An Attempt to Simulate the Outbreak of World War I

Charles F. Hermann; Margaret G. Hermann

The American Political Science Review, Vol. 61, No. 2 (Jun., 1967), 400-416.

Stable URL:
http://links.jstor.org/sici?sici=0003-0554%28196706%2961%3A2%3C400%3AAATSTO%3E2.0.CO%3B2-E

The American Political Science Review is currently published by American Political Science Association.

AN ATTEMPT TO SIMULATE THE OUTBREAK OF WORLD WAR I[1]

CHARLES F. HERMANN
Princeton University

MARGARET G. HERMANN
Educational Testing Service

Political games and simulations are models or representations of particular political systems and their associated processes. They are techniques for reproducing in a simplified form selected aspects of one system, A, in some independent system, A'. Games and simulations have a dynamic quality produced by the complex interaction of properties in the model. This feature enables them to generate states of the system that differ radically from those present originally. The kinds of transformations that may occur between the initial and final states of a simulation or game are difficult to represent by other means, despite a diversity in modeling procedures ranging from verbal descriptions to differential equations. Because of their apparent applicability to many problems of politics, as well as their novelty, games and simulations have been developed in a variety of areas in political science.[2] They have been used in research, instruction, and policy formation. Although the application of these techniques has been increasing, systematic evaluation of their performance is only now beginning. This essay reports one type of evaluation.

The researchers sought to structure a simulation of international politics so it would reproduce features of the political crisis that preceded the beginning of the First World War. Two separate trials or runs of the simulation were performed as a pilot project. With two runs, the data are sufficient only to illustrate what might be done in an expanded research program.[3]

[1] This research was conducted under Contract N123(60530)25875A from Project Michelson, U. S. Naval Ordnance Test Station, China Lake, California. An earlier report on this project was distributed by the contractor as *Studies in Deterrence X: Validation Studies of the Inter-Nation Simulation*, NOTS Technical Paper 3351, December, 1963. The authors wish to acknowledge their indebtedness to Harold Guetzkow, principal investigator and mentor; Thomas W. Milburn, director of Project Michelson; and Robert C. North and his colleagues at the Stanford Studies in Conflict and Integration who generously shared their document collection and data analysis on the outbreak of World War I. The Center of International Studies at Princeton University supported the first author during the preparation of the present article.

[2] In addition to the studies cited elsewhere in this paper, the variety is suggested by the following illustrations: Oliver Benson, "A Simple Diplomatic Game," in James A. Rosenau (ed.), *International Politics and Foreign Policy* (New York: Free Press, 1961), 504–511; William P. Davison, "A Public Opinion Game," *Public Opinion Quarterly*, 25 (1961), 210–220; Robert P. Abelson and Alex Bernstein, "A Computer Simulation Model of Community Referendum Controversies," *Public Opinion Quarterly*, 27 (1963), 93–122; Lincoln P. Bloomfield and Barton Whaley, "The Political-Military Exercise: A Progress Report," *Orbis*, 8 (1965), 854–870; Andrew M. Scott with William A. Lucas and Trudi M. Lucas, *Simulation and National Development* (New York: Wiley, 1966); J. David Singer and Hirohide Hinomoto, "Inspecting for Weapons Production: A Modest Computer Simulation," *Journal of Peace Research* (1965), 18–38; James A. Robinson, Lee F. Anderson, Margaret G. Hermann, and Richard C. Snyder, "Teaching with Inter-Nation Simulation and Case Studies," *American Political Science Review*, 60 (1966), 53–65.

[3] The simulation runs were conducted in the summer of 1961 at Northwestern University. The exploratory nature of these runs led the authors to question whether the pilot study should be published. The supply of the original Navy report, however, is now exhausted. Moreover, no more complete set of historical runs has been conducted to date. Because a number of other published materials have discussed these pilot runs, it seems appropriate to make a fuller description of the World War I simulation more widely available. In doing so, the authors wish to caution that the work is primarily an examination of a means of evaluating simulations rather than a direct validation of the Inter-Nation Simulation. For examples of how this pilot project has been discussed elsewhere, see Arthur Herzog, *The War-Peace Establishment* (New York: Harper, 1963), esp. 183–184; Sidney Verba, "Simulation, Reality, and Theory in International Relations," *World Politics*, 16 (1964), esp. 507–515; James A. Robinson and Richard C. Snyder,

The study was undertaken to investigate the use of a historical situation as a means of validating simulations. The problems of model validity are critical in determining the value of the simulation-gaming technique not only to political science, but to all the social sciences. In a fuller discussion of simulation validity elsewhere,[4] one of the authors has indicated that model validity is always a matter of degree and is affected by (1) the purpose for which the model is used, (2) whether or not human participants are involved, and (3) the types of criteria employed. The World War I simulation explores the third area—criteria for estimating validity. It focuses on possible standards or criteria for establishing the goodness of fit between the simulation and the system represented. To what extent do features of a political system or its processes correspond to their simplified representation in a model? One means of investigating this question is to ascertain if a simulation produces events similar to those reported in a historical situation. Another approach is to determine whether the simulation supports more general hypotheses about political phenomena which previously have been confirmed by independent methods. Both events and hypotheses are used as validity criteria in the simulation of the 1914 crisis.

Although the validity issue is the primary reason for conducting the simulation of World War I, several other purposes are served by the exercise. First, it provides a milieu in which to explore the relative effect on political actions of personality characteristics as compared to variables more frequently associated with political analysis. Second, the simulation of past events offers a possible device for teaching and studying history.

I. PROCEDURE

The Inter-Nation Simulation. Researchers differ as to the distinction between games and simulations. A number of experimenters, however, have associated "games" with operating models that involve human participants and "simulations" with models which do not.[5] Usually human participants are involved when the procedures or rules for designating the interplay of all components in the model have not been explicitly determined. When the model's relationships are incompletely programmed, human players and administrators are required to make judgments during the game. If the relationships are programmed, the need for human decision makers is reduced. In this essay an operating model will be defined as a simulation rather than a game, if a separate staff or computer is required to execute the programmed features. Thus, it is possible to have a simulation that is partially programmed and partially determined by human participants.

The model of international politics used in this study is such a hybrid. Developed by Harold Guetzkow and his associates,[6] the Inter-Nation Simulation incorporates both human participants and programmed calculations. In its usual format the Inter-Nation Simulation involves five or more nations. The government of each simulated nation is represented by human participants who assume one of several decision-making positions. During the 50-to-70 minute periods into which the simulation is divided, the decision makers allocate the military, consumer, and natural-industrial resources available to their nation. These various types of resources have different functions in domestic and international affairs. Using their resources the participants make decisions about internal matters such as economic growth, government stability, defense preparations, and research and development programs. At the international level nations may enter alliances, negotiate trades or aids, engage in various kinds of hostilities, and participate in international organizations.

Every period, which represents approximately one year of "real" time, the decision makers record their actions on a standardized decision form. Then, either a calculation staff or a computer applies the programmed rules to the decision form to determine the net gain or loss in the various types of resources. The structured part of the model also establishes whether the decision makers have maintained

---

"Decision-Making in International Politics," in Herbert C. Kelman (ed.), *International Behavior* (New York: Holt, Rinehart and Winston, 1965), 445, 512; and J. David Singer, "Data-Making in International Relations," *Behavioral Science*, 10 (1965), p. 77.

[4] Charles F. Hermann, "Validation Problems in Games and Simulations with Special Reference to Models of International Politics," *Behavorial Science*, in press.

[5] For a discussion of the distinctions made between games and simulations, see Charles F. Hermann, "Games and Simulations of Political Processes," *International Encyclopedia of the Social Sciences*, in press.

[6] Harold Guetzkow, Chadwick F. Alger, Richard A. Brody, Robert C. Noel, and Richard C. Snyder, *Simulation in International Relations* (Englewood Cliffs, N. J.: Prentice-Hall, 1963).

TABLE 1. FOUR SIMULATION PARAMETERS FOR 1914 NATIONAL PROFILES

Nation	Basic Capability Units[a]	Force Capability Units[b]	Validator Satisfaction[c]	Decision Latitude[d]
Austria-Hungary	45,540	14,560	4	7
England	86,940	25,000	6	3
France	62,100	20,800	4	3
Germany	120,000	24,500	7.5	8
Russia	78,660	23,000	4.5	9
Serbia	less than 4,140	less than 1,700	not given	not given

Note.—With the exception of decision latitude, the 1914 data and procedures for estimating these parameters were derived from James A. Winnefeld, "The Power Equation Europe, 1914," Stanford University (1960).

[a] Basic capability units represent the human, natural, and industrial resources available in a country. For the 1914 period the following indices were combined: steel production, national income, and total population weighted by the rate of male literacy.

[b] Force capability units are the military component in the simulation and were calculated by combining two indices, regular peacetime armies and capital ships.

[c] Validator satisfaction is the degree to which a decision-maker's policies are acceptable to those elite groups with power to authenticate his office holding. A "crisis coefficient" composed of the frequencies of certain types of events (e.g., civil disturbances and insurrections, assassinations) in a given country in the 50 years preceding 1914 was combined with an indicator of national security (relative military strength). The integrated estimates of satisfaction were placed along a 10-point scale. Higher values represent more satisfaction with the government.

[d] Decision latitude is an ideological element. It is defined as the degree to which probability of office holding is responsive to changes in validator satisfaction. To estimate decision latitude, judges rated the nations on a 10-point scale. Higher values represent greater latitude for the government (i.e., less sensitivity to the demands of validators).

the support of the politically relevant sectors of the nation whose endorsement is required for them to remain in office. The calculated results are fed back to each nation, thus beginning a new period of interactions and decisions by the participants.

Adapting the Simulation. Five nations were represented in the simulation runs of the 1914 crisis. Each government was staffed by two decision makers. A third participant in each nation acted as a messenger. The five simulated nations were intended to replicate features of Austria-Hungary, England, France, Germany, and Russia. Italy was excluded altogether and Serbia was represented symbolically by the researchers without participants. Several reasons can be offered for this treatment of Italy and Serbia. Reliable records of their diplomatic communications (a major input in the simulation of the other nations) were not available. Secondly, one can argue that although Serbia seemingly precipitated the immediate conflict and Italy was a member of the Triple Alliance, both nations were on the periphery in determining the question of world war when compared to the five other countries. Their exclusion, therefore, did not hinder the purposes of the exploratory runs. Although frequently other runs of the Inter-Nation Simulation have included more nations and more participants per nation, these changes are not as fundamental as several others.

Two major modifications were made in the basic simulation model. The first alteration established the initial conditions which the experimenters deemed important to characterize the international situation in the summer of 1914. Participants were introduced to some of the attributes of the historical setting by means of (1) a brief history of selected international affairs prior to the beginning of the crisis, (2) a statement of the current domestic and foreign policies of the participant's nation and the reasons they were being pursued, (3) a sketch of several personality traits of the historical policy maker whose role the participant occupied, and (4) a set of relevant historical diplomatic messages, conversations, and newspapers for the time between the assassination of the Austro-Hungarian Archduke on June 28 and the Serbian reply to the ultimatum on July 25, 1914. In addition, an effort was made to fit the programmed parameters and variables of the simulation to the national profiles of the countries involved.[7]

Several of these inputs require elaboration. Values for most of the components in the programmed part of the simulation were based on 1914 statistical indices (e.g., population, gross national product, size of armed forces) that approximated the meaning of the model's parameters. The 1914 indices were multiplied by an arbitrary constant to convert them to amounts convenient for use in the simulation. Individuals familiar with recent European history were asked to estimate decision latitude, the parameter that indicates how sensitive policy makers must be to the politically-relevant segments of their nation. These judges rated the decision latitude of each nation on a 10-point scale. Higher values were assigned to

[7] The international history, the statements of domestic and foreign policy, and the personality sketches appear as appendices in Technical Paper 3351, *op. cit.* A complete set of the diplomatic messages in the form in which they were used in the simulation is on file with the contracting agency and with the International Relations Program, Northwestern University. For values assigned the basic parameters, see Table 1

nations whose policy makers enjoyed greater freedom of action. Table 1 displays the values of each nation's basic parameters as they were reported to the simulation participants at the beginning of both runs. The historical diplomatic materials used in the simulation were compiled from hundreds of communications and documents which had been translated, edited, and verified by the Stanford Studies in Conflict and Integration under the editorship of Howard Koch.[8] The procedures for determining the personality characteristics are described in the section on participants and their historical counterparts.

All of the structured inputs were masked to avoid revealing to the participants that an actual historical situation was being modeled. Proper nouns, e.g., the names of individuals, countries, and alliances, were falsified. Misleading cues casting events in the future were introduced into the world history. In addition, after a pretest the assassination at Sarajevo was modified to avoid disclosing the identity of the historical setting. In the revised simulation account, several major Austro-Hungarian officials were killed by a strafing aircraft while they were on a reviewing stand in Serbia. These precautions were taken in order that the participants' knowledge of history would not bias their responses. The introduction of an extensive amount of structured material in the initial phase of the operating model represented an important modification of the usual practice in the Inter-Nation Simulation.

The second major alteration in the Inter-Nation Simulation affected the time units represented in the model. Usually the simulation is divided into 50-to-70 minute periods which constitute the equivalent of a year in the "real" world. These intervals are associated with a year because of the programmed calculations. Every period the policy makers allocate their national resources. The consequences of their allocations are determined by the relationships between a number of variables in the programmed part of the simulation. Examples of these variables include the rate of depreciation in existing military equipment, the amount of lead time required for a new research program, and the extent of shift in popular support for the government. The equations used to calculate these and other variables are designed to reflect changes that might occur on roughly an annual basis. As a result, decisions taken by the participants and submitted for calculation normally represent the allocation of resources for approximately twelve months.

The present exercise, however, required the representation of not years, but the few critical days in late July, 1914. This reduction in time necessitated several seemingly contradictory changes. On the one hand, the existing programmed calculations had to be made relevant to the participants. If the basic model was to be maintained, persons involved in the exercise had to experience the constraints and demands imposed on their immediate behavior by the programmed features of the simulation as it is usually constituted. On the other hand, participants required a time framework that would allow them to deal with the kinds of decisions that policy makers might encounter on a daily rather than annual basis. In sum, the individuals in the simulation had to be able to make short-term decisions, while being aware of the long-term consequences as represented in the programmed calculations.

To meet these requirements only individuals who had previous experience with the Inter-Nation Simulation were invited to participate in the exercise. These experienced participants were told that the first few simulation periods would represent days. Moreover, there was the implication that these short time frames would be followed by periods which would be equivalent to years as in previous simulations. To further the impression that the initial periods were to be embedded in a series of longer time units, the participants received an annual decision form and were instructed to submit an updated version for calculations when the simulation periods began to represent years. These arrangements were made to encourage the participants to take account of the basic programmed variables and parameters in conducting their immediate interactions. A decision maker is more likely to act in a short-term situation so as not to damage such programmed variables as the probability of his continuing in office or the annual amount of consumer goods available to his nation, if he believes that the simulation is going to continue for "years." Notwithstanding the information given the participants, both simulation runs were terminated before the anticipated conversion to the longer time units had occurred. Hence no calculations were made in any of the programmed components of the model.

In addition to designating the periods as days, the shortened character of the time units was promoted by the diplomatic messages. These historical documents reflected events

[8] Howard E. Koch with the staff of the Stanford Studies in Conflict and Integration, *Documentary Chronology of Events Preceding the Outbreak of the First World War: 28 June-6 August, 1914* (Stanford University, mimeographed, 1959).

and decisions that were developing day by day, if not hour by hour. Finally, as an aid to short-term decision-making, a new form was introduced into the simulation. On the new instrument each nation's decision makers were able to indicate more immediate changes in the intensity of their action toward other nations. Participants were advised that daily variations in their nation's level of commitment would influence such annually computed variables as total available resources and the likelihood of office-holding. Thus, the intensity scale provided an explicit link with the long-term elements present in the programmed calculations of most Inter-Nation Simulation exercises.

Historical Figures and Their Simulation Counterparts. An attempt was made to select participants with personality traits similar to some of those manifested by political leaders in the crisis of 1914. This task involved three subproblems. First, a restricted number of historical figures who were active in the crisis had to be selected. Second, a judgment had to be made regarding which personality characteristics of these men were salient in their political behavior. Finally, a method had to be devised for selecting simulation participants with similar personality profiles.

The resources available to the researchers necessitated that the total number of decision makers in each run be limited to 10—two participants for each nation. Consequently, we sought the two policy makers in each of the five European nations who were major contributors to the critical decisions during the crisis. More specifically, three criteria guided the selection: (1) which persons had a dominating influence on the foreign policy decisions of their nation at the time of the crisis? This criterion recognizes that the loci of decision making may not correspond with the "legitimate authority" to make such decisions. (2) Which persons received and dispatched (or at least read) diplomatic cables and related foreign policy documents? The historical figures whom the simulation participants represented should have occupied a reasonably central position in the diplomatic communication net because diplomatic messages acted as a major source of simulation inputs. (3) About which persons was autobiographical and/or biographical material available to help in the assessment of personality traits? Utilizing these criteria, the following historical personages were chosen for representation in the simulation:

Austria-Hungary
- Berchtold (Minister for Foreign Affairs)
- Conrad (Chief of General Staff)

England
- Grey (Secretary of State for Foreign Affairs)
- Nicolson (Permanent Under-Secretary of State for Foreign Affairs)

France
- Poincaré (President of the Republic)
- Berthelot (Acting Political Director of Foreign Ministry)

Germany
- Wilhelm II (Kaiser)
- Bethmann-Hollweg (Chancellor)

Russia
- Nicholas II (Czar)
- Sazonov (Minister for Foreign Affairs)[9]

Once the political leaders judged to have assumed critical roles in the crisis were selected, it was necessary to establish which personality characteristics were salient to their political behavior. This determination was made by a cursory content analysis of personal letters, autobiographies, and biographies of the chosen policy makers.[10] Each document was content

[9] Some illustration of how the criteria were employed is appropriate. The selection of Berthelot provides a good example of the application of the first two criteria. Although Berthelot did not have the legitimate authority to make binding decisions for his government, he nevertheless extensively influenced French foreign policy during July of 1914. With the president, premier (who was also the foreign minister), and political director of the foreign ministry on a mission to Russia in July, Berthelot was placed in charge of the foreign ministry. He became the chief advisor during this time to Bienvenu-Martin, acting Premier and Foreign Minister. Bienvenu-Martin, officially the Minister of Justice, was a novice at foreign affairs, and relied heavily on Berthelot. For further evidence on the role of Berthelot, see Richard D. Challener, "The French Foreign Office: The Era of Philippe Berthelot," in Gordon A. Craig and Felix Gilbert (eds.), *The Diplomats: 1919–1939* (Princeton, N. J.: Princeton University Press, 1953), 49–85. The third criterion was important in the choice of Conrad. Far less autobiographical and biographical material is available on the Austro-Hungarian political leaders than on those from the other countries. Most information concerns Emperor Franz Joseph, a logical selection in addition to Berchtold for the simulation. However, because Franz Joseph was quite old and recovering from a serious illness, he was not as influential on the decisions as other officials. Among the key figures, Conrad had the most available material.

[10] No inter-coder reliability was performed in the content analysis of the personality traits. For this reason, as well as because of the very limited

analyzed for psychological characteristics or traits identified by one of several tests for measuring personality.[11] On the basis of the content analysis, dominance, self-acceptance, and self-control appeared to be characteristics which differentiated among all 10 of the political leaders. Furthermore, the personality of one or more of the selected individuals was strongly characterized by his attitude toward such concepts as fate, frankness, making decisions, his own country, peace, self-confidence, success, suspicion, and war. No claim is made that these traits provide a complete personality profile of the historical individuals under examination. Undoubtedly some important characteristics have been overlooked. This list of traits, however, yielded a distinctive profile for every leader which was consistent with features stressed in documents describing that individual.

Two psychological tests were used to measure the personality characteristics that the researchers associated with the selected historical figures. The California Psychological Inventory (CPI)[12] not only measured the three traits judged to be relevant to all the policy makers, but it also contained measures of some secondary characteristics identified in the content analysis. The second instrument was the semantic differential.[13] The nine attitudes were estimated with this testing device. Utilizing a suggestion made by Gough,[14] one of the researchers responded to both instruments as if she were the historical policy makers. The tests were completed after the biographical and autobiographical material on each of the chosen figures had been read. Test profiles for all 10 individuals were made in this way. These 10 profiles provided a standard against which to compare the responses of potential simulation participants on the same personality tests.

Some 101 high school students, who had participated in the Inter-Nation Simulation experiment in the summer of 1960,[15] were tested. As previously noted, persons already acquainted with the operation of the simulation were used in order to facilitate the shift in the time dimension. Furthermore, experienced participants freed the experimenters from training participants in basic simulation skills. The CPI profile of each prospective participant was compared with that of each historical figure. Particular attention was paid to profiles that matched exactly on dominance, self-acceptance, and self-control and were within one standard deviation of the other traits measured by the CPI. The semantic differential was utilized as a final selection step. In other words, those individuals chosen through the CPI matching were screened further by means of the semantic differential until three individuals per role were selected. Of the 10 individuals best matched to the historical figures, six were able to participate in the runs—five in the first run and one in the second. The balance of the participants in each run were second and third choices. Because of the interest in using individuals whose profiles most closely corresponded to those prepared for the actual leaders, the participants were not controlled on sex. Four women participated in the first run and three in the second.

Conducting the Revised Simulation. To provide an overview of the procedures used in the modified Inter-Nation Simulation, we will describe the operations in the two attempted replications of the pre-World War I crisis. Both simulation exercises were conducted in two days. On the first day, the participants in the runs assembled together for a general introduction and review of simulation procedures. After these activities, the remainder of the first day was used to introduce the structured input for each run.

At the beginning of the input phase, each participant was assigned to a separate cubicle. There they started by reading a disguised statement of pre-1914 "world history" and a description of the individual whose role they would assume. Upon completing this material, every participant was given a set of masked diplomatic communiques, newspapers, and memoranda which might reasonably have come

---

sample of materials that could be examined for each figure, the selected traits must be considered only as tentative approximations. A list of the sources used in the content analysis appears in Technical Paper 3351, *op. cit.*

[11] Sources for the personality categories were Raymond B. Cattell, *The Sixteen Personality Factor Questionnaire*, rev. ed. (Champaign, Ill.: IPAT, 1957); Allen L. Edwards, *Edwards Personal Preference Schedule* (New York: Psychological Corp., 1953); Harrison G. Gough, *California Psychological Inventory* (Palo Alto, Calif.: Consulting Psychologists Press, 1956).

[12] Harrison G. Gough, *ibid.*

[13] Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum, *The Measurement of Meaning* (Urbana, Ill.: University of Illinois Press, 1957).

[14] Personal correspondence from Harrison G. Gough, dated July 25, 1961.

[15] For a description of this earlier simulation research, see Richard A. Brody, "Some Systemic Effects of the Spread of Nuclear Weapons Technology: A Study through Simulation of a Multi-Nuclear Future," *Journal of Conflict Resolution*, 7 (1963), 663–753.

to the attention of his historical counterpart. To the extent that such information was available, messages were ordered in the sequence in which they were received by the historical figure. For example, although the simulation began with an incident representing the assassination of the Archduke on June 28 (known to the participants as Day 1), some participants learned about that event the same day; others did not. In general, a participant received an incoming message and then the recorded reaction, if any, which his counterpart had made to that message. If evidence indicated that the historical figure had been aware of responses made to a message by other members of his government, the simulation participant also received this information. Thus, the Kaiser's marginal notes on diplomatic communiques were seen by the individual assuming the role of the German Chancellor, Bethmann-Hollweg. The input phase terminated with the Austro-Hungarian ultimatum on July 25 (Simulated Day 28). As the historical events became more rapid and complex in the latter part of the input phase, the order in which participants received information took on considerable importance.

The participants were allowed to read the set of messages available to them at their own pace. They were encouraged to write their reactions to each communication in a space opposite the message. The researchers believed that the participants would become more involved in the situation if they recorded their thoughts about the messages as they read. Moreover, these reactions provided a source of data. For the same reasons, after Day 15 and again after Day 28, every individual was required to write a summary of the events that had occurred. On these two simulated days, the participants also were provided sheets of paper and asked to draw a map of the world as they conceived it at that time.[16] After completing the second history and map, the participants were dismissed until the following day. They were cautioned not to discuss the material they had read.

The first run was concluded on the morning of the second day (hereafter designated the M-run) and the second was finished that afternoon (the A-run). When the participants returned to their respective simulations, they were informed that the world would continue with the situation as it had evolved up to that point. They were told that there would be no more structured messages. Thereafter, the situation could be handled in whatever manner they chose. Each nation's decision makers were given an annual decision form and an intensity of action scale to indicate their country's resources and commitments at the beginning of the free activity phase of the simulation. They were instructed to designate each 50-minute period as one simulated day, but the number of such periods that would occur remained unspecified. Following these initial instructions, the two participants in each nation were assigned to a room in which they were separated by a partition. A messenger or courier sat at one end of the partition to prevent unauthorized conversations, to relay messages, and to operate a tape recorder during conferences between the decision makers. Written messages and periodic conferences were used as means of communication both within and between nations.

At the end of three periods (representing July 26–28, 1914), each participant was asked to draw another map and update his statement of world events. Much to the surprise of the participants, upon completing this task they were informed that the simulation was over. A post-mortem or debriefing session was held with each simulation run. The participants first completed a questionnaire and then described their plans and reactions to events in the simulation.

II. RESULTS

Two different standards or criteria were employed in estimating the validity of actions taken by the decision makers in the free activity phase of the simulation runs. First, both macro and micro level events in the 1914 crisis were used as standards with which to compare incidents that occurred in the simulation. Second, two general hypotheses, previously tested with documents from the outbreak of the First World War, were explored using simulation data. Findings on the hypotheses from the two data sources were checked for comparability.

Macro and Micro Events. In the present context, "macro event" refers to the occurrence of

[16] Geographical features are not incorporated in the Inter-Nation Simulation and no explicit geographical statements were included in the masked historical communiques. Therefore, the participants' maps provided data about the changing conception of relationships among nations. The distance between allied and hostile nations as well as the relative size of the nations represented in the maps were analyzed. Although the measuring device is worthy of further exploration, the results from these pilot runs proved to be quite ambiguous and are not reported here. They are included in Technical Paper 3351, *op. cit.*

general war. Did war break out in the simulation runs as it had in Europe in 1914? In neither run did war—or more accurately, the representation of war—occur during the last three periods. Historically, Austria-Hungary declared war on Serbia on July 28th, the final day represented in the simulation runs. In 1914, the declarations of general war between the other European nations did not occur until after the last time period portrayed in the simulation.[17] Although no war had been declared when the simulations were terminated, hostilities were imminent in the M-run.

In the researchers' opinion, if the M-run simulation had been continued for another 50-to-100 minutes (one or two more simulated days), war would have been declared along lines similar to the historical situation. This position is confirmed by 10 of the 15 M-run participants and messengers in their debriefing questionnaires. Throughout the free activity phase of that run the two alliance structures held. Germany[18] was prepared to give secret aid to Austria-Hungary for an attack on Serbia while signing neutrality pacts with England and France to keep them out of the war. This effort was intended to localize the conflict and to assure victory for the Dual Monarchy. If Russia then considered assisting Serbia, she would lack the support of her allies—a fact which Germany and Austria-Hungary believed would deter such action. France and England, however, had plans to go to war if Austria-Hungary attacked Serbia. That attack by Austria-Hungary was being planned at the close of the simulation run. An internal conference between the German participants on the next to last simulated day illustrates the direction in which the M-run was moving.

Central Decision Maker [Kaiser]: Is their war started yet? [Austria-Hungary's attack on Serbia]

External Decision Maker [Bethmann-Hollweg]: I don't think so.

CDM: Here's the problem—we can certainly give them [Austria-Hungary] the aid but the aid must be kept secret because if monetary aid isn't kept secret then the promises—written promises—which I am giving to Bega [England] and Colo [France] will mean absolutely nothing; total war is bound to break out. . . . If Enuk [Russia] enters it, we have to send secret aid; if Enuk [Russia] doesn't enter it, then he [Austria-Hungary] can defeat Gior [Serbia] by himself and that will be a thorn out of our side. We have already driven a wedge into the Tri-Agreement [Triple Entente].

EDM: Another question: Suppose Colo [France] enters, then we go in?

CDM: If Colo [France] enters the war we go in. . . .

In the A-run the participant representing Lord Grey of England called for an international conference on the first simulated day of the free activity phase. Thereafter, most of the simulation time was spent in obtaining agreement from all nations to attend this meeting. At the conference Austria-Hungary was charged with making far too extreme demands against Serbia and was pressured into withdrawing them. This retraction resulted in substantial conflict among the decision makers in Austria-Hungary as well as a bitter dispute between Germany and the Dual Monarchy. Germany had pledged complete support to Austria-Hungary and was highly irritated at the failure of Austria-Hungary to consult privately before changing its policy.

Two observations should be made about these developments in the A-run. First, some historians have expressed the opinion that England should have taken a stronger position more quickly, forcing moderation upon the Triple Alliance.[19] Second, and perhaps of greater importance, if the input phase had been extended for one more historical day, the participants would have received England's diplomatic communiques proposing an international conference and the prompt rejection of this suggestion by several nations. Thus, an alternative actually considered and subsequently

[17] The chronology of hostilities during this critical period in 1914 was as follows: July 28—Austria-Hungary declares war on Serbia; July 29—Russia orders and then cancels general mobilization; July 30—Russia again orders general mobilization, France alerts troops along German border; July 31—Austria-Hungary begins general mobilization, Germany issues ultimatum demanding Russia stop mobilizing within 12 hours or Germany will mobilize; August 1—France and Germany start general mobilization, Germany declares war on Russia. After August 1, formal declarations of war follow in quick succession from the other major European states.

[18] It is to be understood that the references to nations here are to the student decision makers in the simulated nations.

[19] See, for example, Luigi Albertini, *The Origins of the War of 1914*, edited and translated by Isabella M. Massy (London: Oxford University Press, 1953), Vol. 2, p. 514; and Sidney B. Fay, *The Origins of the World War* (New York: Macmillan, 1930), Vol. 2, p. 556.

excluded by the historical figures provided the avenue which the simulation participants followed for the resolution of the imposed situation.

Several major divergences from the outbreak of World War I developed in the A-run. Least parallel to historical events is the indication in some messages that England was considering the initiation of war on Germany while the advantage appeared on her side. This war was represented in communications as a defensive strategy resulting from the military and economic threat of Germany. A second significant variation was the agreement of Austria-Hungary to withdraw her ultimatum. The Austro-Hungarian decision maker, intended to represent Conrad, revealed pacifistic tendencies and readily accepted the objections to his nation's militaristic actions. This behavior suggests a need to match more closely the social-political attitudes of the historical figures and the simulation participants.

This account of the A-run's divergences fittingly introduces the analysis of smaller, more specific events that occurred in the simulation and in the last days of July, 1914. By comparing numerous micro occurrences, validity depends not on the correspondence between isolated events (war or no war) but on the overall goodness of fit between a distribution of events. In other words, we are interested in whether the overall pattern of simulation occurrences is more or less like the pattern of reported incidents prior to World War I.

To illustrate this approach a sample of micro events was drawn from the simulation (the M-run) that at the macro level displayed a higher degree of correspondence to 1914. In this analysis an action mentioned by the participants in their written communication constituted a micro event. Eighteen separate micro events were identified in the M-run messages during the first simulated day of the free activity phase. A somewhat longer series of events would probably have been discovered for that day if transcripts of conferences had been examined in addition to the written communications. No reason, however, has been found to suggest that the types of events produced in face-to-face interactions differ for validity purposes from those indicated in written messages.

After the simulation events had been abstracted, a major historical study of the beginnings of the First World War was examined for comparable events.[20] In Table 2 the simulation events are indicated in the left-hand column and the reported historical events are listed in the corresponding space in the right-hand column.

Several alternative means of comparing the events were identified. Comparable activity is one way to match events in both the pre-World War I crisis and the simulation. We describe this as occurrence similarity. A second basis of comparison involves intent or purpose. Men's purposes in initiating or manipulating events may be similar, regardless of differences in format or activity. For example, in Item 16 in Table 2 the simulation action is the reverse of that reported in the observable world of 1914, but the intentions of the decision makers in both actions were probably similar. Germany was concerned that any war between Austria-Hungary and Serbia be kept localized without the intervention of other countries. To achieve this objective the simulation decision makers chose to elicit pledges of neutrality from two important nations. In actuality, the German decision makers decided that quick action, before commitments and reactions could be made, was the preferable policy. Another way of comparing events is through temporal equivalence. The timing of some simulation events closely matches that of similar events in reality; others deviate sharply in this temporal aspect.

One of the researchers rated the 18 simulation events on each of the three bases of comparison. For each simulation event, temporal equivalence was given a score of one (i.e., assumed similarity) if it occurred within one day on either side of the reported 1914 event to which it was compared. If the time disparity was greater than the arbitrary threshold, it was coded as nonsimilar and assigned a zero. The same zero or one scoring was used for comparing events on intention and on occurrence similarity. The estimated scores for each event are presented in the center column of Table 2.

If the scores for the three categories are combined for the 18 events, the highest possible goodness of fit score is 54. The overall score for the distribution of events in the first simulated day of the M-run is 31. More specifically, all 18 events were judged similar in intent to developments that occurred in the 1914 crisis. Weighed against the significance of this finding must be the recognition that it is difficult to ascertain the intent of simulation participants, to say nothing of historical policy makers. Rules for guiding such coding are exceedingly hard to devise. For half of the simulation events (9 of 18), the researchers were able to

[20] Luigi Albertini, *ibid.*, Vol. 2.

TABLE 2. COMPARISON OF SAMPLE OF SIMULATION EVENTS WITH HISTORICAL EVENTS

Simulated July 26, M-run	Estimated Score[a]	Reported Historical Events[b]
1. Germany requests Russia to demobilize.	T = 0 I = 1 O = 1 — 2	*July 29:* German note to Russia: "... further progress of Russian mobilization measures would compel us to mobilize and then European war would scarcely be prevented" (p. 491).
2. Russia notifies Germany that no demobilization will occur until safeguards are established for Serbia.	T = 1 I = 1 O = 0 — 2	*July 26:* Russian foreign minister informs German ambassador that Russia is ready to assist in "procuring legitimate satisfaction for Austria-Hungary without abandoning the standpoint, to which Russia must firmly adhere, that Serbia's sovereignty must not be infringed" (p. 404). There is no explicit mention of making demobilization conditional on such safeguards.
3. In a conference Russia asks England and France to mobilize; both refuse.	T = 0 I = 1 O = 0 — 1	*July 24:* Conversation between Russian foreign minister and ambassadors from France and England. France agrees to fulfill all obligations of her treaty. England more evasive but conciliatory. (Pp. 294–96.)
4. In a conference England and France inquire as to Russian military needs.	T = 0 I = 1 O = 1 — 2	
5. Germany notifies Russia that the latter's support of Serbia would be aggression.	T = 0 I = 1 O = 0 — 1	*July 28:* No evidence found of direct communication between Germany and Russia on this matter; however, Albertini concludes that this was a German objective. In support he cites several telegrams emanating from the German foreign office on July 28 including a note from the Kaiser: "It is an imperative necessity that the responsibility for a possible extension of the conflict to the Powers not immediately concerned should in all circumstances fall on Russia" (p. 474 and p. 476).
6. Austria-Hungary warns Russia not to mobilize further.	T = 0 I = 1 O = 1 — 2	*July 29:* There is some suggestion that the Austro-Hungarian ambassador warned the Russian foreign minister against Russian intervention on this date, although the conversation was interrupted by information that Austria-Hungary had bombed Belgrade. (Pp. 552–53.) On July 26 Austria-Hungary's alliance partner, Germany, issued warnings through her ambassador to the Russian foreign minister: "I made detailed and urgent representations to the Minister about how dangerous it seemed to me to attempt to strengthen diplomatic action by military pressure" (p. 481).
7. Discussion on need for attack on Serbia in Austro-Hungarian internal conference.	T = 0 I = 1 O = 1 — 2	*July 7:* Austro-Hungarian Council of Ministers for Joint Affairs: one of a number of internal conferences before delivery of the Serbian ultimatum in which plans for attack on Serbia and the expected Russian reaction were discussed. No explicit program to make Russia appear the aggressor is mentioned although this appears to have been part of the Triple Alliance strategy. (P. 166.)
8. Discussion in Austro-Hungarian internal conference on making Russia appear the aggressor in an expansion of hostilities.	T = 0 I = 1 O = 0 — 1	
9. Austro-Hungarian note to Russia charges Russia's action is aggressive.	T = 0 I = 1 O = 1 — 2	*July 30:* Austro-Hungarian foreign minister tells Russian ambassador: "Austria-Hungary had mobilized solely against Serbia, not against Russia. . . . By the fact that Russia is obviously mobilizing against us we should also have to extend our mobilization. . . . " (p. 662).

[a] "T" stands for temporal equivalence, "I" for intention, and "O" for occurrence similarity. Each of the three categories was assigned a "1" if the simulation and 1914 event were judged as similar on that criterion; or a "O" if they were judged dissimilar. Simulation and historical events rated as similar on each of the criteria received an overall score of three.

[b] The source for the reported historical events is Luigi Albertini, *The Origins of the War of 1914*, edited and translated by Isabella M. Massy (London: Oxford University Press, 1953), Vol. 2. All page references are to this volume.

TABLE 2—(*Continued*)

Simulated July 26, M-run	Estimated Score[a]	Reported Historical Events[b]
10. Austro-Hungarian note to Russia states Austria-Hungary must hold Serbia accountable.	T = 0 I = 1 O = 1 — 2	*July 30:* On the basis of above conversation with Berchtold, the Russian ambassador telegraphs his own foreign minister: "[Berchtold] is of the opinion that it is impossible for Austria to stop her operations against Serbia without having received full satisfaction and solid guarantees for the future" (pp. 662–63).
11. In note France assures England of friendship.	T = 0 I = 1 O = 0 — 1	*July 29:* France appears to have repeatedly probed England on the issue of unified action and support. For example, the July 24 three-way conversation in St. Petersburg (pp. 294–96) or the inquiry to the British ambassador by his French colleague in Berlin, July 29 (p. 520n). A major discussion on this point took place between the British foreign minister and the French ambassador on July 29. In a follow-up conversation on July 31, the ambassador asked Grey if England would "help France if Germany made an attack on her." (P. 646.) However, the type of military aid was not specified.
12. France requests military aid in note to England.	T = 0 I = 1 O = 1 — 2	
13. France urges a united front in a note to England.	T = 0 I = 1 O = 1 — 2	
14. Germany makes internal decision to give secret aid to Austria-Hungary.	T = 0 I = 1 O = 1 — 2	*July 5–6:* The Hoyos mission occurred, and Germany agreed to the so-called "blank check." Subsequent conversations between the Kaiser, his Chancellor, and military staff also reveal a discussion of military assistance to Austria-Hungary. (Pp. 133–59.)
15. Germany's internal decision to seek neutrality pact with France and England to keep them from aiding Russia should latter intervene in Austria-Hungary's war against Serbia.	T = 0 I = 1 O = 0 — 1	While no direct documentation was found for verifying this internal decision, Albertini concludes: "The study of the German documents shows beyond all shadow of doubt (1) that in allowing Austria to attack Serbia, Germany started from the assumption that, if the attack developed into a European war, England would remain neutral; (2) that Grey's conduct until the afternoon of July 29 . . . strengthened the German leaders in this opinion. . . . ." (p. 514).
16. Germany urges Austria-Hungary not to attack until neutrality of France and England is assured.	T = 1 I = 1 O = 0 — 2	*July 26:* Austria receives Germany's message urging military operations against Serbia as quickly as possible: "Any delay in coming military operations is regarded here as a great danger because of the interference of other Powers" (p. 453).
17. Decision in Russian internal conference not to reveal, at present, their plan to attack Austria-Hungary if she attacks Serbia.	T = 1 I = 1 O = 0 — 2	*July 25:* Russian crown council decides on mobilization to be followed by war if Serbia is attacked. (P. 762.)
18. Decision in Russian internal conference to announce their intention to give Serbia full military aid.	T = 1 I = 1 O = 0 — 2	

find historical actions which seemed to be equivalent in physical format. Four of the 18 simulation events took place at approximately the point in time that the simulation was intended to replicate. One possible explanation for the low correspondence on temporal equivalence involves the problem of equating intervals in the simulation to actual days. The simulation may not have been adequately structured to provide the participants with the impression that each 50 minute period represented 24 hours of "real" time.

Hypotheses. In one sense the comparison of simulation and historical events includes the investigation of a hypothesis. The hypothesis is that a macro event or a distribution of events in system A is comparable in a specified characteristic to a given event or distribution of events in system A'. Such a proposition, however, is oriented exclusively to the two systems

involved—in the present case, the outbreak of World War I and the simulation of those events. In this section, attention is directed to more general hypotheses. These statements of relationship are intended to apply not only to the situation as it existed in Europe in the summer of 1914, but to other configurations of nations in international politics. Two hypotheses which have been tested with data from the 1914 crisis will be explored with the simulation data.

The three researchers who have studied the first hypothesis in the 1914 context have stated it as follows:

> If a state's perception of injury (or frustration, dissatisfaction, hostility, or threat) to itself is "sufficiently" great, this perception will offset perceptions of insufficient capability, making the perception of capability much less important a factor in a decision to go to war.[21]

To explore their hypothesis, the members of the Stanford Studies in Conflict and Integration content analyzed 1,165 communications exchanged between the European states in the weeks prior to the beginning of the First World War. The purpose of this analysis was to determine the frequency of perceptions of capability and hostility. In the content analysis a perception of capability was defined as an assertion "concerning the power of another state or a coalition of states, or a statement with regard to the changing power of either a state or a coalition of states." A perception of hostility was defined as "a statement by one country about the hostility directed toward it by a second country, or about the hostility directed by a second country toward a third country. . . . Statements of hostility are defined to include statements of threat, fear and injury."[22]

The number of hostility and capability statements that the Stanford group found for the nations subsequently represented in the simulation are reported in the first columns of Table 3. Perceptions of hostility exceed those of capability for every nation. The overall difference is statistically significant ($p < .05$) by a sign test. Using the same definitions of hostility and capability, the messages written during the two simulation runs were content analyzed in a similar fashion.[23] As shown in the second and third sections of Table 3, the perceptions of hostility were significantly greater ($p < .05$) than those of capability in both runs. Thus, the simulation runs and the 1914 data produce comparable results on this general hypothesis.

TABLE 3. NUMBER OF HOSTILITY AND CAPABILITY STATEMENTS FOR NATIONS IN 1914 AND IN SIMULATION

Nation	1914[a]		M-Run		A-Run	
	Hostility	Capability	Hostility	Capability	Hostility	Capability
Austria-Hungary	179	26	91	38	63	28
England	32	16	17	13	24	7
France	26	18	41	36	21	5
Germany	138	34	53	35	107	26
Russia	50	28	30	16	69	20

[a] The 1914 data are in Dina A. Zinnes, Robert C. North, and Howard E. Koch, "Capabilities, Threat and the Outbreak of War," in James A. Rosenau (ed.), *International Politics and Foreign Policy* (New York: Free Press, 1961), p. 476.

A further check was made on the goodness of fit between the simulation and the 1914 data. The difference between the number of hostility and capability statements was determined for each nation. Within each of the three sets of data these differences were placed in rank order, that is, with the nation having the largest difference being given the first rank, etc. Then the order of the ranks for each of the simulation runs was compared with the order produced by the 1914 data. The resulting rank-order correlation (.90) between the 1914 data and that of the M-run was statistically significant ($p = .05$). The correlation for the A-run was not significant, however.

The second general hypothesis asserts that when opposing alliances or blocs emerge in international politics, the communication between blocs will be much less than that among alliance partners. Zinnes, who has tested this hypothesis with the 1914 data, states the relationship as follows: "Frequency of interaction within the bloc will be greater than the frequency of interaction between blocs."[24] To

[21] Dina A. Zinnes, Robert C. North, and Howard E. Koch, "Capabilities, Threat and the Outbreak of War," in Rosenau, *op. cit.*, p. 470.

[22] *Ibid.*, p. 472.

[23] The inter-coder reliability for the simulation content analysis of hostility was .82 and for capability .83. For a description of the statistics used in this paper (the sign test, rank-order correlation, and Mann-Whitney *U* test) see Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill, 1956).

[24] Dina A. Zinnes, "A Comparison of Hostile Behavior of Decision-Makers in Simulate and Historical Data," *World Politics*, 18 (1966), p. 477. This article is one of the few efforts to explore simulation validity by what we have de-

TABLE 4. EFFECTS OF ALLIANCE STRUCTURE ON COMMUNICATION IN 1914 AND IN SIMULATION

Source of Data	Number of Intra-Bloc Dyads (n_1)	Number of Inter-Bloc Dyads (n_2)	Mann-Whitney U[a]	Significance
1914	8	12	15	<.01
M-Run	8	12	10.5	<.01
A-Run	8	12	43	—

[a] It should be noted that for a given n_1 and n_2, the smaller the U the greater its likelihood of being significant.

explore this hypothesis, a count was made of the number of communications that each nation dispatched to the other four nations primarily involved in the 1914 crisis. The number of messages exchanged between each possible combination of two nations was established. Rates of communication in dyads composed of nations in the same bloc were compared with those in dyads consisting of nations from opposing blocs. A Mann-Whitney U test applied to the rank-order positions of these dyads supports the hypothesis.[25] The results are shown in the first row of Table 4. Identical procedures were followed for the M- and A-runs using the messages sent by the decision makers to other nations. The results for the M-run, shown in the second row of Table 4, are statistically significant. By contrast, as indicated in the third row of Table 4, the findings for the A-run are not significant. Once again, the data from the M-run appear to be a better fit with that drawn from 1914 than do the data produced in the A-run.

III. EVALUATION

Replication of the 1914 Crisis. Can the Inter-Nation Simulation replicate occurrences in the observable universe such as the outbreak of World War I? Regardless of how intriguing this question is, for a number of reasons we cannot fully answer it with the present data. The exploratory nature of this initial research has been emphasized. A more complete study would require many more runs. It would be desirable to engage more mature participants whose previous experiences and backgrounds might lend themselves to the simulation problem. Moreover, the calculations used to set the beginning values of the parameters and the procedures used to code the simulation data need further refinement.

Even if one were tempted to make rough judgments on validity based exclusively on this pilot study, the divergence between the two runs imposes constraints. Using both events and general hypotheses to check the "fit" between 1914 and simulation data, the M-run approximates the political crisis prior to the First World War more closely than the A-run. At the macro event level, the M-run appeared on the verge of war with the same alliance commitments as observed in 1914. On the other hand, the A-run averted the immediate threat of war and involved several incidents in sharp contrast to those reported in the historical situation. In data from both runs, as in the material from 1914, perceptions of hostility were significantly more frequent than perceptions of capability. Only in the M-run, however, did the differences between these two kinds of perceptions correlate with similar data from the actual nations of 1914. Furthermore, the hypothesis that communication would be greater within alliances than between them was supported both by the 1914 data and the M-run data; but it was not confirmed in the A-run. These differences occurred despite a common introductory briefing, identical materials in the input phase, and the same set of initial values for the simulation parameters.

The two runs, however, did not share the same human participants. Although an effort was made to select participants in both runs with similar personality characteristics, it will be recalled that the matching was limited to a few traits. Moreover, on those characteristics for which an attempt at correspondence was made, the personality profiles of some individuals more closely matched their historical counterparts than others. As we noted previously, five of the six individuals whose profiles corresponded most completely with those designed to represent the historical figures participated in the M-run and only one in the A-run. In other words, half of the participants in the run that best approximated the 1914 crisis were first choices on the personality matching, whereas only one of the 10 decision makers in the more divergent run was a first choice. (In the A-run, six participants were second choices and three were third choices. Of the other five participants in the M-run,

scribed as the hypothesis approach. Statistical tests are conducted on hypotheses using data from World War I and from another series of Inter-Nation Simulation runs which made no attempt to replicate those historical events.

[25] The 1914 data for this analysis were obtained from Zinnes, *ibid.* She performed a similar statistical test using Serbia as well as the five nations represented in the simulation. It is interesting that with the addition of Serbia, the result is not significant.

TABLE 5. SITUATIONS VIEWED AS SIMILAR TO THE SIMULATION ON THE DEBRIEFING QUESTIONNAIRES

Situation	M-Run Decision Makers (N = 10)		A-Run Decision Makers (N = 7[a])		Totals (N = 17)		
	Similar	Almost Exact	Similar	Almost Exact	Similar	Almost Exact	Overall Total
World War I	3	2	1	1	4	3	7
Hitler's Ultimatum to Poland	1	2	2	1	3	3	6
Berlin Crisis	2	2	1	1	3	3	6
Hitler's Ultimatum to Sudetenland	2	1	2	1	4	2	6
Israeli-Arab Conflict	3	0	2	1	5	1	6
French-Algerian Conflict	1	1	2	1	3	2	5
Korean War	3	0	1	1	4	1	5
19th Century Colonialism	4	0	0	1	4	1	5
Spanish-American War	1	0	2	1	3	1	4
Attack on Pearl Harbor	1	0	0	0	1	0	1
World War II	1	0	0	0	1	0	1

[a] Three of the A-run participants were unable to attend the debriefing session.

three were second choices and two were third choices.) Correspondence between a simulation and its reference system appears to have been facilitated by closer matching of the personalities of the participant and the historical figure.

Procedural Issues. These initial runs do not establish whether the Inter-Nation Simulation can replicate aspects of historical events, but they do uncover several procedural issues that a more complete study would confront. One problem is the participants' awareness of the historical events being simulated. We have described the efforts to mask the historical clues to avoid biasing participant responses. Despite these attempts, the disguise was not fully successful. On the debriefing questionnaire, the participants were asked to check from a list of historical situations those incidents that appeared "somewhat similar" and those that seemed "almost exactly like" their simulation run.[26] Table 5 indicates that five of the 10 participants in the M-run and two of the seven who completed the questionnaire in the A-run perceived some degree of similarity between their simulation and World War I. No other historical situation was identified by seven participants, but four others were each chosen by six individuals.[27]

Even though the results from this questionnaire item are not conclusive, they suggest the need for more vigilance in making past events unrecognizable. When, however, the content of historical situations is changed or misleading clues are added, there is the danger that essential attributes of the actual events will be so distorted they will not provide a validity

[26] In retrospect, this questionnaire item was not totally satisfactory. Estimating the difference between situations rated "exactly alike" and those rated "somewhat alike" is difficult. Because most participants checked several historical situations, it is not clear how much more the simulation resembled the 1914 crisis than, for example, Hitler's ultimatum to Poland. Furthermore, participants were not asked to state when they became aware of the apparent similarity between the simulation and a past event. Some evidence indicates that a messenger suggested the parallel to World War I to several individuals near the end of the M-run. An unstructured question, which provided no list of historical events, might have reduced the number of references to World War I. The alternative situations in the questionnaire item were selected from events mentioned by high school students who pretested the material for the input phase.

[27] The messengers in both runs also completed this questionnaire item. Six out of nine messengers reported a similarity to World War I. It may be that the overall view of events provided by this role increased their awareness of the similarities.

check. This difficulty can be handled by selecting a less well-known situation—providing sufficient information exists on it to construct the necessary inputs. Alternatively, means of comparing the simulation and the actual event can be chosen which are unlikely to be affected if the participants identify the general situation. For example, in the present runs even those participants who saw a similarity to World War I were less likely to know the relationship between perceptions of hostility and capability, or to be aware of the micro events that occurred on July 26, 1914.

Another issue is the distinction between "self-structured activity" and "role playing." Should a simulation participant be required to play the role of a historical decision maker whose characteristic behaviors have been described to him or should the participant be free to structure his own activity? This problem is illustrated by an incident in one of the runs. A participant confronted a decision which he felt his assigned historical figure would answer one way, but he personally would answer another. Should he assume the role or play himself? Given difficulties in constructing profiles of historical figures and problems in matching participants on numerous characteristics, role playing may supplement incomplete personality correspondence. On the other hand, without detailed information one individual's interpretation of the probable behavior of another individual is likely to produce major distortions. Supplying such information would disclose the identity of the historical figure. Investigations probably should be limited to a few personality traits that are part of the natural disposition of the participant. In the event, however, that participants are asked to "play" a historical personality, selection procedures might include some indicators of the empathic qualities required for role playing or acting.

Some explicit reference should be made to the procedural problem of selecting the kind of historical situation that the simulation is intended to represent. All models—including games and simulations—are simplifications of the systems they are designed to replicate. This simplification is achieved in part by completely excluding certain properties of the referent system from the model or by combining numerous detailed elements in an aggregated form. Obviously, a simulation cannot be validated by comparing its output with historical phenomena the model was not designed to represent.

In several ways the effectiveness of the present research for validating the Inter-Nation Simulation is reduced because the 1914 crisis was selected as the historical situation to be replicated. The conversion of the basic time units in the simulation from years to days furnishes one example. With the time units scaled-down, the components in the programmed part of the model acted as constants, though many of them normally operate as variables. In these exploratory runs, then, the part of the simulated world contained in the structured programs affected the behavior of the participants only insofar as they recalled its impact on their previous simulation experience. The 1914 crisis did not provide a means of investigating the programmed relationships as they operate in the usual Inter-Nation Simulation. Although such short-term crises may be excellent for determining the validity of other simulation models, they seem somewhat less appropriate for the Inter-Nation Simulation.

The micro event analysis provides a second illustration of the problem. The variables in the programmed portion of the simulation are broad representations of properties within a nation (e.g., the sum of all human and natural resources, or overall military capability). The events in the micro analysis, however, were at the more specific level of a diplomatic conference or a decision to mobilize ground forces in a certain district. The Inter-Nation Simulation was able to produce events at this level of specificity, but it is not clear that they are produced by the aggregate variables that compose the model.[28]

IV. HISTORICAL SITUATIONS AS AN APPROACH TO VALIDITY

The procedural problems that emerged in the exploratory attempts to replicate the 1914 crisis are not insurmountable obstacles to the use of historical situations as validity tests for simulations. With careful attention to the selection of both participants and the past occurrences to be simulated, these difficulties can be minimized. Broader concerns about the use of historical data for verifying models, however, also must be considered. One challenge to this validity technique is raised by the developers of a simulated underdeveloped economy.

[28] For further discussion of the level of specificity issue, see Harold Guetzkow, "Simulation in International Relations," in *Proceedings of the IBM Scientific Computing Symposium on Simulation Models and Gaming*, December, 1964, Thomas J. Watson Research Center, Yorktown Heights, New York, esp. pp. 264–267.

When a simulation of a particular economy has been formulated, it should be subjected to shocks and exogenous trends like those that have impinged on the actual economy in the past so that its responses may be compared with historical records. It would seem that such tests would yield an independent verification of the model. Actually, in most cases it is unlikely that this ideal can be fully realized. . . . The difficulty, fundamentally, is that the information available in almost all countries is insufficient to establish a model without using all the relevant historical data in the formulation and in the adjustment of parameter values.[29]

The economic model mentioned above is a completely programmed computer simulation and, accordingly, requires more data for establishing its initial values. Nevertheless, the problem is applicable to all historical validation attempts. Clearly, a simulation cannot be validated against the same historical material used to determine its parameters and beginning variable values. Historical replication always will be limited by the record of past events. Not only must there be a detailed account of decisions and actions, but the sequence of events must continue over a sufficient period of time without major changes which require resetting the model's parameters. For this reason historical verification may be more feasible for some types of simulations than others. The smaller the number of variables and parameters which require historical data, the more uncommitted historical material is available for establishing the goodness of fit. Moreover, if the content of the simulation deals with the kind of phenomena that recur frequently in the observable world, then historical validation is more applicable. Thus, a simulation of American judicial decision making is more readily verified with historical data than an operating model of disarmament processes.

Let us assume for the moment that sufficient historical data exist and that the content of a given simulation model permits the use of the validity approach described in this essay. Furthermore, let us assume that a comparison of events in simulation and in history reveals a high degree of correspondence. This correspondence does not demonstrate that the simulation correctly represents the structure and processes that were operative in the historical occurrence. We are speculating on the similarity between the historical and simulated inputs on the basis of the similarity of their outputs. Different relationships among various combinations of properties in the simulation conceivably could produce outcomes like those in the historical situation.

A simulation of the 1960 national Presidential election predicted the percentage of the vote for each candidate—the outcome—with considerable success. The designers of that simulation observe, however, that "it may legitimately be asked what in the equations accounted for this success, and whether there were parts of the equations that contributed nothing or even did harm."[30] Further analysis of the equations in the simulation revealed that the outcome was predicted despite the fact that at least one equation misrepresented aspects of voter turnout. Part of the structure of the simulation was incorrect, but the simulated result still matched the actual outcome. Despite this difficulty, our confidence that the simulation has captured some aspects of the voting process is greater than it would have been if the simulation had failed to replicate the campaign outcome. Confidence in the simulation would increase further as the operating model demonstrated ability to produce outcomes that corresponded with various elections. In sum, the similarity between simulation and historical events can provide at best only indirect and partial evidence for the correctness of the simulated structures and processes that produced the outcome.

Historical material can be used for validity purposes in other ways than by providing events for simulations to reproduce. In the exploratory runs, two validation techniques were tried—event comparisons and hypothesis testing. Although general hypotheses can include events as variables, they also can involve the processes by which events are produced. When events do become variables in a hypothesis, they tend to be more the micro events that occur with sufficient frequency to permit an adequate test. Hypothesis testing, therefore, is less susceptible to the criticisms of event comparisons made above. Verba has argued that difficulties in validly simulating a macro event (such as the outbreak of war) may exceed a model's potential contribution to theories of international politics. "Even if one could design a successful simulation in that respect, it

[29] Edward P. Holland with Robert W. Gillespie, *Experiments on a Simulated Underdeveloped Economy* (Cambridge, Mass.: MIT Press, 1963), 207–208.

[30] Ithiel Pool, Robert P. Abelson, and Samuel L. Popkin, *Candidates, Issues, and Strategies* (Cambridge, Mass.: MIT Press, 1965) rev. ed., p. 64.

might not be useful for future situations, which would not match the historical one in many important ways." Instead, he proposes that "if the situation can be decomposed into many subprocesses, such as communications flows, emotional states of decision-makers, and so forth . . . it may be possible to develop more widely applicable principles that can deal with many political situations."[31] Historical material may prove more useful for simulation validity explorations if it yields frequency distributions of events and processes that can be employed in hypothesis testing.

We have described some important procedural problems as well as two major limitations to using the replication of historical situations as a means of validating political simulations. In addition to event comparisons and hypothesis testing, alternative ways of verifying simulations are available. But they also have substantial liabilities.[32] In part because no one approach can fully establish the correspondence between a simulation and its intended referent system, simulation validity is always a matter of degree. Yet we cannot abandon the efforts to determine the goodness of fit between verifiable empirical observations and our conceptualizations—be they stated as verbal theories, mathematical equations, or simulation models. To improve our estimates of simulation validity, a strategy is required that includes multiple methods for discerning the degree of correspondence. In such a multimethod strategy one approach is historical replication. Until more validation exercises are conducted, it is premature to accept or reject simulation as an important new tool for studying political phenomena.

[31] Sidney Verba, *op. cit.*, 511, 513.

[32] For a survey of various validity approaches and a discussion of their assets and liabilities, see Charles F. Hermann, *Behavioral Science*, *op. cit.*