing the Correlates of War data on militarized interstate disputes (MIDs) and military expenditures, a debate has been waged for a number of years to answer the question of whether or not a positive correlation exists between arms races and disputes escalation. This simple question is vital to our understanding of conflict because if a positive correlation be-

* The author thanks J. David Singer, Stuart Bremer, and members of the Correlates of War Project for access to the data used in this study. I am grateful to Dan Smith, John Vazquez, Nils Petter Gleditsch, Paul Diehl, Michael Wallace, and the anonymous reviewers for their helpful comments and suggestions on drafts of this paper. Naturally, the responsibility for the content is my own.

tween arms racing and dispute escalation is found, then we will know that in militarized disputes a dynamic is a work other than calculations based simply on the absolute level of the opponent's arms. Until now, attempts to find a direct link between arms races and war have met with mixed success, suggesting that conceptual, measurement, and sampling differences in quantitative studies have affected the results of the studies, and that the arms policies of states cannot legitimately be considered separate from the political relationships between the states.

The studies I am looking at have failed to resolve the issue because they have failed to focus the strong theoretical light on their critical differences that would lead to accumulation of knowledge, and have instead engaged in largely parallel research. By examining the areas of disagreement in the research thus far, and by considering the implications of introducing a temporal element into the dyadic relationship, I intend to resolve the debate over the effects of arms races in dispute escalation.

## 2. *Empirical Tests of the Arms Race–Dispute Escalation Connection*

In 1979, Michael Wallace published a path-breaking study in which he found a very strong relationship between the existence of an ongoing arms race between two countries and the escalation of a militarized dispute between them to war. Looking at 99 militarized disputes among the major states between 1816 and 1965, he found that 23 of the 26 disputes that escalated to war were characterized by an ongoing arms race (Q = 0.98 and phi = 0.80). In other words, if an arms race was going on between two countries, there was an 82% chance that a militarized dispute between them went to war. On the other hand, only about 4% of the disputes that occurred when both countries were not arming at abnormally high rates escalated to war (Wallace, 1979).

### 2.1 *The Para Bellum Hypothesis*

Wallace's conclusions were an empirical challenge to many long-held assumptions about politics among nations, including some of the foundations of deterrence theory. However, as Weede asserted, Wallace's results were not inconsistent with the *para bellum* hypothesis: 'according to most realist or conservative thinking in the West ... the most dangerous occurrence is not a runaway arms race but the status quo powers losing the arms race' (Weede, 1980, pp. 286–287). In other words, if a revisionist power strives to overturn the balance of capabilities, it is likely that both the revisionist and the status quo states will show high growth rates, but the relevant change is that of the power balance, not the rate of arming. In answer, Wallace used the ratio of expenditure growth rates between revisionist and status quo states to determine who 'won' the race. The results of this test are reported in Table I.

Table I. Test of *Para Bellum* Hypothesis

| | Revisionist 'Wins' | | Revisionist Losses |
|---|---|---|---|
| Escalation | 12 | | 12 |
| No Escalation | 18 | | 53 |
| | Q = 0.49 | $\chi^2 = 5.0$ | $n = 95^{12}$ |

There was some small difference in escalation between a 'victory' on the part of the status quo state or the revisionist one, but the relationship is much weaker than Wallace's original test (Wallace, 1980). A further test, using a static measure of comparative capabilities (the ratio of absolute military expenditures between the revisionist and status quo states) rather than the dynamic ratio of growth rates, achieved even weaker results for the *para bellum* hypothesis (Wallace, 1981). These findings lend support to Wallace's conclusion that it is the arms race itself, rather than the power balance, which is dangerous.

### 2.2 *Focal Points of Criticism*

The two supports of Wallace's study, the disputes he used and the index he employed, both suffered massive criticism. The problem with the sample was that the strength of the findings seemed dependent on the division of the obviously multilateral World Wars into dispute dyads, hence overweighting the arms races that preceded those wars. If a small number of arms races make up the bulk of the supportive findings, it is evident that the certainty with which one is able to make general conclusions about the relationship between arms races and war is undermined.

The index that Wallace used to determine the

existence of an arms race was also challenged. On theoretical grounds, it has been attacked as placing too great an emphasis on the last two years of military spending before a crisis, therefore capturing war preparations rather than an independent arming process. A further problem is that the index is multiplicative – the final index score for the dyad is arrived at through multiplying individual country scores together (Altfeld, 1983). It is conceivable that heavy unilateral buildups could be mistaken for arms races. Certainly in the former case, but possibly in the latter as well, conclusions drawn about the relationship between arms racing and dispute escalation might be overstated. A truly complete study would demand that the index be replicated, but that has yet to be done; although failure to aid in the replication efforts of others is scientifically unforgivable, a useful discussion of Wallace in the present context is still possible. It has been a mistake and an oversight in the critical literature to assume that many of the questions cannot be answered. By comparing Wallace's findings to an analysis of the same dispute sample using another index, a great deal *can* be deduced about the nature of the index that the ongoing debate has neglected to systematically discuss.

Diehl's work (1983a, b, 1985) on the subject of the relation of arms race and dispute escalation is an independent alternative to that of Wallace (1979, 1980, 1982, 1983, 1990). Although he criticizes Wallace at length, he does remain focused on the real issue by offering an alternative means of testing the identical proposition. Unlike Michael Wallace, he found little connection between arms races and dispute escalation. Fewer than 30% of the disputes that he finds occurring during an arms race escalate to war. The corresponding correlation coefficients are very different: Yule's Q drops from 0.98 to 0.42, and phi is an unimpressive 0.13 (Diehl, 1985, p. 338).

Diehl objects to Wallace's findings on both grounds: that his dispute sample was overweighting the World Wars, and that reliability and validity of his index were questionable. Therefore, he offers both a new index and a new set of disputes. In doing this, the criticisms of Wallace's work fail to be compelling because it is impossible to determine the extent of the effects of each weakness. We can certainly guess that both the dispute set and the index are flawed, but without being able to compare them to anything it is impossible to tell whether or not the assertion is true, or, more importantly, what particular influence these factors have had on the conclusions Wallace has drawn. Without such a comparison, it is impossible not to consider the possibility that Wallace has just found a superior means of dividing dangerous from non-dangerous arms buildups.

By looking more closely at the criticisms of the index that Wallace used, and by determining what impact the dispute sample has had on the conclusions, it should be possible to reach a conclusion about the strength of the index, and, consequently, the strength of the relationship between arms races and dispute initiation. In order to do this, two things are needed: first, an alternative measure of what constitutes an arms race apart from either Wallace or Diehl's index, and, second, good consistent data. The latter is taken care of by using the recently updated Correlations of War project data on Militarized Interstate Disputes and on capabilities. This data set now gives access to information from 1816 through 1993. For the former, the measure created by Michael Horn is used.

Horn (1987) offers an independent means of determining when two countries are engaged in an arms race. His relatively straightforward measure has two qualifications for definition as an arms race. First, that the growth rates of a country's military expenditures are on average higher in the period preceding a dispute than in the whole period under study (for him, the Correlates of War period from 1816 to 1980). Second, in order to be classified as an arms race, the growth rate must be higher in the second half of the period than in the first. In this, he attempts to include the criterion of acceleration which is so vital to the Richardsonian distinction between stable and unstable arms races (Richardson, 1960, pp. 74–75). Only when growth is speeding up or seemingly out of control is there an arms race.

He finds that of longer arms races (he looks at Huntington's twelve-year period, cf. Huntington, 1958) over one-half end in war (Horn, 1987, p. 56). Experimenting with a shorter time period, Richardson's suggested six years (Richardson, 1960) yielded no significant relationship. He concludes from this that an arms

race over the longer period is indicative of a continuing conflict of interests and/or hostile relationship between countries, but the shorter measure could easily catch incidental or short-term conflicts that never reach the level of war. He determines that arms races are considerably rarer than others have thought, but they do have some relation to the escalation of disputes to war, though his findings are less robust than Wallace's (Horn, 1987, p. 60; Morrow, 1989, p. 502).

In using this alternative measure and by analysing the previous literature as well as the relationship between arming and dispute escalation in this comparative way, I believe I can put to rest many of the questions still open many years after Wallace's initial article was published, and begin to move past the basic question of the link between arms races and dispute escalation to a more complete analysis of the role of arms races in conjunction with other factors in the process of escalation.

### 3. *Assessing Wallace's Index*

Even without the ability to replicate Wallace's index, I can determine to what extent the index itself affects the results by simply holding the sample constant and calculating the arms race–dispute correlation using another index. As indicated, Wallace found a very strong link between arms races and dispute escalation. His principal findings are reproduced here as Table II.

In order to determine to what extent these findings result from the set of disputes that Wallace used, and what part from the peculiar characteristics of Wallace's index, I calculated Diehl's index for the same set of disputes that Wallace used in his original study.[1] As Table III shows, the results are impressively different.

With Diehl's index, one counts four fewer arms races overall, and only half of the disputes characterized by an ongoing arms race escalate to war.[2] The results are considerably weaker than those obtained by Wallace, although there is still a noteworthy relationship between arming and going to war. While 50% of the arms race disputes led to war, only 19% of those disputes where there was no arms race did so.

Table II. Relationship between Ongoing Arms Races and Dispute Escalation: Wallace's Findings (1979)

| | Arms Race | | No Arms Race |
|---|---|---|---|
| Escalation | 23 | | 3 |
| No Escalation | 5 | | 68 |
| | Q = 0.98 | $\chi^2$=62.95 | $n$ = 99 |
| | phi = 0.80 | | $p < 0.001$ |

Table III. Relationship between Ongoing Arms Races and Dispute Escalation: Wallace's Disputes and Diehl's Index

| | Arms Race | | No Arms Race |
|---|---|---|---|
| War | 12 | | 14 |
| No War | 12 | | 61 |
| | Q = 0.63 | $\chi^2$ = 9.2 | $n$ = 0.99 |
| | phi = 0.31 | | $p < 0.001$ |

Clearly, using the same set of disputes, the two indices yield very different results. The relationship is positive in either case, but Wallace's index shows a markedly more significant relationship between an ongoing arms race between countries and the escalation of a militarized dispute between them to war. Phi (a more conservative correlation coefficient) for Wallace's study was 0.80, and a much lower 0.31 when using Diehl's index. Wallace's index appears to do an extraordinarily good job of distinguishing between episodes of dangerously high abnormal spending on arms and those that are neither abnormal nor dangerous.

I mentioned earlier that the criticisms of the index included the fact that by multiplying countries' growth scores together it may find an arms race where only a unilateral buildup has occurred, and that the index, by heavily weighting the last few years before the dispute, is probably capturing preparations for war which would naturally inflate the results.

The first seems possible. A unilateral buildup would have to be very intense to counteract a non-arming competitor, but it is certainly conceivable that it could do so. Wallace might have been better served by establishing his criteria as Diehl did – each country individually achieving a certain index score that, when considered together, constitutes an arms race. If Wallace is merely capturing unilateral buildups, they must also be extraordinarily dangerous given his findings. This conclusion would be in contradiction to Diehl's findings that unilateral buildups are not associated with escalation (Diehl, 1985a, p. 339).

Theoretically, if this were the secret of

Wallace's strong findings, it would deflate the argument that arms races lead to escalation through the dynamic of a hostile spiral. Escalation would not be the result of *mutual* hostility and perceived threat deriving from high rates of military expenditure growth. Rather, the same effect on the likelihood that a dyadic dispute would escalate to war would be had if both countries were arming or if a single country was massively increasing its arsenal. The willingness displayed by one country and the threat perceived by the other would seem to have the same explosive combination as mutual threat and hostility.

Another possibility is that discovering patterns of unilateral arming only captures planned events. Diehl does not find arms races between England and Germany in 1914 or France and Germany in 1939. If German expenditures alone account for the arms races that Wallace finds in these cases, it is not theoretical support for the conclusion that arms races lead to war. Considering that each of these escalating disputes was initiated by the country that had been spending (at whatever rate of growth) absolutely more on its military,[3] one cannot argue that deterrence failed if these are cases of just unilateral arming.

In any case, it is possible to discern if the addition of unilateral buildups explains the difference in findings. Using Diehl's criteria for an arms race and a unilateral buildup (the country that is arming has an index score of 10% or higher), the hypothesized correlation actually declines when one includes both unilateral and mutual buildups in the calculations.

Wallace's findings are not the result of capturing unilateral buildups with his index. Using Diehl's measure of a unilateral buildup, inclusion of these cases in the test causes a decrease in the strength of the hypothesized relationship. Whereas one half of the cases in which there was an ongoing arms race escalated before, now 35% do. This discovery seems to confirm Diehl's finding that unilateral buildups are less closely related to dispute escalation than mutual buildups are.

Table IV. Relationship between Unilateral and/or Mutual Buildups and Dispute Escalation: Wallace's Disputes and Diehl's Index

| | Unilateral or Mutual Arms Buildup | | No Arms Buildup |
|---|---|---|---|
| Escalation | 20 | | 6 |
| No Escalation | 37 | | 36 |
| | $Q = 0.53$ | $\chi^2 = 5.4$ | $n = 0.99$<br>$p < 0.025$ |

Theoretically, this finding does not make sense if countries base their policies primarily on an assessment of the absolute capabilities that they face. If a strict assessment of capabilities drove policy, the rates of arming of each party would only be considered material when they altered the capability distribution. Mutually high levels of arming should not be highly relevant if they do not change the fundamental position of the countries; in fact, unilateral arming would seem to be more dangerous because it is more likely to significantly challenge the current distribution of power. However, the empirical evidence clearly shows otherwise.

Another explanation for Wallace's exceedingly strong results could be that the value of his index rests so heavily on the last years before a dispute that he is merely capturing war preparation. This problem, too, would undermine his argument that arms races are much more likely to lead to dispute escalation than effective deterrence. Arms races would be primarily artefacts of measurement, not a true independent phenomenon.

As others have pointed out, it is clear that Wallace's measure does heavily weight the two or three years just prior to the dispute (Altfeld, 1983, p. 228). Given the weights that Wallace presents, the last year appears to account for about 50% of the value of k.[4] The last three years account for over 90%. If Wallace's findings are the result of overweighting these years, then his conclusions about the link between arms racing and dispute escalation become an unfortunate product of investigating what amounts to war preparation. However, it is apparent that this factor is not sufficient to explain Wallace's results either.

Wallace finds a much stronger connection between arms races and dispute escalation than does Diehl. When the index formulas are compared, however, one sees that the weights Diehl assigns to the military expenditure growth rates of the years prior to the dispute are different from the weights that Wallace assigns, but not

by a great deal. In Diehl's formula, the last year of growth before the dispute carries 45% of the total, and the three years before carry 90% of the total value of the index for each country. The difference exists, but it is implausible that a difference of this small magnitude could account for the vast differences in the findings obtained using the two indices. This assessment is not consistent with the fact that Horn found that the longer arms races in his study were more likely to escalate than the shorter, which should not be the case if racing were merely war preparation.

4. *Assessing Wallace's Dispute Set*

The second major criticism of Wallace's work was that his dispute set determined his findings. Although a comparison of the indices that Wallace and Diehl each used has shown that there is reason to doubt this conclusion, it is important to look at the issue of the dispute sample used more closely. The criticisms have not been unimportant, as they are based on the concern that the original test cases that Wallace used were not independent, that they determined each other, and by implication could not validly be used to determine the relationship between arming and escalation. Weede (1980) removes a number of dyads that he believes are non-independent, and the relationship between arming and escalation, though still significant, drops considerably. Rather than 82% of arms racing dyads leading to escalation, 55% do.

This problem of the disputes has been a very large obstacle to the resolution of the debate over the role of arming in dispute escalation. The actual effects of the sample used cannot be measured using Wallace's index. Since it has not been replicated, I can hardly apply it to other sets of disputes to determine the change. However, the specific issue of how the disputes themselves affect the findings can be investigated using any two indices and comparing the results using different sets of disputes.[5]

Using the different indices, we already have an idea of how Diehl's findings relate to Wallace's. In order to look at the dimension of the effects of the sampling issue, I use Diehl's index and compare the results obtained by applying it to three different sets of disputes with the results obtained by bringing in Horn's measure of an arms race as a third valid alternative.[6]

Table V. Diehl's Index and Enduring Rivalry Sample

| | Arms Race | No Arms Race | |
|---|---|---|---|
| Escalation | 5 | 8 | |
| No Escalation | 9 | 80 | |
| | Q = 0.69 | $\chi^2 = 7.7$ | $n = 102$ |
| | phi = 0.27 | | $p < 0.01$ |

The first dispute set I look at is that used by Diehl (1985a), which includes 102 cases of major state disputes within enduring rivalries, (eliminating his two disputes between the Soviet Union and Great Britain and the United States in 1946). Each dispute in this enduring rivalry sample was one in a series of disputes between the same two nations. As a result, the set does occasionally contain more than one dyad for each World War; however, the implicit and compelling argument is that these are independent events since they are chosen based on the fact that each dyad had a history of militarized conflict over a period of years.

Using this set of disputes and applying the two different measures of an arms race, one obtains the following results. The Appendices provide tables listing the disputes by participants and year.

Using this dispute set and sample, approximately 36% of disputes occurring during an ongoing arms race escalated toward using either measure. Using Diehl's index, five of fourteen arms racing disputes escalate (Q = 0.68, phi = 0.27). Horn finds eleven racing disputes, four of which escalate (Q = 0.68, phi = 0.25). The correlations in looking at the results of Diehl's index are slightly weaker than those I obtained using Wallace's dispute set, suggesting either that the two sets are only somewhat different or lending added impetus to the argument that it is indeed the index which distinguishes them. In either case, the relationship, although positive, is considerably weaker than Wallace found.

Table VI. Horn's Measure and Enduring Rivalries Sample

| | Arms Race | No Arms Race | |
|---|---|---|---|
| Escalation | 4 | 9 | |
| No Escalation | 7 | 82 | |
| | Q = 0.68 | $\chi^2 = 6.2$ | $n = 102$ |
| | phi = 0.25 | | $p < .0.025$ |

Following Wallace's method of dividing the updated militarized dispute data into bilateral disputes yields a total of 264 dispute dyads, seven of which I eliminate because their proximity to the end of a world war makes calculating a valid arms race measure impossible. I examine the association of arms races and escalation based on this total group as one dispute set, and I remove two classes of dyadic disputes from it to make the third sample. Those two classes are: first, all dyads that were part of a multilateral war that was already underway when they joined; and, second, all escalating dyads that became part of a multilateral war despite the fact that this was the first militarized dispute between the two countries.

The first rule eliminates, among other cases, United States' entry into the World Wars, because we cannot be certain of the independence of events, including the bombing of Pearl Harbor after Japan solidified its alliance with Germany, once the war was underway.[7] It leaves both French and English dyads against Germany at the commencement of those wars.

This is not the tack that Wallace takes in resolving this problem. He chooses instead to merge dyads that entered the war simultaneously. However, he is able to do that because he finds arms races in each instance; using Diehl's index, one does not. If the dyads were merged, it would not be clear whether it would be considered racing or not. Both Wallace and Weede claim to be eliminating non-independent dyads at the beginning of the World Wars, but they do not make it clear precisely which dyads they eliminate, and they obtain very different results (Wallace, 1980; Weede, 1980).

The second rule is meant to eliminate dyadic disputes between countries that end up at war largely through alliances. Because I am not explicitly addressing the role of alliances in war diffusion, I must find an alternative means of allowing for their effects on the findings in the arms race debate. Wallace finds that Austria–Hungary and the United Kingdom were engaging in an arms race at the beginning of World War I, and the dispute escalated. However, Austria–Hungary and the UK had no militarized disputes prior to that. It seems a better argument that this war dyad is the artefact of alliance that each country had with disputants who did have a recent history of militarized disputes, like Germany and France.

Table VII. Diehl's Index and Dispute Set Eliminating Controversial War Dyads

| | Arms Race | | No Arms Race |
|---|---|---|---|
| War | 7 | | 10 |
| No War | 32 | | 183 |
| | Q = 0.60 | $\chi^2 = 7.8$ | $n = 232$ |
| | phi = 0.19 | | $p < 0.01$ |

I am reversing the order in which I look at the other two dispute sets, looking at results obtained through complete desegregation of multilateral disputes into dyads last. In the set which eliminates post-onset dyads and dyads between those with no recent dispute, there is a total of 232 militarized disputes between major states in the period between 1816 and 1993.

Use of this dispute set, as indicated in Table VII and Table VIII, shows quite different results for the two measures. Employing Diehl's index to discern abnormal arms growth from normal yields only a modest relationship between arms races and dispute escalation. Of the 39 total cases of disputes occurring when there was an ongoing arms race, only seven escalated to war. And there remain ten cases of war breaking out between dyads in which both countries were not arming at a rapid pace.

Horn's measure of an arms race is more stringent. Overall, he counts fewer arms buildups as races, and the difference is consistent with distinguishing better between those that escalate and those that do not.[8] Of eighteen arms race disputes, eight escalate to war (or 44%). Nearly half of all disputes that erupt into war do so when the countries are racing. Despite the criticism levelled at the original study by Wallace that his positive results came from having too many non-independent cases, one discovers that with the simple measure of an arms race that

Table VIII. Horn's Measure and Dispute Set Eliminating Controversial War Dyads

| | Arms Race | | No Arms Race |
|---|---|---|---|
| War | 8 | | 9 |
| No War | 10 | | 205 |
| | Q = 0.90 | $\chi^2 = 39.6$ | $n = 232$ |
| | phi = 0.41 | | $p < 0.001$ |

Horn uses, the *elimination* of controversial cases from the total set of dispute dyads actually yields the strongest relationship we find in the data between arms races and dispute escalation.

Before I discuss what these findings mean for the debate over the role of arms races in dispute escalation, I look at the results of the final dispute set. The desegregation of all multilateral disputes into dispute dyads yields 257 disputes. Despite the criticism, levelled at Wallace for his use of all bilateral dyads, I do not believe that it is really illegitimate to do so. In such a small company of states over a period of less than 200 years, we cannot expect international events to take on the same sense of independence that one can realistically claim in an experimental science.

The attempts to solve this problem (Wallace, 1980; Weede, 1980) have shown the difficulty in coming up with generally acceptable criteria for the elimination of controversial dyads. My rules are clear and justifiable, but they still leave room for doubts on a case to case basis. Italy's entry into World War I in 1915 is eliminated because the war had already begun, even though it followed several prewar militarized disputes between Italy and Austria indicating that the dispute between them *was* independent of other relationships. The rules also make no particular allowance for the two theatres of World War II being distinct. However, deciding which cases to remove or include on a case to case basis has not yielded an acceptable standard for the last sixteen years either.

In addition to thinking that the deduction of a sample of truly independent cases that is in itself not controversial is nearly impossible (therefore serving to lead the argument in yet another circle that goes nowhere), I confess that I am not entirely convinced that every controversial case is a non-independent one. The conviction that the war dyads are not independent, I believe, is largely the result of the surety of hindsight. When Lewis Richardson conducted his classic study on the pre-World War I arms race, he did not even include the arms expenditures of Great Britain as part of the arms race. His rationale for this was that he could only legitimately assume interdependence and combine the expenditures of definite allies, and the only allies that he could be sure of before the war were Germany and Austria–Hungary, on the one hand, and France and Russia on the other (Richardson, 1960, p. 31).

But his treatment of World War II indicates that even that was the assumption of hindsight: he was sure of the pre-World War I alliances *because* they had survived onto the battlefield. His study on the pre-World War II arms race was originally conducted in 1938. He chose to conduct his study differently from that of the first War. Rather than combining the expenditure figures of the allies, he calculated each country separately. He did this because he was unwilling to say who would end up allied in the end when war came, although (from our perspective, at least) it must have been relatively obvious by that time. France and Britain had spent much of the 1930s coordinating their diplomatic policies, and it was not until after Munich that the Soviet Union decided to attempt a separate peace with Germany, but Richardson gave no indication of believing before the war that the policies of the allies were excessively interdependent given that he was unwilling to claim, on the eve of war, that certain alliances were set in stone (Richardson, 1960, p. 113). As a consequence, I consider it legitimate to assume that countries have generally relied, in the years before a conflict breaks out, on their own arms policies rather than on their war allies to determine their security.[9]

Therefore, I feel that the use of all 257 bilateral militarized disputes between major states is a valid test of the arms race–escalation connection.

Table IX. Diehl's Index and All Major State Militarized Disputes

| | Arms Race | | No Arms Race |
|---|---|---|---|
| Escalation | 14 | | 17 |
| No Escalation | 39 | | 187 |
| | Q = 0.60 | $\chi^2 = 13.0$ | $n = 257$ |
| | phi = 0.23 | | $p < 0.001$ |

Table X. Horn's Measure and All Major State Militarized Disputes

| | Arms Race | | No Arms Race |
|---|---|---|---|
| War | 12 | | 19 |
| No War | 17 | | 209 |
| | Q = 0.77 | $\chi^2 = 26.5$ | $n = 257$ |
| | phi = 0.32 | | $p < 0.001$ |

As with the second sample explored, the relationship between arms races and disputes escalation is stronger for Horn that for Diehl. Using Diehl's measure, the relationship is fairly weak (phi of 0.23), only about 26% of disputes occurring during arms races escalate to war. Horn, on the other hand, finds that approximately 41% (twelve of 29) of such dyads escalate to war while only 9% of non-racing dyads do. This is a slightly weaker correlation (using Horn's measure) than obtained by eliminating post-onset and war disputes between previous quiet dyads.

This difference between the two measures largely reflects the more conservative definition of an arms race that Horn employs: the level of arms buildup must be quite high and relatively enduring to be classified as an arms race, and that distinction appears to correspond fairly well with the likelihood that a dispute will escalate. Although Diehl does not always find an arms race in Cold War militarized disputes, either, military spending patterns during the Cold War almost never satisfied Horn's stringent criteria for what constitutes arms racing behaviour.

Looking at the results of the four different sets of disputes employing Diehl's criteria for arms racing, we see that changing the sample does have an effect on the results, but it is not very dramatic. The range of Yule's Q is between 0.60 and 0.69, and the range of phi is between 0.19 and 0.31. By comparison, the difference between Diehl's index and Wallace's is markedly more impressive: Yule's Q of 0.63 vs. 0.98, and a phi of 0.31 vs. 0.80. Wallace's results are clearly not determined by his sample, though they are probably mildly increased by it. Rather, Wallace appears to have found a measure that distinguishes remarkably well between dangerous and non-dangerous military buildups.

Without replicating the index with the updated Correlates of War data, I can make no firm assertions about what Wallace is capturing that others are not, but through a careful comparison it is possible to determine what he is *not* doing. The evidence suggests that Wallace's findings are *not* primarily the artefact of either the set of disputes that he uses or that his index is merely capturing unilateral buildups or war preparations. Consequently, it is possible that were Wallace to make his procedure for calculating arms races generally available to the scientific community, there might be grounds for concluding that a much stronger link exists between arms races and war than this study is able to suggest.

It is still possible, however, to draw conclusions about the impact of arming on dispute escalation. Looking at the question in this strictly bivariate way, we know from this investigation that it is possible to devise a measure or index that distinguishes reasonably well between periods of dangerous arms buildups and non-dangerous ones. Looking at the results of Horn's measure, a moderate positive relationship between arming and dispute escalation is clearly discernible. Arms races do not appear from this study to be necessary or sufficient causes of war, but they do increase the probability that war will break out between disputing states. This should not be the case if the old adage 'if you want peace, prepare for war' is good policy.

Looking at the question in this strictly bivariate way, however, is a mistake. The whole debate has tended to focus single-mindedly on this question of an arms race–dispute escalation question without moving on to answer the questions that the results beg. If, for instance, there is a positive relationship between arms races and dispute escalation, why are there so many arms races that do not escalate? And are there discernible patterns in the disputes of other factors that would affect the bivariate analysis?

## 5. *Assessing the Role of Arms Races*

Diehl finds a modest positive relationship between arming and war, but still has a group of 39 arms race disputes which do not escalate. Wallace (1990), when replicating Diehl's index, found that of 43 arms race dyads, 24 went to war, leaving 9 disputes unaccounted for. Even granting that the relationship between arms races and escalation is probabilistic does not justify ignoring the question of those disputes that do not escalate, although both parties are arming at abnormally high rates. Do those non-escalating arms races follow a pattern of their own?

Narrowly focusing on the original question, neither Wallace nor Diehl really addressed the issue. However, certain possibilities have been

offered as explanations for the existence of such a large number of non-escalating arms races.

5.1 *The Impact of Allowing a Time Lag*

Deriving several potential explanations for these cases from the empirical literature, Vasquez & Henehan (1992, p. 104) propose that even given a high level of arming by the participants, a dispute might not escalate if it is only the first or second much militarized conflict between the countries in the recent past, if one or both parties have nuclear weapons, or if the dispute is over a low-salience issue. The research on each possibility suggests that, in any case, the chances of dispute escalation decline.

Elsewhere, I have looked at each of the relationships in more depth (Sample, 1996), but for the purposes of this article I focus on the results obtained when one looks at the relationship of two countries over a period of several years. A first or second dispute between countries is said to be less likely to escalate than a later one, because each dispute does not occur in a contextual vacuum. The character of relations in the past between two countries naturally affects their behaviour in the present. If countries are engaging in repeated militarized disputes over a relatively short period of time, it may begin to appear to the decision-makers that the issues dividing them are unresolvable through normal political methods; the display of resolve evident in the militarization of previous disputes was unsuccessful in heading off future ones (Houweling & Siccama, 1981). Leng (1986, 1993) found that each successive dispute is characterized by more coercive policy and a more conflictive outcome. Hence, it is possible that these disputes that are characterized by arms races but do not escalate are early disputes in rivalries that later became more volatile.

5.2 *Patterns in Non-Escalating Arms Races*

With this possibility in mind, it is necessary to consider more closely the specific disputes that make up the findings. This specification of dyads was not done in the work that followed Wallace's original study, making it more difficult to assess the relationship. Referring to the tables in the Appendix, one sees that the patterns in the disputes that occur during arms races, but fail to escalate, are quite clear. There are two groups of disputes that do not escalate to war. The first group, found by both Wallace's and Diehl's index, is Cold War Disputes, nearly all of which were also characterized by the presence of nuclear weapons in the arsenal of one, if not both, of the participants. This suggests that nuclear weapons may have the effect of raising the threshold of provocation between states – the level of threat at which war is seen as a legitimate way to resolve a conflict (Lebow, 1985). However, using Horn's measure, one finds very few periods of arming during the Cold War sufficiently intense to be designated a race.

More importantly, the other major group of non-escalating disputes consists of those that were clearly early disputes in longer term conflicts. Regardless of which measure is used to count arms races, it is apparent that, with the exception of the Cold War, virtually every case in which two countries are both arming at abnormally high levels was at war within five years. Not all of the disputes are the first or second in a pattern of disputes between particular countries, but in nearly every case, a subsequent dispute did escalate to war. Arms races have not historically been a necessary condition for two or more states to engage in war, but it is rare that an arms race between countries engaging in militarized disputes has not been indicative of a deep-seated antagonistic relationship that fails to be resolved short of war.[10]

Both Wallace and Diehl in their work determined the correlation between arming and escalation by simply looking at the number of arming disputes that escalated immediately. However, when a time lag is introduced it can be shown that 26 out of 28 (or 93%) arms racing disputes in Wallace's 1979 study were between countries that were at war with each other within five years. We can look at the same question using the arms race measure of Diehl and Horn and the complete set of bilateral dispute dyads ($n$ = 257). Diehl finds a total of 52 disputes between arms-racing countries; of those, 62% (32 of 52) were at war within five years (including 47% of those disputes that did not escalate immediately). The numbers are even more striking for Horn. Of 29 disputes that are characterized by an ongoing arms race, 25 of them (or 86%) are between countries that were at war within five years (77% of those that did not escalate immediately did so within five years). Arms

races appear to be more closely associated with dispute escalation than the narrowly focused ongoing debate would suggest.[11]

### 5.3 *Arming and the World Wars*

When counting arms races, regardless of the measure or sample used in the study, it is difficult not to see that the major powers have engaged in few periods of intense arming in the last two centuries. In the arms race category (both escalating and not) a large proportion of the arming was occurring proximate to the World Wars. If one were to argue that there were only two or three independent arms races, the general conclusion that arms races significantly increase the probability that a militarized dispute will escalate to war would necessarily be weakened.

I would argue that each arms 'race' before the World Wars was really several arms races, and the relationship is not substantially weakened. In the 1930s, there were rivalries between Germany and the Soviet Union, Japan and the Soviet Union, the United States and Japan, and Germany and England and France. The war that resulted was the explosive convergence of multiple conflicts that were diffused in large part through alliance structures. To assume that the arms races cannot be counted as independent events is to assume that the rivalries were not themselves independent, and that would be a mistake. Probably, the alliance structures themselves increased the probability that disputes would escalate, and the arms race–escalation correlation may be overstated in this study by that fact. However, it would be as unrealistic to assume that the alliances accounted for all of the increase in probability of escalation as it would be to assume the opposite, especially since alliances did exist in other eras, and arms races did not.

## 6. *Conclusion*

Regardless of the measure used to determine racing behaviour, and regardless of the dispute set used, the proposition that arms races are positively associated with the escalation of disputes to war is upheld by the data. The primary focus of this study has been the resolution of the arms race–dispute escalation debate that began when Michael Wallace found such a strong link in his controversial 1979 article. Although the controversy is impossible to resolve completely while his index remains unperfected, it is clear that many of the criticisms were misdirected. I have shown that his findings resulted from his index dividing well between dangerous and non-dangerous patterns of arming, rather than being determined by either his sample or practical problems with his index.[12] Exploration of the issue using the measures offered by Diehl and Horn indicates that a moderate relationship between arms racing and escalation does exist.

Certain questions have remained unanswered because of the narrow focus of the studies that have characterized the debate, including the reason for the large number of arming, but non-escalating, disputes. The evidence of the empirical literature suggests that disputes that occur later on in continuing rivalries are going to be more likely to escalate. Considering that possibility, I have shown by introducing a five-year lag between a given dispute and the onset of war between two countries that both Wallace and Diehl, in focusing their question so specifically, actually missed the fact that many of the races they considered non-dangerous were, in fact, harbingers of future wars. The inescapable conclusion is that when two states engaged in ongoing arms buildups confront each other in a militarized dispute or a series of militarized disputes, they are far more likely to end up at war with one another than are disputing countries that are not involved in an arms race.

## NOTES

1. Differences between Diehl's reported findings and those reported here can be attributed to updates in the military expenditure data, particularly in both the Soviet and Japanese cases in the 1930s.
2. The nature of the relationship between the Wallace and Diehl indices is not clear. There is considerable overlap in the cases that they designate as overlapping, but those that do not agree do not follow an obvious mathematical pattern (specifically, one cannot reciprocate the other study's results by altering the cutoff point of either index).
3. Dispute initiators taken from Diehl (1983b); military expenditure data from COW capabilities data.
4. The arms race index is the value of $\Delta y \approx (k - y9)/(k + y9)$ for each participant multiplied together. $K = 0.000025y_0 + 0.00014y_1 - 0.00059y_2 + 0.00317y_3 - 0.01929y_4 + 0.0307y_5 - 0.1148y_6 + 0.43097y_7 - 1.60773y_8 + 2.26581y_9$, with $y_0 - y_9$ being military expenditure values for the years before the dispute (Wallace, 1979).

5. The Weede–Wallace exchange in 1980 attempted to resolve this problem. Weede accepted Wallace's definition of arming and altered the dispute set to arrive at new results. However, as a solution, it was unsuccessful. From their descriptions of the non-independent dyads each chose to remove to retest the proposition, one might assume that they had removed the same dyads. Their results show that they clearly did not, and as neither specified precisely which disputes were removed or offered a clear justification for the obvious disagreement, the reader is still left without a satisfactory answer to the questions raised.
6. I have calculated Horn's index over ten years (rather than twelve) using the Correlates of War military expenditure data in current pounds and dollars and controlling for inflation.
7. Consideration of the Pacific theatre of World War II as a separate war leads to the reinclusion of three eliminated dyads: Japan–UK (1940) and the USA–Japan (1940), both of which were racing, but not escalating dyads; and the Japan–USA (1941) war dyad. The Correlations between arming and escalation when these three disputes are included or excluded are not significantly different from each other (Yule's Q remains 0.60 either way).
8. I used all years, including war years, to determine the average military expenditure growth rate of the country. Horn played around with the inclusion of years, using six different categories, based on numbers of battle deaths in that year. He found the strongest correlation between racing and escalation using all years. Although it is true that war expenditures cannot be considered a part of an arms race, the point is to distinguish dangerous patterns of arming from non-dangerous ones. Including all years, military expenditure growth rates in the calculation of an average growth rate does that. And patterns of growth that meet that standard are unquestionably arms races.
9. Sorokin (1994) argues that countries depend on a combination of alliances and arms to gain security. Therefore, allies' arms policies may be related; however, he shows that the fact of an alliance between two countries does not determine in which direction that relationship will lead. Allies may be weaker or stronger, more or less trustworthy or demanding, leaving countries to depend once again on their own assessments of the international situation to determine their own arms policies. Since the existence of an alliance itself is not a good predictor of arms policy, the relationship he asserts does not substantially damage this argument.
10. This does not eliminate the possibility of aborted arms races over the course of the period studied. There were years in which countries were mutually arming, but without tying this arming to disputes in any way it is difficult to make any assertion about the interdependence and salience of their relationships, and hence to make any claims about what that mutual arming meant in terms of increased or decreased chance of conflict between them. The fact that aborted races may have occurred means that one cannot conclusively state that arms races are sufficient conditions for war.
11. Discovering that the vast majority of arming dyads that engage in disputes are at war within five years casts some doubt on finding a clear and meaningful division between stable and unstable arms races.
12. Wallace conceded considerable difficulty in determining which state was the revisionist one on many occasions; four ambiguous cases were removed.

## REFERENCES


Altfield, Michael F., 1983. 'Arms Races? – And Escalation? A Comment on Wallace', *International Studies Quarterly*, vol. 27, no. 2, June, pp. 25–231.

Diehl, Paul F., 1983a. 'Arms Races and Escalation: A Closer Look', *Journal of Peace Research*, vol. 20, no. 3, September, pp. 205–212.

Diehl, Paul F., 1983b. 'Arms Races and the Outbreak of War, 1816–1980'. PhD diss., Ann Arbor, MI: Department of Political Science, University of Michigan.

Diehl, Paul F., 1985. 'Arms Races to War: Testing Some Empirical Linkages', *Sociological Quarterly*, vol. 26, no. 3, September, pp. 331–349.

Horn, Michael Dean, 1987. 'Arms Races and the International System'. PhD diss., Rochester, NY: Department of Political Science, University of Rochester.

Houweling, Henk W. & Jan Siccama, 1981. 'The Arms Race–War Relationship: Why Serious Disputes Matter', *Arms Control*, vol. 2, no. 2, June, pp. 157–197.

Lebow, Richard Ned, 1985. 'Conclusions', pp. 203–232 in Robert Jervis, Richard Ned Lebow & Janice G. Stein, eds, *Psychology and Deterrence*. Baltimore, MD: Johns Hopkins University Press.

Leng, Russell J., 1986. 'Realism and Crisis Bargaining: A Report on Five Empirical Studies', pp. 39–57 in John A. Vasquez, ed., *Evaluating U.S. Foreign Policy*. New York: Praeger.

Leng, Russell J., 1993. *Interstate Crisis Behavior, 1816–1980: Realism versus Reciprocity*. Cambridge: Cambridge University Press.

Morrow, James, 1989. 'A Twist of Truth: A Reexamination of the Effects of Arms Races on the Occurrence of War', *Journal of Conflict Resolution*, vol. 33, no. 3, September, pp. 500–529.

Richardson, Lewis F., 1960. *Arms and Insecurity*. Pacific Grove, CA: Boxwood Press.

Sample, Susan G., 1996. 'Arms Races and the Escalation of Disputes to War'. PhD diss., Nashville, TN: Department of Political Science, Vanderbilt University.

Sorokin, Gerald L., 1994. 'Arms, Alliances, and Security Tradeoffs in Enduring Rivalries', *International Studies Quarterly*, vol. 38, no. 3, September, pp. 421–446.

Vasquez, John A., 1993. *The War Puzzle*. Cambridge: Cambridge University Press.

Vasquez, John A. & Marie T. Henehan, 1992. 'Editors' Commentary on "Arms Races and Escalation: A Closer Look" by Paul Diehl', pp. 103–108 in John A. Vasquez and Marie T. Henehan, eds, *The Scientific Study of Peace and War: A Text Reader*. New York: Lexington Books.

Wallace, Michael D., 1979. 'Arms Races and Escalation: Some New Evidence', *Journal of Conflict Resolution*, vol. 23, no. 1, March, pp. 3–16.

Wallace, Michael D., 1980. 'Some Persisting Findings: A Reply to Professor Weede', *Journal of Conflict Resolution*, vol. 24, no. 2, June, pp. 289–292.

Wallace, Michael D., 1981. 'Old Nails in New Coffins: The

Para Bellum Hypothesis Revisisted', *Journal of Peace Research*, vol. 18, no. 1, March, pp. 91–95.
Wallace, Michael D., 1982. 'Armaments and Escalation: Two Competing Hypotheses', *International Studies Quarterly*, vol. 26, no. 1, March, pp. 37–56.
Wallace, Michael D., 1983. 'Arms Races and Escalation – A Reply to Altfeld', *International Studies Quarterly*, vol. 27, no. 2, June, pp. 233–235.
Wallace, Michael D., 1990. 'Racing Redux: The Arms Race – Escalation Debate Revisited', pp. 115–122 in Charles S. Gochman & Ian Ned Sabrosky, eds, *Prisoners of War? Nation–States in the Modern Era*. Lexington, MA: Lexington Books.
Weede, Erich, 1980. 'Arms Races and Escalation: Some Persisting Doubts', *Journal of Conflict Resolution*, vol. 24, no. 2, June, pp. 285–287.

Appendix A. Enduring Rivalry Disputes: Comparison of Diehl and Horn Measures

| | Arms Race | No Arms Race | | |
|---|---|---|---|---|
| War | 1914 Fra–Ger | 1853 Fra–Rus | | 1939 Fra–Ger (D) |
| | 1938 USR–Jpn | 1853 Rus-Eng | | 1940 USA–Jpn (H) |
| | 1939 USR–Jpn | 1859 Fra–AuH | | 1950 USA–Chn (H) |
| | 1939 Fra–Ger (H) | 1870 Fra–Ger | | |
| | 1940 USA–Jpn (D) | 1903 Rus-Jpn | | |
| | 1950 USA–Chn (D) | 1914 Eng–Ger | | |
| | | 1950 Eng–Chn | | |
| No War | 1895 Rus–Jpn (H) | 1825 Eng–Fra | 1900 Eng–Ger | 1958 Eng–USR |
| | 1935 USR–Jpn (D) | 1831 Eng–Fr | 1904 Eng–Rus | 1958 Fra–USR |
| | 1936 Fra–Ger (D) | 1831 Fra–Rus | 1905 Fra–Ger | 1958 USA–Chn |
| | 1936 Fra–Ger (D) | 1833 Fra–Rus | 1911 Eng–Ger | 1960 USA–USR |
| | 1937 Ita–Eng | 1840 Fra–AuH | 1911 Fra–Ger | 1960 USA–USR |
| | 1937 Ita–Fra | | | |
| | 1938 Eng–Jpn | 1840 Fra–Rus | 1923 Ita–Eng | 1960 USA–USR |
| | 1939 Ita–Eng (H) | 1840 Eng–Fra | 1925 Fra–Ita | 1961 USA–USR |
| | 1939 Fra–Ita (H) | 1848 Fra–AuH | 1926 Fra–Ita | 1961 USA–USR |
| | 1948 Fra–USR | 1849 Fra–Rus | 1926 Fra–Ita | 1961 Fra–USR |
| | 1953 USA–USR (D) | 1849 Fra–AuH | 1932 USA–Jpn | 1961 Eng–USR |
| | 1953 Eng–USR (D) | | | |
| | | 1849 Rus–Eng | 1932 Eng–Jpn | 1961 USA–Chn |
| | | 1859 Fra–Ger | 1933 Rus–Jpn | 1962 USA–Chn |
| | | 1861 Rus–Eng | 1934 Ita–Eng | 1962 USA–Chn |
| | | 1866 Fra–Ger | 1937 USR–Jpn | 1962 USA–USR |
| | | 1885 Eng–Rus | 1937 USA–Jpn | 1962 Eng–Chn |
| | | 1888 Eng–Fra | 1937 Eng–Jpn | 1965 USA–Chn |
| | | 1893 Eng–Fra | 1937 Eng–Jpn | 1965 USA–Chn |
| | | 1893 Eng–Fra | | 1966 Chn–USR |
| | | 1893 Eng–Rus | | 1967 USA–USR |
| | | 1895 Eng–Rus | 1948 Eng–USR | 1968 USA–Chn |
| | | 1897 Eng–Fra | 1948 USA–USR | 1969 Chn–USR |
| | | 1897 Rus–Jpn | 1954 USA–Chn | 1973 USA–USR |
| | | 1897 Eng–Rus | 1954 Eng–Chn | 1974 USA–Chn |
| | | 1898 Eng–Fra | 1955 Fra–USR | 1974 Chn–USR |
| | | 1899 Rus–Jpn | 1955 Eng–USR | |
| | | 1895 Rus–Jpn (D) | 1956 USA–Chn | |
| | | 1935 USR–Jpn (H) | 1957 USA–USR | |
| | | 1936 Fra–Ger (H) | 1958 USA–USR | |
| | | 1936 Fra–Ger (H) | 1958 USA–USR | |
| | | 1939 Ita–Eng (D) | 1958 Eng–USR | |
| | | 1939 Fra–Ita (D) | | |
| | | 1953 USA–USR (H) | | |
| | | 1953 Eng–USR (H) | | |

Appendix B. All Major State Dispute Dyads: Comparison of Diehl and Horn Measures

| | Arms Race | | No Arms Race | |
|---|---|---|---|---|
| War | AuH–Rus 1914 (D) | | Rus–Eng 1853 | Ger–Rus 1914 (H) |
| | Ger–Rus 1914 (D) | | Rus–Fra 1853 | AuH–Rus 1914 (H) |
| | AuH–Fra 1914 | | Fra–AuH 1859 | USA–Ger 1917 |
| | Ger–Fra 1914 | | AuH–Ger 1865 | USA–AuH 1917 |
| | Ita–Ger 1915 | | AuH–Ita 1866 | Ita–Eng 1939 (D) |
| | Ita–AuH 1915 | | Fra–Ger 1870 | Ita–Fra 1939 (D) |
| | USR–Jpn 1938 | | Rus–Jpn 1903 | Ger–Fra 1939 (D) |
| | Ita–Eng 1939 (H) | | AuH–Eng 1914 | USA–Ger 1941 (H) |
| | Ger–Fra 1939 (H) | | AuH–Jpn 1914 | USA–Jpn 1941 (H) |
| | Ger–Eng 1939 | | Ger–Eng 1914 | Chn–USA 1950 (H) |
| | Ita–Fra 1939 (H) | | Ger–Jpn 1914 | Chn–Eng 1950 |
| | USR–Jpn 1939 | | | |
| | Ger–USA 1941 | | | |
| | USA–Ger 1941 (D) | | | |
| | USA–Jpn 1941 (D) | | | |
| | Chn–USA 1950 (D) | | | |
| | Chn–Fra 1950 | | | |
| No War | Ita–Fra 1860 | Ger–USA 1940 (D) | See following table | |
| | Fra–Ger 1871 | Ger–USR 1940 | | |
| | Fra–Ger 1871 | USA–Ger 1941 (D) | | |
| | Ger–Ita 1934 | USR–Fra 1948 | | |
| | USR–Jpn 1935 | USA–Chn 1953 (D) | | |
| | USR–Jpn 1936 | USA–USR 1953 (D) | | |
| | Fra–Ger 1936 | USR–Eng 1953 (D) | | |
| | Jpn–USR 1936 | USR–USA 1953 (D) | | |
| | Ger-USR 1936 | USA–Chn 1953 (D) | | |
| | Ita–Eng 1937 | USR–Chn 1962 (D) | | |
| | Ita–Fra 1937 | USR–USA 1969 (D) | | |
| | Ger–USA 1938 | USR–Chn 1977 (D) | | |
| | Jpn–Eng 1938 | USR–USA 1980 (D) | | |
| | Ger–Eng 1938 | USA–Eng 1980 (D) | | |
| | Eng–Jpn 1939 | USA–USR 1980 (D) | | |
| | USR–Ger 1939 | USA–USR 1981 (D) | | |
| | Ger–USA 1939 | USA–USR 1982 (D) | | |
| | | USA–USR 1982 (D) | | |
| | | USA–USR 1983 (D) | | |
| | Ita–Fra 1939 (H) | | | |
| | Ita–Eng 1939 (H) | | | |
| | Ita–Eng 1939 (H) | | | |
| | Ger–Ita 1939 (H) | | | |
| | Jpn–USA 1940 | | | |
| | Jpn–Eng 1940 | | | |
| | US–Eng.1940 | | | |

Appendix B, cont. All Major State Disputes: Comparison of Diehl and Horn Measures (no arms race/no escalation cell)

| | | | | |
|---|---|---|---|---|
| Fra–Ger 1830 | Eng–Fra 1893 | Ita–Rus 1937 | USA–Chn 1960 | USR–USA 1968 |
| Fra–Rus 1830 | Eng–Fra 1893 | Jpn–Rus 1937 | USA–USR 1960 | USR–Chn 1968 |
| Fra–Eng 1831 | Jpn–Rus 1895 | Jpn–Eng 1937 | USR–USA 1961 | USA–Chn 1969 |
| Fra–Eng 1831 | Jpn–Ger 1895 | Jpn–USA 1937 | USR–Eng 1961 | USR–Chn 1969 |
| Ger–Fra 1832 | Jpn–Fra 1895 | Ger–Fra 1938 | USR–Fra 1961 | USR–Eng 1970 |
| Ger–Eng 1832 | Fra–Eng 1896 | Jpn–Fra 1938 | USA–USR 1961 | USA–USR 1970 |
| Rus–Fra 1833 | Rus–Jpn 1897 | Ita–Fra 1939 (D) | USA–USR 1961 | USR–USA 1970 |
| Rus–Eng 1833 | Rus–Eng 1897 | Ita–Eng 1939 (D) | USA–Chn 1961 | USA–Chn 1971 |
| Eng–Rus 1836 | Eng–Fra 1898 | Ger–Ita 1939 (D) | USR–USA 1962 | USA–Chn 1972 |
| Fra–Ger 1840 | Eng–Ger 1899 | Ita–Eng 1939 (D) | Chn–USA 1962 | USR–USA 1972 |
| Fra–Eng 1840 | Eng–USA 1902 | USR–USA 1948 | Chn–USA 1962 | USA–USR 1973 |
| Fra–AuH 1840 | USA–Eng 1902 | USR–Eng 1948 | Chn–Eng 1962 | USR–Chn 1973 |
| Fra–Rus 1840 | USA–Ita 1902 | Chn–USA 1949 | USR–USA 1962 | USR–Chn 1974 |
| Fra–AuH 1848 | USA–Ger 1902 | Chn–Fra 1949 | USR–Eng 1962 | USR–Chn 1974 |
| Ger–Eng 1848 | Eng–USA 1903 | USR–Eng 1950 | USA–USR 1963 | USA–USR 1977 |
| Ger–Rus 1848 | AuH–Ita 1904 | USR–Eng 1951 | USA–USR 1963 | USR–Chn 1978 |
| Rus–Fra 1849 | Eng–Rus 1904 | USR–Eng 1952 | USA–USR 1964 | USR–Chn 1978 |
| Rus–Eng 1849 | Jpn–USA 1908 | Eng–Chn 1954 | USA–USR 1964 | USA–USR 1979 |
| Ger–AuH 1850 | Ger–Eng 1911 | Chn–USA 1954 | USA–Chn 1964 | USA–USR 1979 |
| Ger–Rus 1850 | Ger–Fra 1911 | Chn–Eng 1954 | USA–USR 1964 | USR–Chn 1980 |
| Ger–Fra 1850 | Ita–Fra 1911 | Chn–USA 1955 | USR–USA 1964 | USR–Chn 1979 |
| Rus–AuH 1854 | Ita–AuH 1911 | USR–Fra 1956 | USR–Chn 1964 | USA–UST 1986 |
| Fra–Ger 1859 | Eng–Ita 1911 | USR–Eng 1956 | Chn–Eng 1964 | USR–Chn 1986 |
| Eng–Rus 1861 | Fra–Ita 1912 | USR–USA 1956 | USA–Chn 1965 | USA–USR 1986 |
| Fra–Ita 1866 | AuH–Rus 1912 | USA–Chn 1956 | USR–Eng 1965 | Ger–USA 1940 (H) |
| Fra–Ger 1866 | USA–Ger 1915 | USR–USA 1957 | USR–USA 1965 | USA–Ger 1941 (H) |
| Rus–Eng 1876 | USA–Ger 1916 | Chn–USA 1957 | Chn–USA 1965 | USA–Chn 1953 (H) |
| Rus–Eng 1877 | USA–Ger 1916 | USR–Eng 1958 | Chn–USA 1965 | USA–USR 1953 (H) |
| Ita–AuH 1877 | Eng–USR 1923 | USR–USA 1958 | UST–Chn 1965 | USR–Eng 1953 (H) |
| Rus–Eng 1878 | Fra–Ita 1925 | USA–USR 1958 | USR–USA 1966 | USR–USA 1953 (H) |
| Rus–Eng 1885 | Jpn–USR 1926 | USA–Chn 1958 | Chn–USR 1966 | USA–Chn 1953 (H) |
| Ger–Fra 1885 | Fra–Ita 1926 | USA–USR 1958 | Chn–USR 1967 | USR–Chn 1962 (H) |
| Ger–Fra 1887 | Ita–Fra 1927 | Chn–Eng 1958 | USR–USA 1967 | USR–USA 1969 (H) |
| Ger–Fra 1887 | Ita–Eng 1927 | USR–Fra 1958 | USR–USA 1967 | USR–Chn 1977 (H) |
| Ger–Eng 1887 | Jpn–USA 1932 | USR–Eng 1958 | USR–Eng 1967 | USR–USA 1980 (H) |
| Fra–Ita 1888 | Jpn–Eng 1932 | USR–USA 1958 | USA–USR 1967 | USA–Eng 1980 (H) |
| Fra–AuH 1888 | Rus–Jpn 1932 | USA–USR 1958 | USR–USA 1967 | USA–USR 1980 (H) |
| Fra–Eng 1888 | Rus–Jpn 1934 | USR–USA 1960 | USR–Chn 1967 | USA–USR 1981 (H) |
| Eng–Ger 1888 | Ger–Rus 1936 | USR–USA 1960 | USA–Chn 1968 | USA–USR 1982 (H) |
| Eng–Ita 1890 | Rus–Jpn 1937 | Chn–USA 1960 | Eng–Chn 1968 | USA–USR 1982 (H) |
| | USA–USR 1984 | USA–USR 1985 | USR–UKG 1985 | USA–USR 1983 (H) |

Appendix C. Intra-war Disputes and Escalating Dyadic Disputes between States with no Previous Militarized Dispute

| | | | | |
|---|---|---|---|---|
| 1854 Rus–AuH | 1914 AuH–Eng | 1915 USA–Ger | 1940 Jpn–Eng | 1941 USA–Ger (non-escalating) |
| 1859 Fra–Ger | 1914 AuH–Jpn | 1916 USA–Ger | 1940 USR–Eng | 1941 Ger–USA |
| 1866 AuH–Ita | 1914 Ger–Jpn | 1916 USA–Ger | 1940 Jpn–USA | 1941 Ger–USA |
| 1914 Ger–Rus | 1915 Ita–Ger | 1917 USA–Ger | 1940 Ger–USA | 1941 USA–Jpn |
| 1914 AuH–Fra | 1915 Ita–Aus | 1917 USA–AuH | 1940 Ger–Rus | 1950 Chn–Eng |

Removal of these disputes from the full set of major state militarized disputes yields the second sample employed in the study (final $n = 232$).

Appendix D. Non-Escalating Arms Races: Wallace and Diehl Indices

| Wallace's findings (1979) | Using Diehl's index (Wallace, 1990) |
|---|---|
| 1935 Japan–Soviet Union | 1913 Italy–France |
| 1936 Great Britain–Germany | 1935 Japan–Soviet Union |
| 1938 Great Britain–Germany | 1935 Great Britain–Italy |
| 1962 United States–China | 1936 Great Britain–Germany |
| 1962 United States–USSR | 1936 France–Germany |
| | 1938 Soviet Union–Japan |
| | 1940 Japan–Great Britain |
| | 1948 Soviet Union–France |
| | 1953 Soviet Union–US |
| | 1953 Great Britain–USSR |

SUSAN G. SAMPLE, b. 1969, PhD in Political Science (Vanderbilt University, 1996) is Visiting Fellow 1996–97 at the Peace Research Centre, Australian National University.