PREDICTING CHANGES IN USA MILITARY PERSONNEL USING NEURAL NETWORK MODELING -- TWO APPROACHES

Paul R. Williamson, Ph.D.

Research Report 9601
Global Vision, Inc.
P. O. Box 4394
Ann Arbor, MI 48106-4394

May 1996

The data used in this study were provided to me by the Correlates of War project, University of Michigan, J. David Singer, Project Director. Completing the study and writing this report were aided by a grant from the Hoover Institution, Stanford University, which I obtained with the help of Bruce Bueno de Mesquita, Senior Fellow at the Institution. The essential computations were done using software selected for Global Vision, Inc. through the advice and intervention of GV officers Myron Karasik and David Ford, and Scientific Advisory Panel member Julie Dickerson. For this valuable help I thank all these parties.

Contents:

Appendices:

Introduction

In this report I discuss the first results of using artificial neural network modeling techniques to predict changes in material capabilities of nations, with application to changes in numbers of military personnel of the United States. This use is part of a larger study, the goal of which is "global forecasting"-- that is, to predict behaviors and changes in all parts of the world as realistic and therefore probably complex functions of many inter-connected societal and environmental factors. However, I focus here on a useful initial task permitted by the modest resources currently available: to try out the indicated techniques on a few readily available variables and cases, then gradually to add more of them. This was begun in the summer of 1995 and resumed in March through April 1996. Such an incremental approach means the initial predictions are likely to be inaccurate if only because of too few cases and variables; indeed the initial results are not highly accurate. Accordingly, often I will use the term "projection" to describe these results, so as to distinguish the mechanical act of putting out trial numbers from the judgment (not yet warranted) that such numbers might reasonably be regarded as predictions.

Why, then, would one use artificial neural network modeling techniques for global forecasting? A short answer is, because one expects that, with additional effort, predictive accuracy will come progressively closer to whatever limits may turn out to be inherent in various aspects of global phenomena. I believe the present results are consistent with that expectation, as we will see below.

So how is the neural net approach advantageous to such progress? Let us consider first the way modeling normally is begun. When one sets out to model a system in order to predict its future behavior, it is very useful not only to know what the pertinent variables are, but also to have an idea of the form of the relationships among them. The value of such knowledge is it indicates the appropriate mathematical and other techniques. For example, if one starts with a belief or assumption that each variable is a linear compound of others, then one knows right away that multivariate regression techniques may be appropriate. The subsequent task of developing the model then revolves around solving for regression weights and doing other things appropriate to those particular techniques.

Unfortunately for ease of progress (but what may add interest to the subject), it is common in modeling human society that one starts with an idea of the relevant variables but with no good idea about the mathematical form of their mutual relationships; so there appears no best starting place for the undertaking. Faced with this ignorance, social modelers have started by assuming a variety of often mutually incommensurate types of relationship to represent what looked like the same thing. (An example is the plenitude of arms race models stemming from Lewis F. Richardson's original work; see the bibliography in Etcheson, 1989, pp. 183 - 219.) Typically also, the assumptions going into such models have seemed either unrealistically simple or unmanageably complex. And often the models have relied on unobserved conceptual variables ("utilities" is common) which then must be related to observable data in a manner (the "auxiliary theory") not prescribed by the original model, indeed not intrinsic to it at all.

Given the above circumstance, the artificial neural network approach offers the following advantage: it does not require one to specify functional relationships among variables in order to get a working predictive model of observables. Instead, the approach "learns" to emulate whatever functional relationships help predict the data, through a process of repeated exposure to examples of past, known initial conditions and resulting outcomes. This process is known as "training." These alleged advantages are elaborated and the network technique, itself, is described in Appendix 1 and in sources cited there. (Herein, I use "network" to refer to an artificial neural network model.)

In this report I discuss the results of training various network schemes to predict annual changes in numbers of United States (herein, USA) military personnel. By "prediction" I mean both forecasts of future changes or events, which we do not yet know, and forecasts which deal with known changes or events occurring at some moment now past, but which attempt to make use only of information available prior to the moment being predicted. The latter exercise is sometimes called "postdiction" to distinguish it from future prediction. While I find this an awkward phrase and have mostly avoided it here, clearly there are some issues to which the distinction is relevant. First, the true interest may well be in future prediction; yet we have only the past on which to practice. So we must try to "postdict" until the outputs seem accurate before turning to the ultimate intended application. In addition, the postdiction effort must truly avoid using that which would not have been available at the time but which now we know. Often this is straightforward but sometimes not; data, seemingly "current" but prepared after the event, may actually reflect knowledge of what followed. Later on we will see this problem afflicts one of the data series encountered in the study.

In the next section I discuss the use of military personnel changes as an output-- the indicator to be predicted --in a "direct approach" to prediction. Next I describe an alternative way of representing changes in numbers of military personnel. I call this way the "change components" approach. It became the usual form of the outputs to be predicted by the networks used in this study. Then I describe the inputs used for training and testing the networks; after that, I present numerical results and, finally, conclusions.

A "Direct" Approach To Prediction

First a word about the form of data on numbers of military personnel and on other national characteristics discussed later: all such data were transformed to their logarithmic form. (For those not familiar with this term, the logarithm of a given number is just another number related to the given number by a certain mathematical formula.) One advantage of logarithms is that proportional changes have the same value regardless of the value of the original number. (This means that, for instance, a 10% increase from 100 is 10, and from 2000 the same % increase is 200; but the difference between log 110 and log 100 equals the difference between log 2200 and log 2000.) So the same proportions of change can be easily spotted in data which vary greatly, such as numbers of military personnel spanning two centuries. Furthermore, while the original changes may well become larger or smaller as the value undergoing change grows or declines, it is quite plausible the proportional changes stay within the same range (e.g. within 10% /year). This boundedness of change is helpful for achieving an accurate network (Lawrence and Fredrickson, 1993).

One approach is to estimate the change in logarithm of military personnel directly; that is, the change in logarithm is the output to be predicted. The reasons for trying such an approach are twofold: its simplicity and the possibility that a neural network can do better without being presented with a breakdown of the output into components, such as in the manner discussed below. It might well turn out that such a component approach does better than direct; even if so, it is of interest to see by how much. In other words the direct approach can serve as a benchmark. As we will see, the direct approach performed with comparable (or somewhat greater) accuracy in this study.

Defining Changes Using "Change Components"

An important choice in devising a network is to decide what to tell it, to begin with, about the dynamics of the system it is supposed to model. (You tell it by constraining its form one way or another-- that is, by allowing a certain number of elements, permitting or not permitting various connections among them, and deciding what variables to submit as inputs, to be trained against what outputs.) If you tell it as little as possible, then it will not so likely be confused by your mistaken ideas; but it will not be helped by your correct ideas, either, so it may require many more training examples from which to acquire an acceptable degree of predictive accuracy. Also, if you tell it very little, it may end up with a solution bearing no apparent relationship to your previous ideas; which may leave you frustrated and curious, still wondering “do these phenomena behave as I thought, or not?”

I was moved by such considerations in favor of certain constraints reflecting my idea that the actual military and other national capability values probably are approximated, over a sufficiently long period, by the law of *compound growth*. So I geared most of my neural network studies to try to resolve each annual change into some parts that are fit by such a law and to model those parts; then to model separately the remaining change part that does not fit.

Let us see how that works, starting with the compound growth part. Compound growth refers to change, over some time period, that is proportional to the value at the beginning of the period-- or, for finite time-increments, in the previous period. One could call this beginning time or the time of the previous period, the “reference time.” In this study the time period was one year, so let us refer to this period as the “reference year”. Compound interest on money savings is an example of such growth. Many capability indicators plausibly exhibit approximate compound growth. Total population does so, if there is a constant expectation of number of children born per female in the population (and if the proportion of boy to girl children is constant, etc.); military personnel, if there is a tendency to maintain military personnel in numbers proportional to the population; and energy consumption, if the means of energy production grow at a constant rate due to reinvestment of productive assets and if the energy is consumed by those who produce it. Again, my approach sought to represent such a growth pattern to the extent it is true.

As suggested earlier, the first step in doing so was to transform the initial data values by using their logarithms to make a logarithmic plot of the data. The effect of this transformation is to make compound growth more easily recognizable, which will happen because a plot of logarithms of values obeying such growth forms a straight line. This transformation also makes compound growth easier to analyze because the formula for a straight line is very simple, depending only on numbers describing intercept (the height) and slope of (how steep) the line. The next step is to find the trend line, the straight line which fits as closely as possible to the series of logarithmic-transformed data.

So far as the compound growth part is concerned, the predicted change in value from one period to the next will have two aspects. The first aspect is simply the straightforward extrapolation of the trend line to the first unknown period.

To see the second aspect, note that this line is represented numerically as a series of estimated values, one for each year of observation. Calling this series the trend line series, it would begin with some initial year and end with some final year. Moreover, I chose this final year to be the reference year (see above) of the value change to be predicted. For instance, prediction of the year 1836 from the reference year 1835 was based on the trend line series 1816 through 1835. The idea is that one wants to make the trend line reflect the experience of a certain preceding number of years. Notice, however: when one gets to predicting change from the next reference year, say 1837 from reference year 1836, one moves the trend line series forward by one year, so now it begins in 1817 and ends in 1836; thus there is a new trend line. One moves the series forward because the trend itself may have changed and one wants to detect that change. This change in trend line itself -- which is to say changes in its parameters (its slope and intercept) --is the second aspect of change from one year to the next which one needs to anticipate in the change model I was using. That one is trying to predict the change in the line, also means one needs two series in succession to start the prediction process. For instance, the trend 1816 through 1835 produces a set of parameters, then the trend 1817 through 1836 produces a second set; subtracting the first set from the second gives the parameter changes for the year 1836. These changes, together with the parameter values themselves and sometimes other indicators as well, is the information which one will then seek to use to predict parameter changes for the year 1837. Given the procedure as just described and the data which were available, you can see that 1837 is also the first year for which a prediction could be attempted or which could provide the output values for training a neural network. (1990 was the final reference year.)

The actual procedure did not directly estimate changes in the intercept, however. The reason is, changes in the intercept value are correlated with changes in the slope. By visualizing how a trend line would change with slope, you can see why. If this trend line is drawn on a graph then its intercept is the distance above the horizontal axis of this graph at which the line crosses the vertical axis of the graph. Now increase the slope to make the line steeper. This change occurs at the point which marks the reference time, typically at some place to the right of the vertical axis; thus as the straight line becomes steeper its tail end to the left dips down and strikes the vertical axis at a lower spot, corresponding to a decreased intercept. Clearly, then, changes in slope and intercept are negatively correlated (an increase in one with a decrease in the other); but when the contribution of the slope is removed, an independent component of intercept change (not correlated with slope) remains. So as not to make two neural networks do the work of one, my approach estimated first the intercept change, then separately the change in this independent component. The component, which I call long-term shift, equals the observed value of the second series final year minus the same of the first year of the first data series. (In the example discussed above, long-term shift for reference year 1836 would equal the logarithm of the 1836 value of military personnel minus that of the 1816 value.)

In addition to dealing with the logarithmic trend line and changes in it, there is also the need to deal with deviations from it. That is, even if a predictably changing straight line could be seen as the trend, one reasonably expects also to see plenty of variation above and below such a line. These deviations or “residuals” were computed simply by subtracting the trend line estimates from the true logarithmic values.

In sum, these considerations boil down to predicting three values describing logarithmic variation of military personnel:

- change in slope of trend line (dm_pr)
- long-term shift of observed values (s_pr)
- deviation of logarithmic values from trend line (r_pr).

I refer to each of these values as a change component. If these components are accurately predicted, then their values can be used, together with values taken from the reference year, to compute an accurate prediction of the overall change the following year in the capability indicator of interest. This computation is explained in Appendix 2. The main approach of this study was to predict each of these three change components separately using its own network; then to combine them to predict the overall change. Within parentheses after the bulleted items are change component abbreviations used in the remainder of the report. These and other symbols are described in Appendix 3.

Inputs Used In The Study

Inputs to the neural nets consisted of three groups: information on (1) certain military, industrial, and demographic characteristics of the USA-- referred to as “material capabilities”, (2) domestic and foreign involvement of the USA in conflict, and (3) military personnel and conflict involvement of a group of nations frequently classed by international scholars as “major powers” (other than the USA) during part or all the period of study. Time lags of varying degree were tried for each input. These lags consisted of presenting as inputs the value of the input for 1st, 2nd, 3rd, ... nth years preceding the reference year. Closely related is presentation of both the value in the reference year and its increment from the prior year, which gives the same information as a one year lag. The inputs are described below and, more completely together with increments and time lags, in Appendix 4.

(1) USA Material Capability Indicators: These were total population, energy consumption, pig iron or (after 1900) steel production, military expenditure, and numbers of military personnel, itself. These inputs were used in two different ways corresponding to the change component and direct methods mentioned earlier. In the first way, total population, energy consumption, and military expenditure itself (all in the reference year or previous) are broken into 18 change components which supply the basis for inputs. In the second way, the original values are used to generate inputs, using all 5 variables in logarithmic form.

(2) USA Foreign and Domestic Conflict Involvement Indicators: These inputs were based on two types of data series. The first covers national involvement in “militarized interstate disputes” (MIDs)-- events in which one or more of threats to use military force or displays of military force (alerts, mobilizations, etc.) are directed from one sovereign state government to another state, or in which there is an armed clash or war between two such states. The second data series covers involvement in civil wars, of which one sub-type is involvement by a state in such a war on its own domestic territory. One set of indicators used these data series to list the total number of cases of involvement in MIDs by the United States by year, in the form of entry to a new or ongoing MID. In this set, continued involvement carried over from prior years was omitted. A second set of indicators was based on the number of interstate wars plus domestic civil wars underway involving the United States. (In addition to the War Between the States, the “domestic” category included the war against the Sioux nation.) In this set, involvement was listed whether begun or carried over from a prior year.

(3) Military Personnel and Conflict Involvement of Major Powers: The parties contributing to the list of non-USA “major powers” are described in Appendix 4. The inputs of this group were collective; in that they were all summations of indicators across all parties on the list, combined. One of the inputs was yearly change in logarithm of the total number of military personnel for all these states. A second input was total number of MIDs starting in each year. This number was presented in two ways: In the first, distinction was made between MIDs in which the action included a threat or display of force but not armed clash, and those in which armed clash or war resulted. The respective totals were presented as separate inputs to the networks. These totals appear in an input set labeled ext_1 . In the second way, no distinction was made among types of MIDs. These appear in a set labeled ext_2 .

Results Of Data Analysis

Each of several combinations of input and output variables was employed in its own training session, following which a test was made using data not previously used for training in the session. The more important of these steps are summarized in Appendix 4. In what follows, the results of each training session are labeled as *training* cases and, of each use of the reserved test data, as *test* cases. The results of each such test were summarized by scatter plot and a product moment correlation value, based on historically recorded versus projected outputs. Some training results, selected because the test results were unusually accurate, were also scatter plotted and correlated.

Before viewing the numbers, some further words are in order about what is shown. I am seeking “good” predictions which means, at this juncture, closeness of agreement between projected and historically observed values. To quantify this agreement, I have chosen for each session to compute the Pearson product moment correlation index between projections and observations. The variety of results obtained span a range from zero, even slightly negative, correlations to near-perfection (as great as 0.99 out of a maximum possible 1.00). At this juncture there is no intent to measure statistical reliability of findings. (Indeed, it is unclear what that would mean in a network model.) Instead the purpose is to use the modeling as a search procedure to find the best-fitting results, subject to the condition that they seem otherwise valid for subsequent study; so the emphasis is on such best results. (A more complete listing of sessions is given in Appendix 4, Table 5.) Above each table is the name of the output variable evaluated. For example in Table 1, the notation dl_pr tells us the output to be projected was changes in logarithm of military personnel. Below each table at the caption “inputs” are labels naming the sets of inputs used in that session. These sets are defined in Appendix 3. Below that is a “procedure” caption; it indicates the criteria-- number of iterations of the complete collection of cases through which the back propagation algorithm cycled and, possibly also, an index of accuracy --used to terminate the training session. Thus, in Table 1 the caption indicates the network resulted from running through the complete list of training cases 500 times.

In Table 1 we see the best results of the direct approach. The input sets are indicators of capability, the set dir_1, and of USA conflict, cus_1. Notice the fit for test cases (0.92) is very good; yet for training cases it is only moderately accurate. Examination of projected versus true outputs of test cases, Figure 1 and Table 6, Appendix 4, suggests the high correlation is largely due to two extreme cases: large postwar decreases which occurred in 1865 and 1945. The positive aspect is, the network was able to anticipate these two demobilizations with fair accuracy. (The inputs give no explicit coding of termination of Civil War or Second World War hostilities; either the decreases were found by chance or were implicit in variations of past military personnel levels and timing of conflict onsets.) Remove these extreme cases, and the picture appears consistent with the inferior result obtained in the larger number of training cases.

Table 1. Summary of best results, direct approach

Product moment correlations, projected vs. observed changes in log military personnel, best fitting results:

	dl_pr
test cases	0.92
training cases	0.47

Inputs: dir_1, cus_1
Procedure: training was terminated at run #500.

Figure 1. Observed versus Projected Increments
Logarithm of USA Military Personnel
Direct Approach, Test Cases

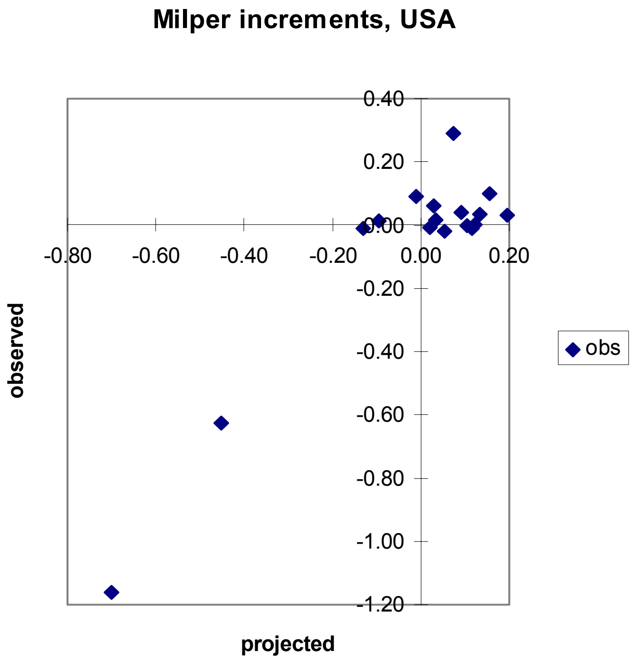


Table 2. Change component summary.

			Component test case correlations:		
Line	Inputs	Procedure	dm_pr	r_pr	s_pr
1	cap	1	0.93	0.89	0.90
2	cap, cus_1	1	0.92	0.88	0.95
3	cap, cus_2	1	0.98	0.85	0.96
4	cap, cus_1, ext_1	1	none	0.93	none
5	cap, cus_1, ext_1	2	0.98	none	0.94
6	cap, cus_1, ext_1	3	none	0.98	0.99
7	cap, cus_1, ext_2	3	none	0.95	0.98

Inputs: see Appendix 3.

Procedures:
Neural net was trained to error criterion 0.x / run number y whichever came first:

1	0.20 / 300
2	0.15 / 500
3	0.10 / 500

In Table 2 we turn to correlations of projected versus observed change components as outputs, using various inputs and training procedures. (These are projections of the type which then subsequently were combined to project the net change in logarithm of military personnel, discussed in a moment.) The inputs consist of the standard set of 18 change components cap, two alternative sets of USA conflict indicators cus_1 and cus_2, and two alternative sets of foreign major power indicators ext_1 and ext_2. Correlations marked “none” refer to networks not developed. The supplementary table below describes the numbered procedures.

At the risk of belaboring the obvious, the correlations vary only in degree of excellence. One of the surprises of the study is how easily these high values were obtained; at the start, I thought this would be the hard part.

Finally, we turn to how these change components translated to projections of overall or net change in military personnel. Four trials, each drawing on somewhat different component estimates are depicted in Table 3. (As a check on arithmetic, I put the *true* change component values for the test cases into the net change estimation formula. The resulting numbers exactly equal the true net change values for those cases, which validates the formula.) The “component source” numbers show the line in Table 2 from which the corresponding estimate was drawn.

Table 4 focuses further on trials 3 and 4, displaying product moment correlations, projected versus observed changes, of change components along side the resulting estimated annual (logarithmic) increments in numbers of military personnel, net_pr . Notice the favorable test correlations of trial 3 held for the training cases as well. Figure 2 presents, for trial 4, scatter plots of projected versus true (“observed”) outputs for each of the three change variables and for the resulting computed versus observed net change in logarithm of military personnel; Table 7, Appendix 4 presents the corresponding data. (For reasons to follow, the much better appearing plots for trial 3 are not reliable, thus I omit them.)

While trial 3 looks like a better result than trial 4, the looks may be deceiving. This is because the input difference is solely the use of foreign conflict set ext_1 in trial 3 versus ext_2 in trial 4, but ext_1 *should not* be relied upon without further work to re-code the MID input variables. As we saw earlier, ext_1 breaks the MID onsets into those which did, versus those which did not present armed clash or war; whereas ext_2 does not distinguish the disputes. Now the method of coding assigns an armed clash or war code to a dispute case without indicating in which year of the dispute the clash or war first occurred. (That information is available in narrative summaries but not as machine-readable code.) Of 507 MID armed clash / war acts of participation by the non-USA majors during the study period, 195 lasted longer than one year. Of the latter, some unknown number of them may first have shown such violence in some year later than the reference year-- that is, the disputes began as threats or displays of force only and later intensified. For such cases, the input is making use of information still in the future thus not yet known in the reference year. (This is the postdiction pitfall problem mentioned earlier.) Thus the inferior result of trial 4 actually represents the best result in which we can be confident at this juncture.

Table 3. Net change in military personnel.

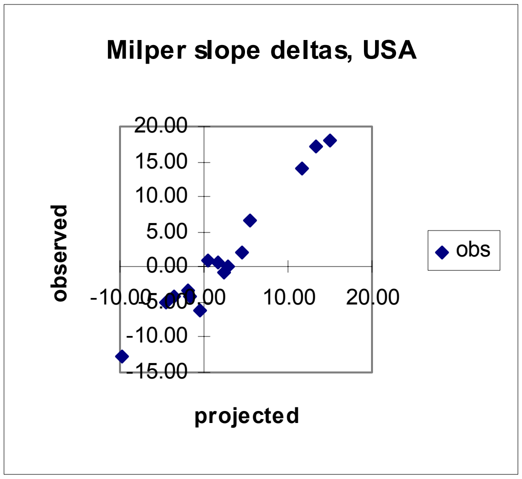
	Component source (Table 2 line nbr):			net_pr test case correlations:
Trial	dm_pr	r_pr	s_pr	
1	1	1	1	-0.10
2	3	4	5	-0.03
3	3	6	6	0.70
4	3	7	7	0.38

Table 4. Summary of product moment correlations

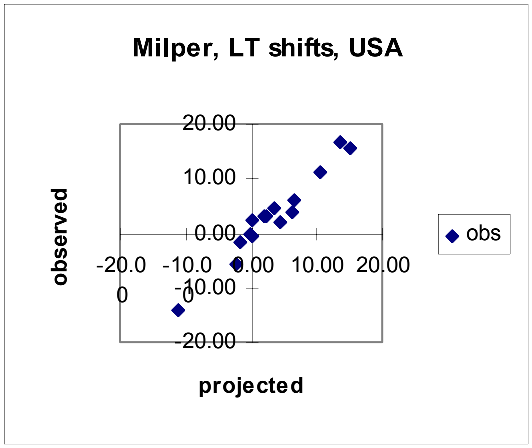
	dm_pr	r_pr	s_pr	net_pr
test cases, trial 3	0.98	0.98	0.99	0.70
training cases, trial 3	0.89	0.91	0.95	0.74
test cases, trial 4	0.98	0.95	0.98	0.38

Figure 2. Observed versus Projected Increments, Trial 4
Logarithm of USA Military Personnel
Separate Components and Net Change

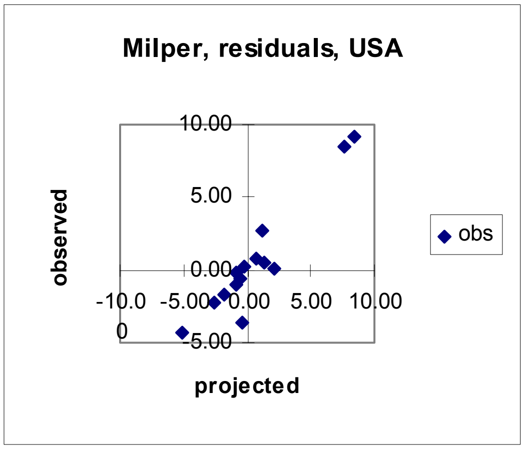
dm_pr :



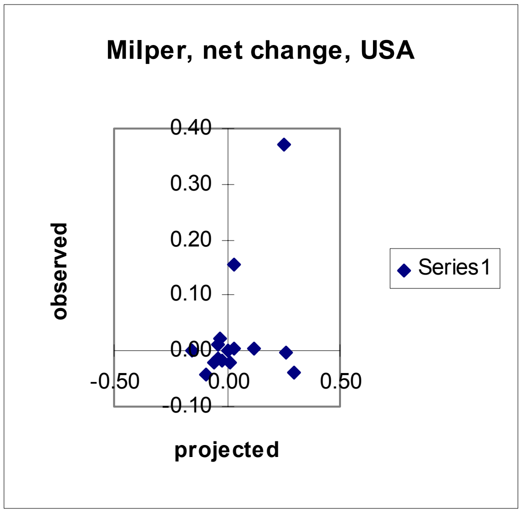
s_pr :



r_pr :



net_pr :



The trial 3 results still are of interest, however. More than half the armed clash / war participation cases are coded at less than a year duration; and if the 195 suspect cases were to be re-coded according to the actual year in which violence first appeared, surely some of these start years would remain unchanged. (For instance, in a cursory check of the 7 USA disputes lasting more than 1 year and ending in war participation, I found only two of them-- Mexican and Persian Gulf Wars --did not involve military violence the first year. Note those two cases compromise the wuw variables of data sets cus_1 and cus_2, albeit to a minute extent.) So, after re-coding, one might expect the correlation of trial 3 to be weakened somewhat but not completely.

Now let us look at the findings as a whole. These projections both succeeded and failed in a way exactly contrary to my original expectation, which was the change elements would be elusive but, once obtained, would readily yield the net change.

First, the expectation was that, at minimum, additional information about USA involvement in USA Civil, First and Second world wars, and other wars, would be needed in order to reflect changes in numbers of military personnel independent of prior variations in USA material capabilities. However even the capability indicators, alone, seem to model the military change components with reasonable accuracy. (The smallest correlation in Table 2, 0.85 , is pretty respectable for social research.)

Yet the net change projections either succeed only modestly or fail completely. Several points go with this surprising (or seemingly paradoxical) result:

1) The weakness applies even to the direct approach (Table 1), manifest in the previously noted qualifications. This suggests the low correlations are not just (or at all) a problem resulting from the network being confused by the components approach.

2) Among the component results, there are noticeable differences of accuracy; compare trials 1 and 2 with trial 4 (and with 3 to the extent it is true). And the better results of trials 3 and 4 *are* associated with slightly greater accuracy in the components, as comparison of tables 2 and 3 reveals. The same is true between trails 3 and 4 alone, as the side-by-side comparison of Table 4 shows.

3) These results show extreme sensitivity in accuracy to *very slight* changes in component accuracy. (For example, the minor differences between trial 3, using ext_1, and trial 4, using ext_2 in reproducing the change variables of Table 2 translates into the considerable difference in accuracy of final result of 0.74 versus 0.38.) Why, is unclear. (Non-linearity is not the answer; the net change estimate is a linear compound of the components, equation (2.12) , Appendix 2.) Two possibilities: First, residual r_pr errors are somewhat correlated with the other two component errors (0.47 and 0.41 with dm_pr and s_pr, respectively); the years of greatest error tending to be the same, the errors cumulate. Second, the least accurate component, r_pr (Table 2), tends to make larger contributions than the other components.

To elaborate, somewhat, the third point, extremely high component accuracy was needed, to which the USA conflict and foreign indicators were helpful but not sufficient. While comparison of trial 1 with the others fits the prior expectation that capabilities alone would not be sufficient, projection failed completely (Table 3, trial 2) even after the additional data were introduced. What finally distinguishes trials 3 and 4 is that the correlations were no less than 0.95 in trial 4 and no less than 0.98 in trial 3, among all three components. Just why those networks were more accurate, again is unclear. Worth noting: for residual and long term shift (r_pr and s_pr) this accuracy was achieved by training the networks to the more exacting standard of procedure 3. (In fact, I switched to procedure 2, then 3 when I saw the higher correlations which followed. Introduction of starting connection randomization and noise techniques in training may also have helped.) A further interpretation is, the result illustrates the rapid accumulation of errors. This difficulty, sharply declining accuracy with subsequent steps, will no doubt be encountered on many occasions in modeling global change.

Conclusions

We can summarize these results in two observations:

- The three change elements were well correlated with calculations based on observed data; the networks appear to be a valid basis for predicting these components.

- The projections of net change, both direct and component approaches, were poorly to moderately well correlated with the observed changes; but the highest correlations are unreliable.

Concerning the second point as well as several of the comments in the previous section, are four possible limitations in the study design:

1) Time interval of prediction: The inherent predictability may be shorter than 1 year; that is, the processes leading to variation may be chaotic over time intervals as great as one year. This comment particularly applies to the residual component.

2) Choice of inputs: Maybe other factors affect military personnel in ways not represented by the present inputs. We have already encountered this problem in the earlier discussion of the MID onset dates. These other factors might be unknown, or suspected but unavailable from historical sources, or not yet developed into data. A fair number of additional indicators are already available, however, in machine-readable data sets or (like the MID data) in forms adaptable from print sources; perhaps some of them need to be included.

In addition to study resource limitations, a reason for limiting the number of inputs was concern with the possibility of “over specialization,” meaning the network simply memorizes answers to the specific cases. My fear was that use of too many variables might produce such a result. (As a matter of algebra, with n many cases, one can always obtain a perfect fit using n many input variables, but this has no substantive meaning-- the meaningful accomplishment is to get a fit with far fewer variables than cases.)

3) Number of training cases: Roughly 150 cases-- the years of experience of the USA over the study period --seems a bit small for neural networks to be successfully trained.

4) Specificity of inputs: The foreign inputs were summations of individual actor experience, which may be too indiscriminate. Perhaps the network needs to see the actions and relationships of foreign parties individually. For instance, United States military personnel changes in response to the actions of the United Kingdom, a friendly state, may be quite different from its response to Nazi Germany or Soviet Russia; and the response may be some compound of situations differentially describing friend, foe, and neutral. One reason for lumping the foreign data was concern with the possibility of over specialization mentioned in item 2, above.

To close the discussion, corresponding to each of these possibilities let us see what steps likely will be taken next in the project:

- Time interval. Very little can be done here without entirely new data development, as the capability data lack discrimination below 1 year. (MID and other discrete data often do have a finer definition, however.)

- Choice of inputs. In the next stage, the MID onset date problem will be corrected for the major nations.

- Number of training cases: The set will be expanded next to include all of the G-7 (Canada, France, Germany, Italy, Japan, United Kingdom, USA) plus Russia--providing roughly 1,000 cases. Later on, we can expand to the entire collection of sovereign state system members. This set comprises approximately 10,000 nation-year cases of data, of which perhaps 5000 will be available after the 20-year run-up to the first change component estimates.

- Specificity of inputs: With more cases, it will be possible to use a greater number of more specific inputs coding for political friendship or hostility, etc., involving the subject nation with other named parties.

Appendix 1. How Social Prediction May Be Aided By Neural Network Modeling

(Note: an earlier version of this appendix previously appeared as part of a paper by Williamson and Karasik, 1994.)

This appendix further describes the rationale for using artificial neural network modeling techniques to help predict the behavior of social systems, in terms of three requirements of social system prediction: to

- represent system complexity
- describe observed phenomena reproducibly and numerically
- reflect change over time.

The first requirement concerns complexity. Social systems have large numbers of different human and non-human phenomena, factors, processes, variables, etc. that mutually affect each other in a variety of ways, typically understood poorly or not at all. The situation is illustrated by Figure 3, which depicts the relationships among a variety of actual and potential problems of world wide scope.

Figure 3. Possible Couplings Among Problem Groups

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
ozone depletion	1						•				•								
CO_2 pollution	2										•							•	
ground / water pollution	3				•		•				•								
deforestation	4		•			•					•								•
desertification	5																	•	•
species extinction	6										•							•	
monetary instability	7							•						•	•		•		•
local war	8								•	•	•	•	•	•	•				•
armaments	9								•	•	•		•		•				•
disinvestment	10	•	•	•				•						•			•		•
disease	11										•	•		•	•	•	•	•	•
instability of global peace	12	+	+	+	+		+	+	+	•	+	+	+	+	•	+		+	+
urban disintegration	13								•		•	•		•	•		•		•
authoritarianism	14								•					•					
population instability	15	•	•	•	•	•	•	•	•		•	•		•	•	•	•	•	•
illiteracy	16										•	•		•	•	•	•	•	•
malnutrition / starvation	17				•		•	•			•	•		•		•	•		•
poverty	18													•	•	•	•	•	•
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

Key: Cell entries represent couplings between row problem group and column group. Number of given column denotes group named in the same numbered row.
• represents the presumption of continuous strong coupling.
\+ represents presumption of contingent strong coupling-- actual coupling only in event of global war.

In this figure, each row names one such problem and tags it with a number; these same problems are also tagged by the same numbers, respectively, one by one in each column. An entry (• or +) in the cell formed by the intersection of a given row and column indicates the row problem arguably has a significant influence on the column problem-- that is, on a second problem (or on itself, in the case of an entry on the diagonal of the figure). By finding the row which names this second problem, then locating an entry in one of *its* cells, one can trace the influence of this problem on a third problem (named in the corresponding column); and on a fourth problem; etc. In this way, the affects of the original problem are seen to bounce around indirectly to other problems, over indefinitely many steps. In the figure, all problems thus directly or indirectly (via intervening problems) influence all others. If we were to widen our attention to all the factors, variables, and processes underlying the problems, then an even more complex pattern composed of exceedingly many more connections would be seen to exist. The first requirement of effective social prediction is to be able to identify and represent this complex pattern of direct and indirect influences among problems, factors, variables, and processes of the global social system; that is, prediction must identify and represent the <u>couplings among the factors</u>.

Neural network modeling is able to receive and combine indefinitely many possible factors as inputs. These inputs can be 0's or 1's , to represent the presence of absence of qualitative conditions, and they can be rational numbers, to represent the direction and amount of a factor. No presumption is required about how to compare incommensurate things ("apples and oranges"); the process of developing ("training") a neural network supplies the conversion constants among disparate social, political, economic, biological, and environmental factors. These conversion constants also function as weights for combining inputs into outputs which serve as predictions; and this process of combination represents the couplings among the factors.

The second problem is the need to use numerical data that realistically describe the social system one seeks to predict. Numbers are important for two reasons. First, if the numbers are developed from facts according to an explicit recipe they will objectively depict the factual situation that is to be modeled. This objectivity consists in the ability of any two persons to derive the same values of data if both use the same recipe to generate them. (This quality is called "reproducibility".) Second, numbers convey much more information than words-- they convey information about quantity (how much, how little) as well as quality. This added information is used by a mathematical model to help make predictions. (Of course, the added information may be false; maybe the precision of numbers is not warranted by what is realistically known of the particular situation being modeled. Avoiding or identifying such false precision is an essential part of mathematical modeling, but in essence the various ways to separate true from false precision necessarily include trying various models to find what does and does not yield accurate predictions.)

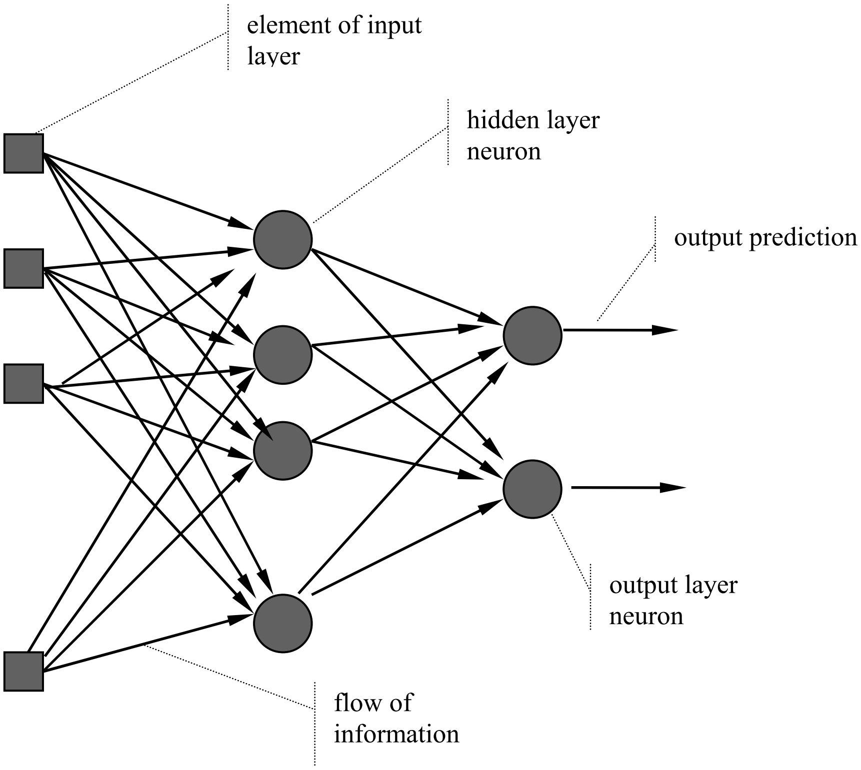
The third problem concerns change over time. The essence of prediction is it captures how something will change as a result of present or past conditions. A computational procedure that is able to represent such change is called a *dynamic model*. Normally, the way one develops a dynamic model is to pick functions (formulae) of the input data that reflect one's belief ("theory") about what mechanisms, causes, processes, etc. are at work to produce the change; one then tests the model by trying it out on historical data, using the functions chosen. (Some details of the functions, called "adjustable parameters", may be selected or narrowed in the process of testing.) If the result is a sufficiently accurate fit between predicted and observed changes, the model is considered to be successfully tested. In social prediction, the problem is that frequently there are no uniquely good ideas about what functions to use in the first place, because there are so many equally plausible ones from which to pick. Also, there may be a great many quite different functions variously appropriate to the various distinct aspects-- human, societal, technological, environmental --of the system being modeled. Here, the advantage of neural networks is they evolve a representation of a best fitting function as they are trained; so it is not necessary to begin with a specific function in mind.

(Note the functions themselves remain implicit or at least well-hidden in the patterns of connections and weights the elements of the network have finally assumed, upon repeated training; so the problem of functional relationships remains to be solved, even though that problem may no longer constitute a barrier to effective prediction. In fact, once a network has been trained, it can be used as a device for simulating relationships between inputs and outputs under simplified, extreme, or other hypothetical circumstances for which historical data are lacking; the results of such simulations may provide clues about the functional relationships themselves and about the underlying causes or other explanations. One might say the artificial neural network may act as a measuring instrument yielding additional useful "data," albeit of an extensively transformed character, in comparison with the original empirical data.)

Since a dynamic model seeks to emulate a real system, it is important that it be trained on data representing events and conditions in the real historical past (at least, as "real" as record keeping and historiography can make it). Since the model seeks to predict change over time, a further essential aspect of its development is that it be trained on data which describe how those events and conditions occurred and changed over time.

A variety of neural network devices are described in Hammerstrom (1993), Haykin (1994), and Rich and Knight (1991, chapter 18). One type, known as a "perceptron", is shown in Figure 4.

Figure 4. Feed Forward Neural Network



A perceptron uses data on past events from which to be trained, which is congruent with the substantive requirement that global models be based on past events. Such a device is composed of "input" and "output layers", and one "hidden layer". It works as follows:

The input layer consists of a series of components, each of which is simply a counter holding the numerical value characterizing one of the pieces of input information.

The hidden layer consists of a second set of components. Let j refer to the jth such hidden layer component. Each such component first computes a value s(j) which is a weighted sum of the values of the input layer, minus a constant reference value: i.e., if x(i) is the value of the ith input component, w1(j,i) is the weight applied to x(i), and w1(j,0) is the reference value, then we define

$$s(j) = \sum_i w1(j,i)x(i) - w1(j,0) \ , \qquad (1.1)$$

across all values of i. Second, the jth component computes a new number based on a sigmoid (inverse logistic) function the argument of which is s(j). The value of this new function, defined as

$$h(j) = 1/[1+e^{-s(j)}] \ , \qquad (1.2)$$

evidently approaches 0 as s(j) becomes more strongly negative and 1 as s(j) becomes more strongly positive, and h(j) = 1/2 for s(j) = 0.

Finally, the output layer consists of a third set of components corresponding to each indicator we seek to predict. (In the present study, there was one such component in each application of the model.) Everything described in the hidden layer has a corresponding element in the output layer: A weighted sum defined as

$$y(k) = \sum_j [w2(k,j)h(j)] - w2(k,0), \qquad (1.3)$$

is computed across all values of j. Then the sigmoid function

$$o(k) = 1/[1+e^{-y(k)}] \ , \qquad (1.4)$$

is computed.

The criterion of the model is: find model w-values w1(j,i) , w1(j,0) , w2(k,j) , and w2(k,0) for all i, j and k such that o(k) equals the observed value corresponding to that output, to an acceptable approximation. These w-values are referred to as “network weights.”

The question then becomes, can one find such numbers? Repeated empirical investigation as well as mathematical analysis suggests that under a very wide set of circumstances the answer will turn out to be in the affirmative, provided one is able to subject the model to a process of calibration called “training”. In training, the initial model w-values are set by some plausible guess (possibly at “random”) and the corresponding outputs are compared with the empirically observed values (criterion values) of a known outcome; then the model w-values are changed as a function of the error vector of output minus criterion values. When this process is repeated using additional criterion values from other known outcomes, the resulting error vectors typically become smaller. That is, the more outcomes known and used in training the model, the more accurate it will become. This technique of making changes in the w-value weights is called “back propagation”.

To emphasize the connection between the above and the earlier discussion: the process just described is a general means for establishing the couplings; the inputs to a neural network are the row entries in coupling matrices such as the matrix of Figure 3; a single output is one of the given column entries; the network itself is the device (corresponding the given column itself) for taking all the row factor antecedents and translating them into column consequence. Of course, as indicated previously and as illustrated for just one fragment of the modeling problem by the work reported in this paper, realistic coupling matrices will be exceedingly larger, more detailed, than Figure 3.

Appendix 2. A Component Approach To Modeling Capability Changes

The following shows the breakdown of capability change into separate "change components". The term "referent capability" means whatever capability one is seeking to predict; in this study, military personnel. We begin with some notation: Let

$t =$ year following the reference year
$m_t =$ slope of trend line in year t
$b_t =$ y-intercept of trend line in year t
$e_t =$ residual of regression estimate of the logarithm of capability in year t
$y_t =$ true logarithmic value of the referent capability in year t
y□ $=$ estimate of y_t based on the trend line version for year t
$x_t =$ independent variable of trend line

where reference is to a trend line determined by least squares bivariate regression of y_t , the logarithm of the referent capability, on x_t . Now, the independent variable x_t actually is t in this approach; however for visual clarity I keep the indicated distinct notation through most of the discussion. However we do make use of this identity to note

$$x_{t-1} = x_t - 1 \quad (2.1);$$

i.e. the independent variable of the prior year is simply the prior year. Note that $t-1$ is the "reference year" in this appendix only; elsewhere the text adheres to the convention that the reference year be denoted by t .

Next define

$$\begin{aligned} \Delta m_t &= m_t - m_{t-1} \\ \Delta b_t &= b_t - b_{t-1} \\ \Delta y_t &= y_t - y_{t-1} \\ \Delta x_t &= x_t - x_{t-1} \quad (2.2). \end{aligned}$$

So the delta's are the increments of the respective values from the prior year. For clarity I include notation for the change in the independent variable, even though $\Delta x_t = 1$ always holds, from (2.1). (In this appendix, numbers within parentheses in text will refer to equation numbers.)

As explained in the main text, the trend line has differing versions for the years t and $t-1$. The equations defining the two versions are

$$\begin{aligned} y_t &= m_t x_t + b_t + e_t \\ y_{t-1} &= m_{t-1} x_{t-1} + b_{t-1} + e_{t-1} \end{aligned} \quad (2.3).$$

From (2.2) and (2.3) we have

$$\Delta y_t = m_t x_t - m_{t-1} x_{t-1} + b_t - b_{t-1} + e_t - e_{t-1}$$

which, after rearrangements and substitutions becomes

$$\Delta y_t = \Delta m_t + m_{t-1} + \Delta m_t x_{t-1} + \Delta b_t + e_t - e_{t-1} \quad (2.4).$$

Now we want to evaluate the y-intercept increment Δb_t . To do so, we need to talk about the method for estimating the slope and y-intercept for a given year t . As described in the main text, this estimate involves fitting the trend line over the series of case years preceding the year to be predicted, starting with some initial year and ending with the reference year. For visual clarity, also we will temporarily omit the subscript t ; so y_t and x_t will be denoted by y and x , respectively in what immediately follows. Let

$N =$ number of case years
i be a subscript which denotes y and x values in the i^{th} case year, $i = 1,2,\ldots N$
$\bar{y} =$ arithmetic mean of the y_i
$\bar{x} =$ arithmetic mean of the x_i .

When the trend line is fit to these N cases, we get an expression exactly like either of equations (2.3), except the case year subscript i replaces the subscript t or t - 1 . When the trend line equation, so modified, is expressed in mean deviation form, the equation becomes

$$y_i - \bar{y} = m(x_i - \bar{x}) + e_i \ ;$$

that is, the intercept term drops out. By rearranging this expression to its equivalent

$$y_i = mx_i + (\bar{y} - m\bar{x}) + e_i$$

and comparing the latter with either of equations (2.3) (using i in place of t or t - 1), evidently the y-intercept must be given by

$$b_t = \bar{y}_t - m_t \bar{x}_t \quad (2.4),$$

where, from this juncture on, we restore the subscript t. From (2.4), we can see the change in y-intercept is given by

$$\Delta b_t = \Delta \bar{y}_t - \Delta(m\bar{x})_t \quad (2.5),$$

where the final delta expression on the right denotes the change from t - 1 to t of the quantity within the parentheses.

Continuing hot on the trail, toward evaluation of the y-intercept increment in (2.4) -- let us focus on the right hand terms of (2.5), starting with the second. Definition of the delta expression, insertion of various equalities, and rearrangement, gives

$$\begin{aligned} \Delta(m\bar{x})_t &= m_t \bar{x}_t - m_{t-1}\bar{x}_{t-1} \\ &= (m_{t-1} + \Delta m_t)(\bar{x}_{t-1} + \Delta x_t) - m_{t-1}\bar{x}_{t-1} \\ &= \bar{x}_{t-1}\Delta m_t + m_{t-1}\Delta \bar{x}_t + \Delta m_t \Delta x_t \end{aligned} \quad (2.6).$$

From their definitions, the increments in mean values can be expressed as

$$\Delta \bar{y} = (1/N)[\sum_{i=t-N+1}^{t} y_i - \sum_{i=t-N}^{t-1} y_i] = (1/N)[y_t - y_{t-N}] \quad (2.7)$$

and

$$\Delta\bar{x}_t = (1/N)[\sum_{i=t-N+1}^{t} x_i - \sum_{i=t-N}^{t-1} x_i]$$
$$= (1/N)\sum_{i=t-N}^{t-1}[(x_i+1)-x_i] = 1 \quad (2.8).$$

Using (2.1) and substituting (2.8) into (2.6) gives

$$\Delta(m\bar{x})_t = \bar{x}_{t-1}\Delta m_t + m_{t-1} + \Delta m_t \quad (2.9)$$

from which we get what we have been after,

$$\Delta b_t = (1/N)(y_t - y_{t-N}) - \Delta m_t \bar{x}_{t-1} - \Delta m_t - m_{t-1} \quad (2.10).$$

(Note the second and third terms on the right demonstrate the negative correlation between intercept increments and slope increments.)

Finally, we can substitute the right hand side of (2.10) into (2.4) to get

$$\Delta y_t = \Delta m_t(x_{t-1} - \bar{x}_{t-1}) + (1/N)\{y_t - y_{t-N}\} + e_t - e_{t-1} \quad (2.11).$$

Equation (2.11) is the desired form of the change Δy_t in logarithmic capability value. Three quantities-- the slope increment, the difference quantity within the curly braces, and the residual for year t -- are indexed to the year of prediction t. These three quantities, thus, are the ones to be predicted by neural networks using inputs known in the reference year t - 1 (expressed in the notation of this appendix). All other quantities refer to the reference year, thus are presumed known through prior observation when we evaluate Δy_t.

In the present study $N = 20$. Suppose the first year of the series over which the trend line is estimated was to have a value $x_1 = 1$. Then the value of the final year of the series (i.e. the reference year) would be $x_{t-1} = x_{20} = 20$ and the average value would equal $[20(20+1)/2]/20 = 10.5$. Thus $x_{t-1} - \bar{x}_{t-1} = 9.5$. From equation (2.8), if we increment the reference year by 1, we increment the average value by the same amount; so, incrementing in 1 year steps to the actual reference year (the first was 1836) shows the difference will remain unchanged. Thus in the present study, (2.11) can be written as

$$\Delta y_t = \Delta m_t(9.5) + (1/20)\{y_t - y_{t-N}\} + e_t - e_{t-1} \quad (2.12).$$

I refer to the quantity within curly braces as the “long-term shift”. Due to typographic limitations, I used entirely Latin characters for notation while working in the computational programs; these are used also in the rest of this report. The correspondence between the two of them for the case of military personnel is given by:

$\Delta m_t \leftrightarrow$ dm_pr
$\{y_t - y_{t-N}\} \leftrightarrow$ s_pr
$e_t \leftrightarrow$ r_pr .

Appendix 3. Explanation Of Abbreviations And Symbols.

Note: indicator values describe the reference year (see main text) unless otherwise stated.

Abbreviation	Meaning
cap	3 cap vars (= pr, ng, tp) x 3 change components each (m_... , r_... , s_...), current and d[...] values.
csxt	Sum, non-USA majors, militarized interstate dispute entries at all maximum levels.
csxt-[n]	[n] = 1, 2 or 3; value of csxt in nth year preceding reference year.
cs1xt	Sum, non-USA majors, militarized interstate dispute entries at maximum level short of armed clash or war.
cs2xt	Sum, non-USA majors, MID entries at maximum level of armed clash or war.
cs2xt-[n]	[n] = 1, 2 or 3; value of cs2xt in nth year preceding reference year.
cst	USA conflict starts: militarized interstate disputes begun plus domestic civil wars begun.
cst-[n]	[n] = 1, 2, ... 8 ; value of cst in nth year preceding reference year.
cus_1	USA conflict vars: wuw , d_wuw , cst , cst-1 ... -3 .
cus_2	= cus_1 + USA conflict: cst-4 ... -8 .
d[...]	Denotes change from prior year of quantity within brackets.
d_wuw	Change in wuw from prior year.
dir_[n]	[n] = 1 ... 4: 1st through 4th sets of inputs, direct analysis; see notes 6, 7, 10, 12 to Table D.
dl_pr	Increment in log military personnel, observed or projected by direct neural network analysis.
dm_pr	Change in m_pr from prior year.
dpr_xt	Change from prior year, log total military personnel non-USA majors (scaled by 10).
dr*_pr	Change in r*_pr from prior year.
ext_1	Sum, non-USA majors: dpr_xt, cs1xt, cs2xt, cs2xt-1 ... 3 (from usa_ext.xls).
ext_2	Sum, non-USA majors: dpr_xt, csxt, csxt-1 ... 3 (from usa_ext.xls).
m_[...]	Denotes slope change-component of capability indicator within brackets.
m_pr	Slope of trend line of pr , based on least squares regression over reference and previous 19 yrs.
MID	Militarized interstate dispute.
net_pr	Projected increment in log military personnel, using projected change component increments.
ng	USA log energy consumption (x scaling factor); as suffix, ..._ng , tags its own change components.
pr	USA log military personnel (x scaling factor); as suffix, ..._pr , tags its own change components.
r_[...]	Denotes residual change-component of capability indicator within brackets.
r_pr	Residual = observed value of pr minus regression estimate (leading to m_pr and s_pr).
r*_pr	Residual in log military personnel after subtracting estimate based on dm_pr and s_pr , from observed dl_pr .
s_[...]	Denotes long-term shift change-component of capability indicator within brackets.
s_pr	"Long-term shift": independent part of intercept regression (leading to slope m_pr).
s1_pr	"Pattern" (the var. to be predicted) was first transformed by sigmoid function prior to training.
tp	USA log total population (x scaling factor); as suffix, ..._tp , tags its own change components.
wuw	Sum of USA international and domestic civil wars underway, reference year.
... (t)	Denotes reference yr.: base year of inputs for projections; = year of data unless otherwise stated.
... (t+1)	Denotes year following reference year = year for which projections are attempted.

Appendix 4. Neural Network Data And Training Steps

The study used a commercially available software package called Brainmaker, marketed by California Scientific Software, which operates as a perceptron (Appendix 1). The Brainmaker software does 3 things: First, it trains a neural network to predict numerical outcomes, based on known input information. This is done by having the network guess the outcome to a question for which the true outcome is already known by the human user; then the network adjusts itself using the back propagation technique. Examples are automatically submitted to the network and adjustments made, until it is judged by the user to be as accurate as possible (according to criteria such as item 3, next paragraph). Second, the software then tests the accuracy of the trained network by giving it input information on new examples not previously seen, and checking the accuracy of the output. Third, the software uses a trained, tested network to predict outcomes for which the user knows only the inputs.

The training step works as follows. 1) A list of cases is submitted to the Brainmaker program. Each case consists of the values of inputs and one or more outputs (which the network is to be trained to "correctly" guess). In the present study, only one output variable was presented to each training session. 2) Some of the submitted cases are reserved; that is, they are not used for training but instead are used later, in the testing step, to present the trained network with cases not previously seen. 3) The computer picks the first unreserved case, uses the input values to predict the output value, compares the latter with the true known output value to produce an error value E, and adjusts the (initially arbitrary or random) weights according to the back propagation algorithm, which is a function of E. The computer then picks the next case on the list and repeats the procedure, using the adjusted weights from the prior case. When all unreserved cases have been so used, the procedure either returns to the first such case and again makes its way through the list, or the procedure terminates. Termination occurs if the "error rate" drops below a pre-selected criterion value or if a pre-selected criterion maximum number of iterations of the above training process has been reached. The error rate is defined as follows. The list of known training outputs has a maximum true value and a minimum true value. Let D denote the difference between these two values and let E_{max} denote the maximum absolute value of error E among all the cases. Then the error rate = E_{max} / D . (This criterion is the operative meaning of "correctly" guessing all the outputs.) In this study, the error rate criterion usually was set to 0.2 but occasionally to 0.15 or 0.1. 4) Often, the minimum error variance occurs for some set of weights computed prior to the final set at termination. In that event, the network may be "re-trained" by backing up to an earlier interation and stopping at the point of minimum variance. Re-training was employed for most applications in this study.

In the change component method, the following were presented to the training procedure: For each reference year, each of 3 indicators-- total population, energy consumption, and military personnel --was presented in terms of its three change elements; and for each change element the value for that reference year and the amount of change from the prior year were both presented. Thus a total of $3 \times 3 \times 2 = 18$ candidate predictor variables were input for each case. The cases themselves consisted of the reference years 1837 through 1990.

The external data consist of summations of conflict involvement instances and military personnel changes across the major powers, the rationale being that the behavior of these parties was most likely to have exerted influence on the military personnel changes of the United States. Exactly which parties are on the list of "major powers" is conventional, even if major status is materially relevant. In this study I began with the list developed by J. David Singer's Correlates of War (COW) project at the University of Michigan, which list I modified somewhat. States classed by that project as "major powers" at one time or another during the study period include Austria-Hungary, China, France, Germany, Italy, Japan, Russia, United Kingdom, and the United States. My purpose was to characterize a group of parties exhibiting actions that plausibly may have exerted a continuing influence on the USA; thus I removed the United States itself. Also, I removed China due to missing data in the earlier part of the study period plus my judgment that the impact of that state probably was negligible prior to 1945; in addition, for military personnel, was the fact that Chinese numbers would be very large and probably less reliable, compared with the other states, which would have obscured the contributions of the latter. (The COW project data for pre- Second World War Germany and Italy include data for states judged to be their respective predecessors, Prussia and Sardinia, which data I retained.) Finally, I included the experiences of post-Second World War Japan and West Germany; and I included data for all years in which the states were in existence, regardless whether they were classed as holding major power status in the particular year.

These various inputs together comprise what is tagged as the "cap" data set in Table 5, below, and elsewhere in the report. As explained in the main text, summaries of USA and foreign conflict involvement and of changes in foreign military personnel also were input in later sessions.

Notice that the criterion output indicator (the value of which we are trying to predict for the reference year + 1) also supplies one of the inputs in the reference year (the values of which we presume already to know). For instance, the output labeled dm_pr+1 has the value 3.46 in 1837 and the input labeled dm_pr has the value 4.25 in the same year; but dm_pr+1 is defined as the value of dm_pr for the following year 1838. This aspect simply reflects the reasonable possibility that a variable in year t + 1 is partly a function of itself in the year t .

The total collection of studies is summarized in Table 5 at the end of this appendix. One can see that some of the studies use alternative approaches to modeling capability change, such as the direct approach. One of the approaches tested the affect of compressing the range of outputs from $(+\infty, -\infty)$ to $(+1, 0)$. These are referred to in the notes to Table 5 as s-transformed data.

Reading across the columns, we see indicated: 1) the output variable to be predicted; 2) for the test data, product moment correlation (test r) between output values projected by the network and "observed" values (i.e. values inferred from observed data); 3) for the training data, correlation (train-test r) between projected and observed outputs, if any; 4) input data sets (see Appendix 3); 5) name of the input and criterion output file (in my computer); 6) name of the test output file; 7) name of the train-test output file, if any; 8) error rate criterion (see earlier in this appendix); 9) run number at which the error rate criterion was met; 10) indication whether initial network connections were re-randomized and noise used in the session (techniques possibly improving network accuracy); 11) starting year of the data series; and 12) notes. (I omit the resulting network weights. Difficult to interpret in any case, at this early stage they simply are too provisional to warrant discussion. I would be happy to provide a listing of them to any interested reader.) Item 2 is key, it being a measure of the accuracy when the trained network was applied to the previously unseen or "reserved" cases withheld from the network during training. The computer file names (items 5 - 7) were included to be possibly helpful in answering later questions. A training "run" is one complete cycle through the training data; "run number" means the number of complete runs executed at a given juncture of a training session.

The data-- inputs and criterion output --were processed by the neural network training program in the manner described above, with results summarized in Table 5 and in the main text. In these analyses, 10% of cases were reserved for testing. Unless otherwise stated the error criterion was 0.2 . (This was used, following a suggestion in the Brainmaker literature for relatively few cases involving social data, until it was discovered that lower values would produce more accurate results.)

Tables 5 through 7 and notes follow, next three pages.

Table 5. Summary of studies, USA log military personnel increments.

output	test r	train -test r	inputs	input file	test output file	train - test output file	err rate critn if not 0.2	crit met run #	retrain to run #	cnctn rdmz, noise	st yr if not 1837	notes
dm_pr	0.93		cap	b6_g01_a	b6_g01p1			111	80			1
dm_pr	0.92		cap, cus_1	b7_g01_a	b7_g01p1			229	186			
dm_pr	0.98	0.89	cap, cus_2	b7_g01_b	b7_g01p2	b7_g01p4		151	91			
dm_pr	0.98		cap, cus_1, ext_1	b7_g01_c	b7_g01p3		0.15	416	393	x		
r_pr	0.89		cap	b6_g02_a	b6_g02p1			214	140			
r_pr	0.88		cap, cus_1	b7_g02_a	b7_g02p1			84	17			
r_pr	0.85		cap, cus_2	b7_g02_b	b7_g02p2			40	36			
r_pr	0.93		cap, cus_1, ext_1	b7_g02_c	b7_g02p3			66	56			
r_pr	0.98	0.91	cap, cus_1, ext_1	b7_g02_c	b7_g02p4	b7_g02p5	0.10	203	55	x		
r_pr	0.95		cap, cus_1, ext_2	b7_g02_f	b7_g02p6		0.10	446	410	x		
s_pr	0.90		cap	b6_g04_a	b6_g04p1			276	60			
s_pr	0.95		cap, cus_1	b7_g04_a	b7_g04p1			65	34			
s_pr	0.96		cap, cus_2	b7_g04_b	b7_g04p2			30	19			
s_pr	0.94		cap, cus_1, ext_1	b7_g04_c	b7_g04p3		0.15	80	none	x		
s_pr	0.99	0.95	cap, cus_1, ext_1	b7_g04_d	b7_g04p4	b7_g04p5	0.10	470	143	x		
s_pr	0.98		cap, cus_1, ext_2	b7_g04_f	b7_g04p6		0.10	453	123	x		
net_pr	-0.10		3 compnts	= output	b6_g05p1		none					2, 3
net_pr	-0.03		3 compnts	= output	b7_g05p2/2		none					2, 4
net_pr	0.70	0.74	3 compnts	= output	b7_g05p4	b7_g05p5	none					2, 5
net_pr	0.38		3 compnts	= output	b7_g05p6		none					2, 16
dl_pr	-0.31		dir_1	b6_g06_a	b6_g06p1			499	20		1818	6
dl_pr	-0.17		dir_2	b6_g07_a	b6_g07p1			395	320			7
dl_pr	0.75		dir_1, cus_1	b7_g06_a	b7_g06p1			305	499		1820	8, 9
dl_pr	0.86		dir_1, cus_1	b7_g06_a	b7_g06p2		0.15	none	419		1820	9
dl_pr	0.35		dir_1, cus_2	b7_g06_c	b7_g06p3		0.15	152	60		1825	
dl_pr	0.25		dir_1, cus_1 , ext_1	b7_g06_d	b7_g06p4		0.15	222	46		1825	
dl_pr	0.21		dir_3	b7_g06_f	b7_g06p5		0.15	318	none		1825	10
dl_pr	0.92	0.47	dir_1, cus_1	b7_g06_a	b7_g06p8	b7_g06p9	0.15	none	187	x	1820	9 , 11
s1_pr	0.36		dir_4	b7_g06_h	b7_g06p6		0.15	314	107		1825	12
r*_pr	0.18		cap, cus_1, ext_1	b7_g08_a	b7_g08p2			40	7	x		13
r*_pr	0.56		res_1	b7_g08_b	b7_g08p3		0.15	147	28	x	1839	14
dl_pr	-0.50		r*_pr, dm_pr, s_pr	= output	b7_g08p3		none					15

Notes to Table 5
1. Input and output file references are to the labeling system in author's computer files.
2. Total change log military personnel projected by summing neural net projected change components.
3. Uses projections from output files b6_g01p1 , b6_g02p1 and b6_g04p1 .
4. Uses projections from output files b7_g01p2 , b7_g02p3 and b7_g04p4 .
5. Uses projections from output files b7_g01p2 , b7_g02p4 and b7_g04p4 .
6. Inputs are 0th, 1st, 2nd order increments of l_pr , l_ng , l_tp , plus logs of military expenditure and total population.
7. Inputs are cap plus l_pr , current and delta values.
8. Training resumed beyond criterion due to continued decline of training RMS error; but note discontinuity.
9. Stopped at run #500.
10. Inputs = dir_1, cus_1 , ext_1 less exclusions indicated in b7_g06_e .
11. Done to check result in b7_g06p2 .
12. s-transformed; training inputs = dir_1, cus_1 + s1_pr from b7_g06_g .
13. Estimate of residual of pr increment less contributions from dm_pr and s_pr .
14. Using b7_g08_a except substitute r*_pr(t), dr*_(t) inputs in place of r_pr(t) , dr_pr(t) .
15. Observed, projected values of dl_pr using data from b7_g08p1 / test data.

16. Uses projections from output files b7_g01p2 , b7_g02p6 , and b7_g04p6 .

Table 6. Direct Approach
Test Case Outputs (Values of Figure 1)

Changes in Logarithm USA Military Personnel, **dl_pr (t+1)**
Notes: see Table 7

case	year	proj	obs	error
9	1825	0.09	0.04	0.05
19	1835	-0.01	0.09	-0.10
29	1845	0.07	0.29	-0.22
39	1855	0.12	0.00	0.12
49	1865	-0.70	-1.16	0.46
59	1875	0.13	0.03	0.10
69	1885	0.12	-0.01	0.13
79	1895	0.20	0.03	0.16
89	1905	0.03	0.02	0.02
99	1915	-0.10	0.01	-0.11
109	1925	0.02	-0.00	0.03
119	1935	0.03	0.06	-0.03
129	1945	-0.45	-0.62	0.17
139	1955	0.05	-0.02	0.08
149	1965	0.15	0.10	0.06
159	1975	-0.13	-0.01	-0.12
169	1985	0.10	-0.00	0.11

Table 7. Components Approach
Test Case Outputs (Values of Figure 2)

Changes in Components and Logarithm USA Military Personnel

		dm_pr (t+1) :			r_pr (t+1) :			s_pr (t+1):			net_pr (t+1):	
case	year	proj	obs	err	proj	obs	err	proj	obs	err	proj	obs
26	1842	1.56	0.61	0.95	-0.80	-0.21	-0.58	2.39	3.01	-0.62	-0.09	-0.04
36	1852	-1.71	-4.16	2.44	-1.75	-1.69	-0.06	0.20	2.22	-2.02	0.00	0.00
46	1862	13.25	17.24	-3.99	8.47	9.18	-0.70	13.56	16.80	-3.24	0.03	0.15
56	1872	-9.81	-12.62	2.81	-5.10	-4.37	-0.73	1.97	3.12	-1.15	-0.04	0.01
66	1882	11.72	14.04	-2.31	1.15	2.66	-1.51	-11.06	-14.25	3.19	-0.16	0.00
76	1892	0.44	0.96	-0.53	-0.30	0.17	-0.46	-0.24	-0.33	0.09	-0.03	0.02
86	1902	4.53	2.00	2.52	-0.82	-0.21	-0.61	3.62	4.69	-1.07	-0.06	-0.02
96	1912	-1.95	-3.20	1.25	-0.54	-0.60	0.07	6.69	6.11	0.58	0.03	0.00
106	1922	-3.48	-4.11	0.64	-0.41	-3.60	3.19	6.15	3.68	2.47	0.30	-0.04
116	1932	-0.46	-6.20	5.74	2.05	0.08	1.96	4.36	1.98	2.39	0.26	-0.00
126	1942	15.10	18.09	-2.98	7.68	8.51	-0.83	15.14	15.64	-0.50	0.26	0.37
136	1952	-4.46	-5.10	0.64	-2.62	-2.29	-0.33	10.63	11.32	-0.69	-0.04	-0.01
146	1962	5.37	6.69	-1.32	0.72	0.82	-0.10	-2.29	-5.57	3.28	-0.02	-0.02
156	1972	2.82	-0.05	2.88	-0.92	-1.02	0.10	-1.68	-1.75	0.07	0.02	-0.02
166	1982	2.50	-0.68	3.19	1.32	0.48	0.84	0.20	-0.69	0.90	0.12	0.00

Notation and notes for Tables 6 and 7:

case = case number
year = reference year t
proj = value projected by network in year t+1
obs = value observed or computed from, historical data in year t+1
error = proj - obs values

Reference year is **1 year prior to** year to be predicted. (So, for example the 1985 projected refers to the value predicted for 1986.)

Citations

Etcheson, Craig (1989). *Arms Race Theory: Strategy and Structure of Behavior*. Westport, CT: Greenwood Press.

Hammerstrom, Dan (1993). "Neural Networks at Work". *IEEE Spectrum* (June) , pp. 26-32.

Haykin, Simon (1994). *Neural Networks. A Comprehensive Foundation*. New York: Macmillan.

Lawrence, Jeannette and Janell Fredrickson (1993). *The BrainMaker User's Guide and Reference Manual*. 10024 Newtown Road, Nevada City, CA: California Scientific Software.

E. Rich and K. Knight (1991). *Artificial Intelligence*. New York: McGraw Hill.

Williamson, Paul and Myron S. Karasik (1994). "A Neural Network Approach to Modeling Inter- and Intra-State Conflict." Presented at Peace Science Society (International) annual meeting, Urbana, IL.