

Kimberly Kempf-Leonard, ed., *Encyclopedia of Social Measurement* (San Diego, CA: Academic Press, forthcoming 2003)

INTERNATIONAL RELATIONS

Bruce Russett

Yale University

I. Introduction

II. Variables and Units of Measurement

III. Analytical Issues

GLOSSARY

dyad A pair of units, typically countries; may be balanced (same value for both units) or unbalanced.

international system A set of interacting countries.

hierararchy A system that system can be ordered from most powerful to least powerful units.

level of analysis The degree of aggregation, from individual decision makers to international system.

polarity The number of major power centers in a system.

political stability The longevity of a political system or a leader's rule, or the degree to which that rule is challenged.

power The ability to change another's behavior; also the resources that may be used to produce change.

MEASUREMENT IN INTERNATIONAL RELATIONS scholarship is directed to the development of comparable and valid measures of theoretically important concepts according to international standards of reliability and validity, to developing procedures for their analysis, and to making the data available to a wide range of social scientists.

I. INTRODUCTION

The founding of the scientific study of international relations is properly credited to two scholars, Lewis Frye Richardson and Quincy Wright, both working just before the middle of the twentieth century. Both were major theorists, but both also devoted great energy and ingenuity to devising techniques for measuring many aspects of international relations. In the 1960s the U. S. National Science Foundation made substantial financial support available, and a critical mass of researchers in Europe and North America emerged to improve, extend, and utilize this information. Vast improvements in the quality and quantity of information have combined with comparable improvements in theory and statistical analysis to permit the emergence of international relations as a scientific endeavor. Moreover, the widespread availability of electronic databases for sophisticated analysis is producing greater scholarly convergence on measurement techniques—a necessary though not sufficient condition for scientific success.

II. VARIABLES AND UNITS OF MEASUREMENT

What is to be measured of course depends on theory, and theories of international relations operate at several levels of analysis, commonly identified in increasing aggregation as individual decision makers, political unit (i.e., the nation-state, or country), dyads (pairs of countries), and the entire international system. This review concentrates on measurements that can be compiled, with acceptable comparability across countries and years, for most or all countries that are members of the system of international relations.

A. Individual Decision Makers

Policy-relevant individuals may in principle range from individual voters to heads of state and other key persons responsible for making foreign policy choices. Public opinion surveys measure the direction and intensity of voters' preferences, and special-sample surveys are used to measure the preferences of those citizens more directly involved in shaping opinions and policies. Surveys of either sort are possible only in reasonably democratic countries. Even in democracies, the influence of such preferences on decisions actually made is a topic of intense inquiry and debate.

The public writings and pronouncements of heads of state can be studied through systematic content analysis, not only for their preferences but for their affectual content (friendship or hostility), perception of threat, and complexity of cognition. Formerly a very labor-intensive task, such analysis is now usually carried out on electronic databases compiled from government sources and news organizations. Consequently the information base has become much richer, and while computational routines rarely match the subtlety of human coders, the gains in quantity and reliability of coding are substantial.

B. Country

Heads of state typically speak in the name of their countries, so information about their pronouncements and action can be treated as information about the country itself. Country-level information, however, includes much more, notably about the economic, political, and social characteristics and institutions of the country.

Any analysis of international relations must be concerned with power, as the ability to persuade or coerce another actor to do something. The actual ability to affect decisions in this relational sense is difficult to measure, due to the conceptual problems in determining what the other would have done in the absence of coercion or persuasion. At one extreme, if the leader of country A makes a threat so as to deter country B from attacking country C, the mere fact that B subsequently does not necessarily indicate that A exerted power over B. Possibly B's leader had no intention of attacking C anyway. Such problems can be addressed with sophisticated models of strategic actors, analysis of counterfactuals, and if possible with detailed post hoc documentary information on leaders' deliberations.

More subject to systematic analysis, however, is information on the power bases of nation-states. Much of this involves reasonably straightforward use of pre-existing information, often from national censuses or sample surveys. These include levels and growth rates of demographic data (total population, population in the labor force or of military age, literacy and education, health conditions) and economic data (total national income or income per capita, production of various commodities, value and composition of exports and imports, internal and foreign investment, size and composition of government expenditure), and distributional equality

within the population on valued goods such as income. All of these data are produced by government agencies in most countries, and are subjected to scrutiny and revision by international organizations and by economists, demographers, and other scholars. Analysts of international relations also evaluate the cross-national comparability of these information, but are not primarily responsible for their compilation. Similarly, information on military capabilities (quality and quantity of troops under arms, military expenditures, number and types of weapons) are produced by governments—concerning their own countries, and their allies and rivals—and critiqued by scholars.

Much analysis, however, requires measurement of capabilities and actions within conceptual frameworks derived from theories of international relations. Such measures may be derived from standard measures compiled by other disciplines and agents, but do not exist in any off-the-shelf condition. They may need to be aggregated, disaggregated, or recombined with other existing information or information compiled for the purpose. For example, it is not obvious what is the most useful measure of national power. For nuclear deterrence, of course one counts and evaluates nuclear weapons delivery systems. But for a general-purpose measure of a country's power, one needs a composite indicator of military, economic, and demographic power bases. Moreover, if one wishes to compare all types of countries, and over long time-periods, one must take account of the availability and comparability of the measures that make up the composite. Numbers of nuclear weapons are irrelevant before 1945; despite massive research by economists national income data for many countries in the nineteenth century are unreliable; demographic data are also of uneven quality depending on national information-gathering capability.

Beyond these are information on data not so obviously part of standard demographic and economic statistics. One example concerns distribution of services and goods that may affect national power or political stability: relative equality of access to health care, equality of distribution of land and wealth. Some theories of international conflict concern how ethnically homogeneous the country is in terms of linguistic/racial/religious characteristics; heterogeneous countries may be less politically stable, especially if major ethnic minorities overlap across national borders with neighbors. These measures have to be compiled painstakingly by scholars, and evaluated for their reliability and validity.

More directly political measures concern types of political systems, such as relative size of the government in the economy, or its degree of democracy or authoritarianism. The latter requires some degree of judgment, but great advances have been made in compiling reasonably objective and internationally-accepted codings of type of political system that cover all countries over a time-scale approaching two centuries (Jagers and Gurr 1995). Originally compiled by scholars of comparative politics, they have proved to be very important in testing theories about whether democratic countries are less likely to become involved in international conflict. Still other theories assert that politically unstable countries may be more (or less) likely to engage in international conflict than are stable ones; for example, that unstable governments may attempt to divert their populace toward foreign adversaries and away from problems internal to the country. Depending on the conceptualization, stability may be measured either in terms of the longevity of a particular ruler, party, or constitution, or by the degree of internal conflict manifested in violent protest, rebellion, or civil war. Measures of involvement in international conflict include merely diplomatic disputes in the context of peaceful relations as well as the use of military instruments of violence. The latter may range from relatively low-level military

threats up through borders skirmishes to full-scale war (conventionally defined in the scholarly literature as involving at least 1,000 combat deaths). Hardly any of this kind of information is reported on a comparable basis by national governments. Rather, to be made comparable across countries and time-periods it requires detailed compilation from news media, scattered documentation, and historical documents by teams of scholars (Jones, Bremer, and Singer 1996). The temporal unit of analysis may be a day, or aggregated into monthly or annual slices. Many scholars scrutinize these compilations of political data for accuracy and completeness.

On the whole, country-unit analyses of international behavior have not proved very productive. One exception is the convincing evidence that great powers, with wide-ranging political and economic interests combined with the military power to exert force far from the home country, are many times more likely to become involved in violent conflict than are small or weak states. Other than this, however, strong generalizations about the risk factors for violence by particular kinds of states have proved elusive. There is some evidence that democracies are less likely to engage in violent conflict than are autocracies, but that evidence is contested, the relationship is fairly weak, and the class of autocracies is too varied (communist and fascist regimes, military dictatorships, traditional monarchies) for easy generalization.

C. Dyad

A major conceptual breakthrough, focusing on the characteristics of pairs of states, has however contributed to greater success in identifying, both theoretically and empirically, the risk factors of international conflict. It drew on the insight that it was less the characteristics of individual countries than their relationships with other countries that mattered. Most countries

are neither especially peaceful nor very war-prone in general; their readiness to engage in violent conflict varies over time and with regard to particular other countries. One such hypothesis is that countries will be reluctant to fight other countries of comparable power (with whom the outcome of a war might be unpredictable) but readier to fight weaker countries with which they could be much more confident of victory. This perspective, however, leads to a competing hypothesis for war frequency: the rulers of weak countries, knowing that they would likely lose a war, might make great concessions to powerful ones, making war between very weak and very strong countries unnecessary, and rare. Thus the risk of war might be greatest between two countries of comparable power, leading to great uncertainty as to which would win the war and thus to war as an empirical test of relative capability. Of the two competing hypotheses the logic beyond the latter (power equality leads to wars, power differential discourages war) seems stronger, but there can be no substitute for empirical investigation over many dyads and years. Such empirical analyses are now possible, since an international system of, for example, 100 countries, has 4,850 *pairs* of states $(N \times N-1)/2$, each of which can be compared over many time intervals. Thus taking one-year data slices over a period of a century would give a potential for 485,000 observations--enough observations to make possible the statistical testing of competing hypotheses. In turn the hypotheses could be refined by careful strategic analysis, including the application of game theory to questions of how states would behave under conditions of uncertainty about each other's intentions or capability.

Thus dyadic information—typically compiled initially on a country-year (one observation for each year a country is a member of the system) basis and then converted into dyad-year format—permits the testing of hypotheses about a variety of conditions long thought to affect the risk of violent conflict, with enough cases to make statistical generalizations about relative risks.

In effect, violence and war could be considered as cases in an epidemic, and social scientists can investigate hypotheses from competing theoretical perspectives about what factors might promote the onset of violence.

Relevant variables for dyadic analyses include traditional concerns for power and its projection: relative military and economic strength, geographic proximity (contiguous borders, or distance between countries, often transformed into logarithms to represent the rate of decline of power capability over space), and the presence or absence of formal alliance agreements between members of the dyad. Others concern the similarity or difference between political systems (as by subtracting the score for the less democratic country on the democracy-autocracy scale from that of the more democratic one), or the similarity of their behavior internationally (do they ally with more or less the same other countries?), or vote similarly in international organizations (such as the UN General Assembly). Still others concern a wider range of economic and institutional ties between countries: trade in goods and services or investment flows, preferential trade agreements, the number and types of international organizations in which they share membership. All of these have been found to be related—sometimes just correlationally, often with plausible grounds for inferring causation—to the degree to which the members of the dyad engage in a wide range of cooperative or conflictful activities (Russett and Oneal 2001). Cooperation and conflict in turn are measured as merely diplomatic cooperation/conflict or as militarized disputes up to and including war (Goldstein 1992). With large data bases it is now possible to assess the relative risk of conflict associated with each variable, and to begin to untangle the web of causation. The evidence is that both traditional measures of power and measures of political similarity and political-economic linkage make an impact.

It is not just the static levels of power or economic relationships that matter, of course. Some analysts have addressed changes, such as if the risk of conflict is greater when one member of the dyad is becoming more democratic from an autocratic base, or if the trend in mutual trade in the dyad is downward even though the absolute level may remain fairly high. So far these empirical analyses are inconclusive. More productive may be a focus on changes in relative power between potentially antagonistic countries, on the grounds that a narrowing power gap creates larger uncertainty about which country might win a military contest, and hence raises the risk of such a contest. It is especially useful to concentrate the analysis on particular kinds of dyads, namely those with long-standing territorial disputes that periodically erupt into military confrontations (Diehl and Goertz 2000). It is more productive to concentrate on these cases, and on the associated political, economic, and military changes in their relationship.

D. International System

Other efforts have been directed to understanding risk factors according to different kinds of international systems, including both the global international system and regional subsystems. An international system is identified as a group of interacting countries, in which strategic actors make interdependent decisions. For much of history a truly global international system did not exist, as the level of technology to permit a high level of interaction among distant countries was too low. A system of warring state-like units existed in China at least 3500 years ago, however, and even earlier in Mesopotamia. A regional interstate system existed among Mayan city-states in Central America from about 800 BC to 800 AD, with some revival up to the Spanish conquest. But until the conquest, none of them had had any interaction with or even knowledge

of a European system, nor European states of them. Not until the nineteenth century did a sufficiently high level of interaction develop for the concept of a global system to be very useful.

A global system, like a regional system, can be characterized by its degree of hierarchy, or by its polarity. If it were dominated by one great power, a hegemon, it would be unipolar. By contrast, a system dominated by two big powers would be bipolar, and one of three or more great powers would be multipolar. Different theories led to expectations that some kinds of systems were more prone to large-scale international violence than were others; for example that bipolar systems were likely to experience more small wars than were multipolar systems, but fewer big wars directly between the two big states. The cold war era between the United States and the Soviet Union was such a system. The evidence available for generalization, however, was so mixed as to prevent any consensus on relative risks. One problem was the small number of different international systems from which to attempt to make any reliable generalizations. For example, most international systems in the “modern” era (even as dated from the Treaty of Westphalia in 1648) have been multipolar, with the less than 50-year cold war era the major exception. Another was imprecision or lack of consensus on theory. In measuring polarity some analysts would focus attention more on the structure of alliance systems than on the number of great powers; for example, in 1914 there were perhaps as many as eight great powers in the international system, but only two competing alliance configurations of great powers (Britain, France, and Russia vs. Germany, Austria-Hungary, and Italy). Consequently no near-consensus has emerged from the empirical analyses comparing bipolarity with multipolarity.

There is somewhat more agreement that major wars may be less likely to occur under conditions of unipolarity or hegemony, but the empirical base for that generalization is very

limited, perhaps only to the years since the end of the cold war. Moreover, conceptual agreement as to the relevant measures of unipolarity (military power, economic strength, even cultural hegemony) is elusive. More promising may be attention to changes in the relative power of the leading country or alliance system in the system and that of its putative challenger, as on the dyadic level of analysis.

Another and possibly more productive way to conduct analyses at the system level is to ask whether certain political or economic characteristics become more or less common in the system. Just as systemic measures of power concentration are built up from country-level data on the components of national power, other systemic measures can similarly be constructed. For example, are a larger proportion of the member countries democratic in one period or another, is the level of economic and financial interdependence higher or lower, and is the number and strength of international organizations greater? Certainly such systemic characteristics do vary over time. The proportion of democratic countries in the system has been higher since the 1990s than at any previous time in world history, and by many (but not all) measures the level of economic interdependence has also been at an all-time high. This coincides with the increasing interest in the effects of “globalization” on the world political economy, and specifically on constraints that may reduce the incentives to violent conflict. It suggests that system-level changes in, for instance, the proportion of democracies, might have an effect in addition to the effects at the national or dyadic levels. That is, a need to obtain commercial ties and foreign investment might even those constrain governments that were not yet closely tied into the global economy. Or the growth of international norms and institutions for the peaceful settlement of disputes, deriving from democratic practice, might constrain even governments that were not themselves democratic. It would thus be useful to analyze the international system as a set of

feedback relationships among political, economic, and institutional elements and operating at different levels of analysis.

III. ANALYTICAL ISSUES

A. Data Problems

As noted, great progress has been made in creating worldwide information on structural and behavioral phenomena that were not previously well measured, and the process of international scholarly scrutiny of these measures has narrowed the range of subjectivity and cultural bias that may infect such measures. Yet data problems go beyond the matter of crude or imperfect measures of the relevant phenomena, to situations where data are simply missing for a substantial number of units needed for the analysis. Typically the pattern of missing data is not random, but correlated with and caused by economic underdevelopment or political systems that are closed to external scrutiny. Limiting the analysis to countries where data are reported or can be directly compiled risks introducing serious bias into any “sample” which is subjected to analysis. An analysis confined largely to largely democratic countries because of missing data for most autocracies, for example, might find that the small variation in their degree of democracy made little difference in their international behavior, whereas analyzing the full number and range of political types might show great behavioral differences.

Fortunately, the existence of large data bases on many variables, and the creation of new statistical routines for analysis, often can mitigate these problems. If one knows from previous analyses, for example, that country-level data on income per capita, literacy, life expectancy, and

educational achievement are highly correlated with one another, one can estimate the level of any one variable which is missing for a particular country from its levels on the other three measures. Sophisticated computational routines can impute such missing data from a large body of information, and even supply a range for the probable error in estimation (King et al. 2001). And unlike many substantive analyses which require theories about causation, knowledge of correlational patterns is sufficient for this purpose.

Some kinds of selection biases are more insidious. If, for example, one looks at military crises in which a government must decide whether to come to the defense of an attacked ally, one finds many instances where the alliance is not honored. Does this mean alliances serve no purpose in deterring conflict? To reach that conclusion one would have to look at the many situations where the ally is never attacked. A strong, credible military alliance could deter even the threat of an attack, whereas the alliances that are perceived as weak may be the ones that attract attack. For repeated interactions, as in a process from normal peace to diplomatic dispute, up through militarized challenge to war or settlement, one must control for selection bias at every step (Huth and Allee 2003).

B. Conceptual Problems

The data may be well-measured, but if not matched to appropriate concepts they may be useless at best, and misleading at worst. In principle, interval measures are preferable to binary ones, if the equal-interval assumption of monotonicity is reasonable or can be made so by some transformation. In the dyadic context, some measures are inherently balanced; e.g., the distance from A to B is the same as that from B to A. Other dyadic measures, however, are inherently

unbalanced. To measure the importance of trade to a country's economy and political system the trade total should be divided by the country's gross national product. But the same total trade (exports plus imports) between the countries will produce a highly asymmetric measure if the two countries' GNP totals are very disparate. Presumably the political importance of that trade will be much greater for the smaller country. Behavior is often best measured as that of a directed dyad. Rather than measure the mere existence of a militarized dispute between A and B, one should try to identify the initiator of the dispute. From theory, one should expect a small country to initiate a dispute with a big country far less often than vice versa. Use of directed measures is especially appropriate when studying sequential patterns of strategic behavior. Behavior may be measured as a scale, for example a scale of cooperation or of conflict between countries. Some efforts combine the concepts of cooperation and conflict into a single scale, but this is very problematic. Countries that are politically or economically salient to each other often experience rather high levels of both conflict and cooperation; putting the two concepts on one dimension obscures this reality.

C. Rare Events

Many aspects of international relations represent common behaviors, especially many acts of cooperation. Trade volumes between countries may vary greatly over time, but the individual commercial transactions are numerous. Other events, such as conflict behavior, are quite rare events. Militarized disputes, for example, arise infrequently. In most years, only about three percent of dyads in the international system will experience any such disputes, and

only a tenth of those are at war. Popular statistical procedures, such as logistic regression, can sharply underestimate the probability of rare events (King and Zeng 2001).

D. Cross-sectional vs. Cross-temporal Analysis

Most measures in international relations are utilized both cross-sectionally (comparison across countries at the same point in time) and cross-temporally. Measurement error may be serious in either kind of comparison. Many analyses are done with pooled time-series; e.g., panels of the same countries or dyads measured at regularly repeated time intervals. Such analyses often raise difficult problems of independence of observations across space and time that are increasingly being addressed in this discipline. The behavior of France, for example, is in part dependent on that of its allies as well as that of its adversaries. War, and even changes in political system such as the spread of democracy, may be contagious across space. And of course if France was involved in a militarized dispute with Germany in the past year it is much more likely to be so involved this year than with a country with which it was previously at peace. Most statistical packages contain routines to correct for non-independence of observations, but the choice of correction depends on good theory as well as on statistical options. For instance, controlling current disputes by a term for disputes in the preceding year may obscure the effect of theoretically-important variables that raise the risk of a dispute in both years.

In summary, issues of measurement in international relations are inseparable from theoretical issues. The rise of large-scale datasets of quantitative indicators has forced the

refinement and formalization of theories that were initially expressed verbally; in turn mathematical theories have demanded far greater rigor of measurement.

Bibliography

- Diehl, P., and Goertz, G. (2000). *War and Peace in International Rivalry*. University of Michigan Press, Ann Arbor.
- Goldstein., J. S. (1992). A conflict-cooperation scale for international events data. *Journal of Conflict Resolution* 36, 369-85.
- Huth, P., and Allee, T.. (2003). *The Democratic Peace in the Twentieth Century*. New York, Cambridge University Press.
- Jagers, D., and Gurr, T. R. (1995). Tracking democracy's third wave with the Polity III data. *Journal of Peace Research* 32, 469-82.
- Jones, D. M., Bremer, S. A., and Singer, J.D. (1996). Militarized interstate disputes, 1816-1992: rationale, coding rules, and empirical patterns. *Conflict Management and Peace Science* 15, 163-213.
- King, G., Honaker, J., Joseph, A., and Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*
- King, G., and Zeng, L. (2001). Explaining rare events in international relations. *International Organization*
- Russett, B., and Oneal, J. R. (2001). *Triangulating Peace: Democracy, Interdependence, and International Organizations*. W. W. Norton, New York.