

VALIDITIES

A Political Science Perspective

{PRIVATE}Francis A. Beer
303-492-7802

Francis.Beer@Colorado.EDU
<http://socsci.colorado.edu/~beerball>

[Department of Political Science](#)
[University of Colorado](#)
Boulder CO 80309

in *Social Epistemology* 7, 1 (1993): 85-105.

© *Social Epistemology*, 1993; Web Site Copyright©, Francis A. Beer

Abstract

Validity is a central legitimating word in the lexicon of political science, suggesting the connection of scientific theory and research with the political world. Validity is constrained by uncontrolled and uncontrollable variance in sampling, context, text, and analysis. Judgments of validity include scientific and political dimensions, implying serious anomalies. These may be partly dissolved by more complex definition and decomposition into multiple, parallel, switchable validities.

Post-modern, post-positivist discourse has focused on "the movement from confidence to skepticism about standard foundations, methods, and rational criteria of evaluation" (Gibbons, 1987; Bernstein, 1985: 3 cf. Radnitzky and Bartley, 1987; Fiske and Shweder, 1986; Rorty, 1979). Validity is a central concern in this conversation. Validity is the fulcrum on which positivist method rests, the connection between the work of political science and the world of politics. Validity sums up the ambitions of positive political science for accurate representation and faithful correspondence with reality.

Professional political science and other social and applied sciences have long been concerned with validity. Textbooks on method contain a section discussing validity; specialized books and journal articles in multiple countries and disciplines center on it. Validity is divided into multiple different categories or types – for example, internal validity, external validity, ecological validity, population validity, situation validity, construct validity, measurement validity, variable validity, outcome validity, conclusion validity, cultural validity, interpretive validity. Consideration is given to factors "jeopardizing" validity, "threats" to validity, sources of "bias," or factors that "confound" or "contaminate" experimental and non-experimental research (cf. Leichsenring, 1985; Andrews, 1984; Calder and Phillips, 1983; Calder, Phillips, and Tybout, 1982; Mook, 1983; Cook and Campbell, 1979; Carmines and Zeller, 1979).

Experimental Validity

Problems of validity are perhaps clearest in experimental research. It is hardly news that there is a striking gap between a sample of experimental subjects and the universe of real political actors. The well known college sophomore problem centers on heteromorphism between the undergraduate subjects that are often the subjects of experiments and the non-undergraduate world that they supposedly represent (Greenberg, 1987; Gordon, Slade, and Schmitt, 1987, 1986; Sears, 1986). Unrepresentative sample is, however, only one of many difficulties. As those who work in the laboratory are well aware, other experimental flaws are equally serious and more complex. These include contextual, textual, and analytical effects.

Context

The context of any experiment is vastly different from the context of life as it is usually lived (cf. Rosnow and Georgoudi, 1986). In many psychological experiments, this disparity may appear to be of marginal relevance. For example, rapid eye movements or reading tasks may seem to be relatively unaffected by their surroundings. Nevertheless, even this trivial stability is less clear than we might expect; research suggests that general stress may adversely affect eye functioning. The experimental subject does not appear at the laboratory wrapped in cellophane. He or she exists in a specific space/time context, which is the frame for a particular mental and physical state. The larger political context may be particularly relevant for political research. For example, English language versions of questionnaires, administered abroad, have the potential to activate anti-American attitudes. These attitudes vary in different countries, according to current political conditions.

The subject also exists within the controlled discourse of the immediate testing context. The simple act of observation – as the Hawthorne experiments and Heisenberg's uncertainty principle have suggested – may change the observed and create a qualitatively different situation. Observation changes the configuration of reality, although in ways that may not be immediately visible if we are not looking for these particular effects or do not yet have the instruments with which to measure them. Observation may take different forms depending on the proximity, evidence, and interaction of the observer. Participant observation and hidden observation mark the opposite poles of these observational dimensions. Subjects may react to observation in different and unexpected ways. Further, subjects may be sensitive not only to the present, but also to the past and future. Prior testing experiences are stored in memory. Information about future testing changes subject expectations and perhaps strategies.

Research on subject reactivity suggests that the experimental environment contains crucial semiotic cues that may activate predispositions. There are administration effects resulting from interaction with the tester. Subjects respond to verbal and behavioral hints, the language and voice modulation of instruction, and responses to questions. Body language, a laugh or a frown, can change the meaning of instructions. Subjects generally try to give the researcher what they think he or she wants. Depending on the experiment, even such apparently irrelevant attributes as the age, gender, ethnicity, or even clothing of the experimenter provide pointers to attitudes and possible desirable responses. The temperature or lighting of the room may subtly influence subjects' orientation to the experimental materials (Harris and Lahey, 1982; Haynes and Horn, 1982).

The mode of the experiment may also be an important variable. Paper and pencil tests may give different results than computer modes (Lee, Moreno, and Sympson, 1986). Equivalency between such instruments and others like video simulations, and various internal and external physiological monitors involved in lie detector technology – temperature, brain waves, heart rate, galvanic skin response, hormones-is unclear. Time and place may also include important variables. Length of interview or interviewing period may induce fatigue or boredom (Herzog and Bachman, 1981). The season or time of day may affect attention and orientation.

Text

Many experiments, particularly psychological ones, rely on textual material. The meaning of such text is stable, but only to a degree.

Much experimental procedure is based on an objectivist theory of language in which words of the text have definite, determinate meanings. This corresponds with the common sense notion that words are independent and transparent. Words have meanings that are reliably denoted in terms of a correspondent reality. Words are the building blocks of questionnaire items. The items also have face validity, explicit meanings that are self-contained, invariant across experimenters and subjects.

This view is strongly influenced by direct experience with other speakers, use of dictionaries and thesauri, and, in some cases, language training. Even if the words do not have an essential foundation of meaning, they are stabilized through convention and in ordinary usage so that intersubjective communication occurs. This stability permits a conformity between the intention of the user and the understanding of the receiver (see Ruhl, 1989).

An objectivist theory of language is consistent with the notion that small variations in words of text can produce significant differences in meaning. Some of these are obvious and within the bounds of normal discourse. Others are more subtle. Compare six items:

1. *President Bush drop{ped} nuclear weapons on the USSR.*
2. *President Bush {said that he} drop{ped} nuclear weapons on the USSR.*
3. *President Bush {[said] that he (would)} drop nuclear weapons on the USSR.*
4. *President Bush {[said] that he (might)} drop nuclear weapons on the USSR.*
5. *President Bush {[hinted] that he (might)} drop nuclear weapons on the USSR.*
6. *President Bush{! D}rop nuclear weapons on the USSR.*

These six items seem successively weaker in the conflictual messages that they convey. Yet the lexical differences are not very large. Item 1 differs from the items 2-5 primarily through the later addition of a spacer consisting of four words "... that he ..." (cf. Fauconnier, 1985). Items 2, 3, 4 and 5 basically vary from each other in only one word. Item 3 adds "would." Item 4 changes "would" to "might". Item 5 changes "said" to "hinted". Item 6 is lexically the closest, yet semantically the farthest from Item 1. It differs only in an exclamation mark, a space, and a capital letter. Yet semantically it is perhaps the most different. It suggests a voice in the wilderness, calling on the President to perform an unnatural act. The patterns of ink, letters and

spaces, and even words on the page are statistically very close, yet minute variance produces recognizable, disproportionate effects.

Even within an objectivist perspective, however, text remains indeterminate to some extent. Such indeterminacy comes from complex verbal combinations and multiple possible meanings. The complexity of verbal combinations may escape the control of the researcher. In the example above, it is not clear that most researchers would pay careful attention to the fine switches between items 3, 4, and 5 – "said that he would," "said that he might," "hinted that he might." Even if the researchers did, most subjects probably would not. It would be easy for a subject to blur fine shades of intended meaning in these items.

The indeterminacy of text goes beyond the amount of care and attention paid by researchers and subjects. As the dictionary suggests, many words have more than one possible meaning. The reader of the text must decode the text, at least to the extent of decoding the word on the page into the appropriate one(s) of these possible meanings. Yet, not all dictionary meanings may be equally accessible to all the actors, especially in a society of multicultural diversity. Further, in an experimental situation, extended oral discourse is not possible to negotiate specific, determinate meaning at each step of the way.

The dictionary partly reflects the meanings of words, but it can not provide a fully determinate solution. Ultimately any dictionary, and all dictionaries, are fully circular. Each word is defined in terms of other words; and the other words, in turn, are defined in terms of each other. At the end of the equation, all words are defined in terms of all other words; the words define themselves in a closed system.

When we look outside the words, we are again caught in a tautology. External reality can not provide completely determinate meanings for words. Words are defined by nonverbal aspects of reality. Yet words are also a part of reality and help to create other parts. Reality is as the words describe it, if and only if we accept the words as accurately describing that reality.

Words are defined by non-verbal life experiences and associated naming. Some aspects of these experiences are collectively shared. Words are, at least partly, the result of our common history and customs. They carry with them a heavy freight of collective historical usage and association. Yet they are experienced, to some extent, individually. The experiences are stored and connected within the individual's partly private lexicon. They are interpreted in the idiosyncratic context of the individual's personal biography and circumstance, intention and perception.

The typographical configuration of the text itself – letters, words, spaces, punctuation, sentences, paragraphs, pages – exerts independent semantic priming effects on subjects. Syntax can also be important. Work in survey research has shown that open or closed questions, affirmative or negative expression, can activate pre-existing response sets in subjects or elicit "non-attitudes" (Schuman and Scott, 1987; Brody, 1986; Schuman, Ludwig, and Krosnick, 1986; Norpoth and Lodge, 1985; Bishop, 1984; Taylor, 1983). A whole panoply of related effects – consistency, framing, proximity-suggests that this is just the tip of the iceberg (Fleishman, 1988; Jackson and Jacobs, 1983; Schuman and Ludwig, 1983; Cicourel, 1982; Kahneman, Slovic, and Tversky, 1982; Sudman and Bradburn, 1974).

As Edelman (1988b: 1336; 1988a) suggests, "belief in the objectivity of a form of language is a lethal pitfall for researchers." A more sophisticated interpretation of meaning, derived from the work of writers like Wittgenstein, Saussure, Foucault, and Derrida suggests that each word is a node within a complex, inter-related network of signifiers and signifieds. Words appear to exist discretely, following in serial sequences, separated from each other on the page by blank spaces. But language is deep. Words are multidimensional – densely interconnected in linguistic systems: embedded in culture, laden with assumptions, evoking complexes of feelings.

Modern theory in such diverse areas as literary criticism, semantics, and artificial intelligence clearly suggests that text is at least partly indeterminate. Items are specific types of stimuli that produce responses. If other factors remain constant, the responses generally may fall within bounds defined by a common culture and usage of the experimenter and the subjects. This gives the impression of stable meanings and understandings, but the impression can be very misleading (cf. Herring, 1987; Lakoff, 1987; Norris, 1987; Small, Cottrell, Tanenhaus, 1988; Scheffler, 1979; Neisser, 1976; Fodor, 1975).

Peripheral Text

Experimental text may be divided into core and peripheral text. The discussion so far has centered on the core text, which consists of those written questions or items on which subjects are asked to focus their immediate attention, and to which subjects are asked to respond. Peripheral text surrounds the core text. It is not the focus of direct, immediate subject attention and response. As we have already seen, the surrounding context partly determines the meaning of text, and part of the surrounding context is textual. Questionnaire items are psychological stimuli. The priming effects of peripheral text act upon the core text that the subject is supposed to be processing. Some of the priming effects are anticipated and even included in the experiment, but others may be unintentional and unknowable.

Survey research has been particularly sensitive to the effect of order on meaning (Abramson, Silver, and Anderson, 1987; Bishop, Oldendick, and Tuchfarber, 1982; Sigelman, 1981; Bradburn and Mason, 1964). Serial order effects can occur within items – the sequence of words – or between items, questions or responses. We can see this clearly from the six items above. The word order effects are obvious; the item order effects less so. Nevertheless, if we imagine that the six items above are events that constitute a narrative sequence, then the effect of item placement becomes clearer. If we were to place item 6 at the top of the display, and ask subjects to describe the story behind the text, they might tell us about pressures on the President to drop nuclear bombs, Presidential compliance with those pressures, and Presidential follow-on statements. The story does not physically appear in the text; it is constructed in the mind of the reader. Different chains of items or events may suggest different stories. Responses may also vary with items that are asked but not answered – "don't know," "no opinion," "not sure" – or items that are not asked (Duncan and Stenbeck, 1988). If the experiment involves several rounds of activity, successive turns create different contexts that also impinge on the responses in different ways. The effects of textual order have been used as a powerful literary device, for example in the short story surprise endings of O. Henry.

An Experimental Example

Experimental results clarify the mixed determinate/indeterminate nature of larger segments of text. Beer, Healy, Sinclair, and Bourne (1987) reported an experiment modeled on the Falklands/Malvinas crisis between Britain and Argentina. Schema theory from cognitive psychology had suggested that violent priming should trigger subsequent conflictual productions. Sixty American college students were given background items, including a personality scale measuring dominance-submission; an initial scenario modeled on the Falklands/Malvinas crisis; and one of two vignettes recalling the slaughter of World War I, Allied weakness before World War II. A control group received no priming vignette. The subjects were then asked to choose from an inventory of possible cooperative-conflictual foreign policy acts in five sequential rounds. Both war vignettes produced more subsequent conflictual foreign policy act choices for subjects with dominant (i.e. aggressive) personality attributes and more cooperative choices for subjects with submissive personality attributes. A second, follow-up study adapted and replicated the first experiment a year later in South Korea (Park, 1988). One hundred South Korean college women were given English and Korean language versions of the original materials. The replication established cross cultural and cross-gender stability of the earlier U.S. results.

At the same time, the two studies suggest that the textual understandings of experimenters and subjects may be incongruous. Subjects may see the same materials as being different. At the beginning of each experiment described above, all subjects read the same initial scenario, which the authors had modeled on the Falklands/Malvinas crisis. Twenty seven of the sixty subjects in the American experiment identified it as the experimenters intended. The results for these subjects were different than for those did not recognize the situation. The vignettes only tended to activate underlying personality dispositions for those who did not recognize the situation. Priming of personality did not occur for those who clearly saw a representation of the Falklands/Malvinas crisis. For the group that did recognize the real-world referent of the experiment, there was little difference between high and low dominance groups in the cooperative-conflictual mix of actions that they finally chose. In the South Korean replication, thirty two of the hundred subjects also recognized the situation. As in the first case, such cognitive activation suppressed personality effects that appeared in other subjects; cognition replaced personality for subjects who recognized the action scenario.

South Koreans who recognized the situation recalled a different case than Americans. Instead of the Falklands/Malvinas, they believed that the experiment represented the situation of Tokdo, a small set of islands between Japan and Korea. South Korean recognition thus produced effects on other experimental variables similar to those of the original American experiment, but the cognitive referents were completely different. The same text, the scenario, stimulated the same response, recognition. But, in different contexts, the same response had very different meanings – the Falklands/Malvinas or Tokdo. These meanings were revealed only through subsequent items, asking whether subjects recognized the scenario and what they recognized.

Subjects may also understand different materials as being the same. The experimenters had developed the two original vignettes of World War I and World War II to prime subjects about the costs of war and the costs of not going to war. Each vignette was supposed to convey a dramatically different message. The Korean adaptation added two new vignettes – the costs of the Korean War and the mistake of the Acheson line – to duplicate this distinction in a more familiar Korean context. In neither experiment did subjects seem to discriminate between the two

types of messages. In both experiments, students reacted to all of the priming vignettes in virtually identical ways. The main priming cue, the principal message that they seemed to receive, was war. Subjects were not significantly affected by the flavor, directionality, and difference of the war experience – the costs of action and inaction. The message that the experimenters had intended to send was not received as the experimenters had intended to send it. Different texts stimulated similar responses; the two vignettes apparently primed equivalent psychological schemata. From this perspective, the alternative versions might be considered to be non-discriminant and collapsed theoretically to one vignette.

From these results, it is evident that intended face meanings of different texts do not translate directly into cognitive schemata. Separate groups perceive texts in different ways. Authors and readers, American and South Koreans understood both the introductory scenarios and the priming vignettes quite differently. The exact causes of these differential perceptions are not clear. For example, it is easy to believe that cultural factors determined whether or not individuals recognized the initial scenario as the Falklands/Malvinas or Tokdo. Nevertheless, two South Koreans undertook the adaptation of the original experiment for the South Korean context, and they did not make this connection. Further, it is not immediately evident how two such apparently different vignettes could stimulate similar responses, or what other kinds of vignettes might prime different schemata and give different results. The full dynamics of subtextuality, intertextuality, and supertextuality remain to be explored.

Analysis

A third major set of uncontrolled factors occur in analysis – theoretical setting, identification of key variables, measurement, mathematical and statistical manipulation.

A paradigm, a body of scientific literature and theory, sets the supertext for each experiment. Experimenters locate their research in preceding work. This justifies their choice of variables to examine and helps them to nest their small scientific story within a larger plot. We have already seen how alternative item order can alter the narrative that subjects perceive. In the same way, alternative theoretical story lines can reinterpret experimental data. The experiments described above were conducted and interpreted from the perspective of schema and personality theory, exploring the way in which scenario and vignettes activated underlying cognitive representations and personal predispositions. The same data, however, can also be interpreted from a rational choice outlook. Different personalities might be coded, in rational choice theory, as substitute sets of preferences. From this vantage point the experimental results give us a surprise ending. Rational choice theory assumes stable preferences. Yet in this set of experiments, preferences did not remain stable. They were buffeted by multiple stimuli, switched on by the priming vignettes or switched off by scenario recognition.

Just as different variables can be measured by the same data, the same variables may be assessed through different means. Alternative sets of items may produce apparently similar effects; different scales may profess to measure the same underlying dimension. For example, the CPI, 16PF, and MMPI personality inventories may be cross-validated to give generally similar results in many settings, but they are surely not completely interchangeable (Morey, Blashfield, Webb, and Jewell, 1988; West, 1986. cf. Krosnick and Alwin, 1988).

Some statistical methods give differential weight to extraneous factors, for example size of the sample. Standard methods create a world whose assumptions and attributes have varying applicability in different concrete cases: independence, normality, homogeneity, continuity, linearity, homoscedasticity, stationarity, unrelated error. There are indeterminacies of mathematical formulation and application – multiple alternative choices for cross sections of time and space; variable transformation; model identification, specification, and estimation – that shape final results.

Non-Experimental Validity

Problems similar to those of experimental research appear in non-experimental research as well. Symmetries are clearest in other forms of research that use similar question-and-answer methods – participant observation; unstructured or semi-structured oral history; depth interviews; standardized testing; survey research.

Survey Research

Survey research probably bears a resemblance to experimental research, but it is usually differentiated as being closer to the real world. Schuman and Presser (1981: 313) suggest that "psychological theory has developed largely through a kind of controlled laboratory experimentation that is usually far removed from the encounters with ordinary people that characterize surveys." Survey methods usually include sophisticated techniques of sample selection – multiple stages, stratification, clustering, for example – attempting to represent specific larger populations, to replicate their attributes in microcosm. The results of survey research are then generalized to these populations, assuming common underlying attributes and dynamics. No matter how clever and complex the sampling procedure, however, there exist inevitable residual differences between the sample and other populations. These differences include developmental, social, economic, cultural, or political dimensions. Advances in sampling technique allow us to eliminate and narrow some of these differences. In spite of such progress, it is impossible to say a priori what effect the remaining differences may have on the results. Though we may assume them away as irrelevant, trivial, or insignificant, we can not specify their values in advance. As in the case of experimental research, these disparities open all generalizations across samples and populations to initial question.

Survey research, except for its concern with sampling, is similar to experimental research. As the analysis above suggests in several places, survey research contains the same instabilities. In spite of attempts at technical fixes-e.g. buffers between particularly important items, splitballot questionnaires – there are multiple uncontrolled and uncontrollable contextual, textual, and analytical variables. As practitioners of these techniques are well aware, all question-and-answer methods are subject to distortions of attention, perception, articulation, memory. Neither experimental nor survey techniques are well suited to evoke or capture the dynamics of real-time group discussion and decision, or the stress and emotion of serious personal or public situations (cf. Lodge, McGraw, and Stroh, 1989; Gerstein, Luce, Smelser and Sperlich, 1988; Schuman and Scott, 1987; Schuman, 1986; Turner and Martin, 1984).

Content Analysis

Another set of procedures does not elicit words, but takes them as given, for example, unstructured analysis of documents like those produced by branches and agencies of government and individual authors. The most systematic form of this work is content analysis. The sampling, contextual, textual, and analytical problems discussed earlier are present here as well. A representative sample of the documents may be taken as the basis for research, but the documents themselves hardly mirror all dimensions of the behavior that they may purport to report or reflect. They are produced by specific authors in a particular setting for specific purposes.

One genre of content analysis, events data analysis, involves collecting, coding, and analyzing political events from press sources. Researchers begin with samples of newspapers that, in turn, provide samples of events. The events are created through multiple filters – actors, reporters, coders. The chain grows longer when the reporters talk with other observers, witnesses or commentators, or if their own work is heavily edited before publication (cf. van Dijk 1984; Pollock, 1981; Boorstin, 1985). The actors, reporters, and coders are not independent as they communicate and consult with peers or superiors. Definitions of some events or variables depend on definitions of other events or variables. Further, the actors, reporters, and coders come from particular national, cultural, linguistic backgrounds and training that mold their perspectives.

Early content analytic method placed great emphasis on intercoder reliability, the multiple replication of results as a guarantor of validity. Growing experience and sophistication has made it clear, however, that such reliability is a function of background variables such as those just discussed, as well as coding categories, rules, and error checking. The final events data is thus hardly a reflection of an "actual" event. Like other data, it is an artifact, manufactured by standardized means, and hardly a natural product. Further, like the other data forms discussed earlier, the final product is influenced by sampling, context, text, and analysis.

Aggregate Data

Aggregate data allow us to watch what people do rather than what they say. Aggregate research has different aims and uses different methods, but it implies similar problems. Instead of sampling individuals, aggregate data analysis may rely on samples of larger social units. Like the events data, they are not simply facts, but more importantly artifacts whose creation is more deeply hidden. Aggregate data go through several stages in the production process, and each of these stages involves omission and inclusion. For example, income inequality is an important statistic, particularly in the study of economic development. At the very lowest level, there is an equivalency problem centering on the relation of the basic unit, money, to the activity actually being undertaken, for example growing fruit. This problem is similar to the contextual and textual problems discussed above. The monetary unit translates and transforms this activity from a qualitative to a quantitative phenomenon, takes it from the world of individual experience, processes it, and enters it into the homogenized ledgers of social exchange. Through this process, growing fruit becomes an activity comparable to, and, therefore like, all other activities – flying airplanes, ski racing, or murder for hire. This numerical transformation does more than simply condense or lose information. It translates and exchanges one experience into another. It strips away uniqueness or variance in production and consumption; it leaves behind specificity of time and place. The peasant caring for his date trees by hand in Cyprus of 1910 is the same as the

American agribusinessman harvesting wheat by combine in 1990. Conversion in the other direction – from money into experience – is hardly easier. Even leaving aside questions involving exchange rates and purchasing power, the same commodities are not evenly available at different times and in different places, nor do they have the same social and individual meanings. Louis XIV could not have modern plumbing at the palace of Versailles. Apparently similar activities of two different people – a Cambridge philosophy professor and a Tunisian businessman driving a red Ferrari – may have very different social and individual meanings.

Once the conversion of activity to money is accepted, new problems arise. One of these is similar to the sampling problem discussed above. It centers on the accuracy of income reporting, the extent to which the income reported faithfully represents the income received. The peasant growing dates will not report as income the increase in value of his land, nor what he, his family, or his friends eat themselves. What he sells privately or at market will almost certainly suffer severe attrition before it enters the formal government bookkeeping system. At upper income levels, there will also be under-reporting of income. Land earnings will be hidden behind a curtain of silent appreciation. Individual earnings can similarly be redirected and sheltered behind the corporate veil. Foreign earnings may be protected by bank secrecy. And earnings from bribery, kickbacks, and other improper or illegal activities are of course outside the normal reporting matrix. What numbers are reported must then be further processed by the government, with its own interests. Internal and external practices and pressures may mold the outcome. Nations generate statistics using disparate accounting methods. Different countries have different laws, with different penalties, for income reporting. Bureaucracies vary in sophistication and resources. Cultural norms for collection and payment differ. The regimes of socialist countries hold power on the basis of claims to foster equality. Developing countries must justify various forms of assistance to international agencies and donor countries. These data, like the others, are artifacts, subject to abstraction, distortions, and omissions.

Non-economic data have the same problems. An extensive body of data on wars has been collected and refined over a long period by a dedicated group of scholars. The war data, like the economic data, are very different from the phenomena from which they are derived. The experience of war was recorded in earlier times by poets. The *Odyssey*, the *Iliad*, and the *Chanson de Roland* are epics that have come down to us through history. The poetic form persisted through the 19th century, as in Crane's "The Red Badge of Courage" and Tennyson's "Charge of the Light Brigade." In modern times, war has been recorded through a diversity of narrative forms, including biographies, novels, and journalism. Nonnarrative forms have replaced words with numbers. This has created an entirely different phenomenological reality, replacing the worlds of writers with those of accountants and book-keepers, mathematicians and engineers.

These war data can be misleading, even within their own terms of reference. The identities of the wars and their participants are basic to this type of formulation, yet the boundaries are fuzzier than the innocent might suspect. Wars are more like clouds than tables; they do not have pre-existing sharp edges that give them clear boundaries. Wars can be defined according to different criteria – participants, length, casualties, formal declarations – and the criteria determine where we stick the label war and where we do not. Only once we accept a particular definition, with its underlying presuppositions and assumptions, do wars become commensurable.

Wars are not independent of each other, and this interdependence further blurs identity-boundary issues. World Wars I and II are usually treated as separate wars, but many analysts have suggested that they were actually two separate segments of the same war. In the same way, wars are not independent of other phenomena to which they may be related. For example, crises may be defined in the framework of subsequent wars. If the war had not occurred, what preceded it might not have been defined as a crisis. The error may be compounded if we wish to examine the extent to which crises predict wars. Peace and war may also be conjointly defined and related with similar problems.

Wars do not necessarily have crisp dates on which they neatly begin and end. The same definitional problems surround war initiation and war termination. Dates of entry and exit are, to some extent, arbitrary markers of ill-defined points in the process; belligerents often slide slowly into and out of war (Beer and Mayer, 1986).

The identity and stability of war actors are not constant. National names – Austria, China, France, Germany, Italy, Japan, Russia, Turkey, United States – have meant very different things across time and space, even phonologically and orthographically. Under the names, little has remained absolutely constant.

Casualty statistics are fraught with dangers. The most defensible numbers are direct battle casualties, but the Vietnam experience clearly taught the distortion of body counts. Contemporary data archives may contain multiple estimates of battle casualties, but these estimates are not necessarily independent. A study of the genealogy of the data may trace their derivation back to the same earlier questionable source (Beer, 1974). If direct battle casualties are problematic, indirect casualties are almost impossible. Famine and disease rival weapons, and in many cases supersede them, as causes of related death. Nevertheless, their relationship to war is not necessarily clear or stable, and the statistics are even more ephemeral. Even where there is general consensus on the statistics, statistical homogenization creates illusions of comparability. The same casualties may have very different meanings, depending on different actors in different situations. Korea and Vietnam had roughly the same casualty levels, but they were very different wars.

Analysis

We have already considered some dimensions of analytic indeterminacy in the earlier discussion of experiments. Other aspects occur in non-experimental research. For example, Lewis Richardson and some of his followers have described wars and domestic violence as basically random occurrences, whose distribution in time and space conform to stochastic statistics like the Poisson distribution. Recent advances have created hybrid, combined Poisson regression models (King, 1989). Yet it seems paradoxical, even bizarre, that supposedly random events should conform to a deterministic formula or model. Further, if they do conform, why should it be to one formula rather than another? Subsequent researchers have indeed applied other different formulas to random events, for example binomial or Weibull distributions. Superordinate randomness could easily apply not only to the events, but also to their distributions.

The evolution of modern applied mathematics into areas like fractals and chaos suggests that there are virtually infinite degrees of freedom, turning on possible distributions and equations, lines and curves to which data may conform, creating illusions of shape and movement (cf.

Arrow, 1988; 280-281). The formulas, in turn, can be tested against data divided by an innumerable number of possible cut points or disaggregations in time and space. Goodness of fit is difficult to compare across models beginning from different premises and taking different forms. Some types of models exist in a kind of free zone where problems of empirical fit are, to some extent, finessed. For example, formal models of rational choice that rely on individual expected utilities, or more complex interactions of game theory, can be extended to external political situations using very relaxed testing criteria. In some cases, the extensions are largely metaphorical (Beer, 1986). Such models proceed as if their assumptions were true; they are properly limited to possible worlds where this might actually be the case.

One answer to the indeterminacy dilemma is reliability through replication. A strong form of replication is multimethod research design, for example, "comparison and calibration of results obtained by longitudinal surveys, randomized field experiments, laboratory studies, one-time surveys, and administrative records" (Gerstein, Luce, Smelser, & Sperlich, 1988: 174. cf. Arlin, 1982). In spite of its intuitive appeal, replication is not the final solution to the problems we have uncovered through dissection. It will not lead us to universal commensurability. Reliability supports validity, but the two are not identical. The claim to validity is presumed to be stronger if results are relatively robust and insensitive, if findings are reproducible or converge across research designs, samples, contexts, and designs. For example, validity claims were enhanced in the experiments on war cues cited above, when the results from the South Korean study, with a dramatically different sample from another culture, virtually duplicated the American outcomes. We ask "what if..." and "if...then" comes back the same. Nevertheless, simple repetition falls into the trap of naive verificationism. It applies an algebraic, additive solution to a geometric, combinatorial problem. Replications in different contexts do not measure all the unmeasured or unmeasurable variables; they do not unravel the tangled skein of multiple causality; they do not stabilize analytic indeterminacy. Even if all of these problems were solved, and one could be absolutely certain that some hypothetical finding were absolutely invariant in all forms across all contexts, the invariance would apply only to the past. We should have succeeded at retrodiction, but we should be as far as ever from prediction. We should know the worlds of Christmas past and present, but from this we still could not surely foretell the worlds of Christmas yet to come.

Scientists have long dreamed of a demon that would simulate the mind of God. This demon would know all phenomena of the world and their minute, inter-related causes. A modern version of this demon would have to know the same things for a multitude of possible alternative, parallel, simultaneous worlds, subjective and objective, differing in an infinity of ways, at all points in space and moments of time. And it would have to know their rules of activation and valence, transformation and transition, privilege and power-all at the same time (cf. Casti, 1989; Wolf, 1988; Lewis, 1973). In spite of evolutionary progress, science and political science are closer to the beginning than the end of this road.

Science and Politics

Scientific Validity

This paper began with a conventional view of validity as found in political and other social and applied sciences. This view, in its ideal form, suggests that political science can be a transparent window onto politics. Through it political scientists are separated from, and look out onto, political behavior in the wider community. The more valid is the theory, method, and research,

the more accurate is the reflection or the view. This perspective assumes applicability. It supposes that concepts correspond with, that theory and research accurately map, reality although in simpler, sparer, minimalist form. Following an idealized positivist example of natural science, and in the footsteps of such behavioral pioneers as Charles E. Merriam, the value free "pure science" of politics ultimately aims to provide objective "certain knowledge" through "rigorous scientific method," and the professional "power of prediction and control" (Ross, 1991: 391-400).

This view may be abbreviated as V_{ss}, to designate strong scientific validity. Very few sophisticated contemporary political scientists would probably subscribe to this doctrine in its simplest, most extreme, naive form. We know that there are validity problems. In most cases, we have very little idea of the proportion or direction of variance introduced by a multitude of unmeasured or unmeasurable variables. Samples are not isomorphic with larger populations. Research contexts are different from the situations of political life. Text is not a completely transparent medium that conveys intended and perceived meanings of researchers and subjects. Language and numbers, statistics and general abstract models are too spare, parsimonious, stingy, to carry the rich particularities of specific political situations. Multiple, uncontrolled degrees of freedom lead to a combinatorial explosion of samples, texts, contexts, and modes of analysis, an array of infinite dimensions and possibilities. The expected is never exactly the observed. Political science has long passed the age of innocence. It can not maintain the strongest scientific form of validity because the problems that have emerged are too serious and too numerous to be ignored. Defects of experimental and nonexperimental methods cast serious doubt on validity as an absolutely attainable operational scientific criterion, and on the realistic possibilities for comprehensive validation research (Sussmann and Robertson, 1986).

Standard political science adapts according to "the maxim of minimum mutilation," which enjoins us to "disturb overall science as little as possible" (Quine, 1987: 142). Political science has thus evolved a weaker, more complex form of scientific validity, V_{sw}. This weak form of scientific validity, nevertheless, rests upon the same foundation as the strong form. Political science still aspires to the fidelity of truth. Validity is commonly used in the scientific lexicon as a synonym for verity. Cook and Campbell (1979: 37) define validity as "the best approach to the truth or falsity." Valid research supposedly describes real behavior in the political world. More complexly, Campbell (1987: 85) rejects "pragmatism, utilitarian nominalism, utilitarian subjectivism, utilitarian conventionalism, or instrumentalism" in favor of "hypothetical critical realism". Nothing is perfect. Absolute truth is unattainable. We use a pragmatic standard of evidence; we do the best we can to carry the burden of proof. Bias and its cousins mar the view. We satisfice. We act "as if" our results were true beyond reasonable doubt or within the balance of subjective probabilities; but our confidence level is less than 100%. We try to push back the limits (Singer, 1988; cf. Fienberg, 1989: 191-205).

Political Validity

Validity remains problematic even in this weaker form because it still rests upon the concatenation of validity and verity. In their classical sense, validity and verity are far from synonymous; they connote quite different ideas. Verity derives from the Latin *veritas*, meaning truth. Validity comes from the Latin *validus*, meaning powerful or strong.

Validity, from its origins, thus appears as a political word and a political concept. We may designate this view of validity, political validity, as Vp. The political harmonics of validity are clear both from its genealogy, and also from usages in other contexts. Thomas Hobbes (1958: 120) informs us that "the validity of covenants begins not but with the constitution of a civil power sufficient to compel men to keep them." The validity of the written word itself originally derived from the power of the medieval ecclesiastical or secular authority that heard oral speech and wrote it into formal symbols. (Illich and Sanders, 1988). Today, hearings still lead to writings that provide the validity of law. The United States Senate, through the powers granted it by the Constitution, holds hearings that validate an international treaty; the Court, acting under civil law, hears a case and validates a will. The etymology of validity helps to distinguish and distantiate it from verity; historical context helps to reconstitute validity with a specifically political content. Validity, from this angle, is not just about the political world, but is also a part of it. Our assessment of validity, then, depends not only on the fidelity of the view, but also its power as derived from the power of the viewer.

The claim to validity thus reflects the political power and authority of the broader community. The authority of political science derives from the political authority which is its context, the solidification of secular authority and science after the medieval collapse of the church. The currency of political science is backed by the gold of the polity that supports that political science and directs its aims and methods (see Mukerji, 1990; Brown, 1984; Jagtenberg, 1983; Blume, 1974). The coin of validity is stamped with the approved seal of expert judgment by the scientific elect. Validity is conventional, based on trial and error, the collective guess, the best bet, the informed hunch of the scientific community. Validity, seen in this light, is the name we give to the negotiated product of our scientific satisficing or sufficiency. Political and scientific laws support the sciences of both Sherlock Holmes and Herbert Simon (see Eco and Sebeok, 1988).

Vp also suggests that validity reflects a narrower politics, with a face that turns inward toward political science. Judgments of validity are based on accepted professional standards. In this guise, validity is a mask for scientific politics, a rhetorical trope, evoking widely shared ideologies and orienting actors in the politics of political science. Scientific politics, like the Church Latin of the Middle Ages, flows in a formal pattern of discourse. Its main direction is from the center out, from the top down. This is quite different from the discourse of natural language, a world of subjects that are fully human in all directions, as follows:

Language is ... the possibility of subjectivity because it always contains the linguistic forms appropriate to the expression of subjectivity, and discourse provokes the emergence of subjectivity because it consists of discrete instances. In some way language puts forth "empty" forms which each speaker, in the exercise of discourse, appropriates to himself and which he relates to his "person," at the same time defining himself as I and a partner as you. The instance of discourse is thus constitutive of all the coordinates that define the subject (Benveniste 1971: 236-237).

Political science discourse replaces this natural intersubjective equality with its own graded political structure. Vp implies that validity points towards, and helps to define, the center of scientific power, the mainstream, the politically correct. The members of the central network, the minimum winning coalition, control, in variable degree, the levers of professional inclusion: the prestige of major institutional affiliations; the resources of the granting agencies; access to

central professional associations, presses and journals. Elite consensus defines reality. Validity indexes political position, economic reward, and social status. Those at the periphery suffer validity anxiety (cf. Freund, Kruglanski, and Shpitzajzen, 1985; Devereux, 1967). The psychology of invalidity centers on the fear of exclusion and abandonment, being left outside the magic circle of privilege and survival. There are penalties for non-conformity or dissent. Valid research may or may not be closer to being true, but it is closer to being accepted as true (cf. Hallett, 1988).

Validity in this light also assumes a hierarchical researcher-subject structure. The political scientist constructs the agenda and the terms of the interaction; he or she extracts and shapes the data; he or she forges the results, through advanced technology, into new and significant analytical products. The subject has an I-Thou relationship with the scientist. Like the obedient child of an earlier time, the subject speaks only when spoken to, and is otherwise invisible and mute, a black box or a black hole.

In spite of validity's inward dimension, political science is not self-contained within the political science community. It draws power resources from the entire polity and must finally increase the power resources available to that polity. Political science can easily justify itself by its real virtues. It sees, hears, and reads a common political experience and writes it into a new collective political language. It defines, codifies, combines, and processes phenomena in different ways that re-form our thinking and generate alternative conclusions. At the same time, it is counter-productive to pretend to what can never be achieved. Excessive claims on validity are losing propositions. Political science can not support them scientifically or politically. They draw on surplus political capital and oversell the endeavor. They advertise more than political science can deliver, raise expectations that it can not possibly fulfil. Such large promissory notes overextend the whole enterprise; the scientific reserves are too skimpy to pay the political debts (cf. Glenn, 1989).

Anomalies

Validity thus has a Janus face. Scientifically Vs looks outward for an objective, real, true world. Politically Vp looks inward to those who define the view onto that outside world. This Janus-faced quality is a characteristic of natural as well as social and political science. Latour (1987:4, 23) suggests that there are "two sides, one lively, the other severe, of a two-faced Janus. 'Science in the making' on ... [one] side, 'all made science' or 'ready made science' on the other." The external view is observation from a distance in which the subject of scientific inquiry is placed in a black box, "away from its conditions of production." The internal view is closer to the scientific process. It opens the black box to show the way in which human beings socially construct scientific knowledge.

This bidirectional context implies anomalies that go beyond the simple innocence of Type I and Type II error. There are tautologies. Vp implies an underlying circularity in the definition of truth and reality. What is true is not necessarily true independently of the observers. Truth is what the leaders of the political science community say is true, and what the leaders of the wider political community are willing to accept as true. The reality we have created is not valid in its own right. The politics that create and shape our reality also establish criteria for what we accept as scientifically valid. It is as our science describes it, if and only if the scientific and general

communities accept the results as fitting within the political lexicon, as accurately describing the reality, as valid.

There are further unresolved inconsistencies. Both political and scientific versions of validity suggest that political scientists are concrete human teleological agents, makers of the glass and observers through it. Thus, under both V_{ss} and V_{sw}, political scientists independently create research tools, test hypotheses, weigh evidence. We use free will to make decisions and choices, to control our domain. Our ends determine our means. Sometimes we also see our subjects in this light, in a standardized, testable image of political participation, educational achievement, psychological normality. Yet, paradoxically, the scientific view often sees others quite differently – as objects of our science. They are nodes in a matrix of abstract variables, caught in a determinate, repetitive, general political science where cause and effect, stimulus and response, cues and priming coexist in an organization of stable predictions. Past political behavior of these objects often determines their future political behavior; their means determine their ends. Our own codification of that past behavior supposedly allows us, or will allow us, to predict scientifically the future behavior of political automata.

It is more pleasant to look at others than ourselves in this glass, to see them in our image, not ourselves in theirs. As the authors of the story, we assume that we are freer than the characters we create and observe. We are outside the system of abstract, impersonal variables and not fully reducible; our science is not behavior of the same kind subject to similar constraint and observation. We tend to believe that we are not fully programmed, that our past scientific behavior does not necessarily determine our future science, that we as scientists cannot predict our prospective scientific behavior and our scientific reactions to that eventual scientific behavior. We may guess from clues; we are not primed by cues. This position is dramatically inconsistent. If our science means anything, it must apply to us as well as them, inside as well as out. Political science must be simultaneously the science of external politics and a definition of our own situation.

Validities

A first step toward more complex understanding is gradually to recognize and articulate such existing paradoxes, using them to begin to define a more complex lattice of linked definitions of validity, as is being done for other political words (cf. Flathman, 1987). This may be conceived either as an extension of analytic and scientific method or as a part of a strategy that Cook (1985) has labeled "postpositivist critical multiplism." We see computers simultaneously through different glasses: hardware and software, display and cognition, components and systems, machines and humans. The patterns of light on the video terminal screen have many alternative meanings, all valid in different ways. Understanding of validity anomalies may be advanced by similar decomposition. Multiple validities portray different realities and reflect differently on each other. Alternative frames of reference reveal different possible meanings and inconsistencies.

Combinations

The family of validities includes scientific and political validity, in strong and weak forms. These forms may be joined as general attitudes in varying combinations in different communities and individuals. For example, a plausible modal stance in United States political science might

include weak versions of both scientific and political validity, V_{sw}/V_{pw} . This implies some degree of belief in scientific method as the basis for the study of politics, the independence of evidence, the possibility of uninvolved prediction. At the same time, there is a recognition that we can not attain absolute truth, an appreciation of the subjective, artifactual dimension of our knowledge and the dynamics of symbolic interaction. Other combinations – V_{ss}/V_{pw} , V_{sw}/V_{ps} – give heavier weight to one component and lighter weight to the other. A particularly interesting combination is V_{ss}/V_{ps} where belief in both major forms of validity is held equally strongly. The dynamics of cognitive consistency suggest that this is probably quite rare, if it exists at all. Nevertheless, it is logically possible for science, as for a court of law, to assert that judgment can be simultaneously strongly true and strongly political.

Our assessment of specific validities does not necessarily remain stable over space and time, even within the same individual. As we read and write, observe and create, we differentially validate diverse aspects of our environment. Sometimes we draw attention to the process; Dr. Johnson replies to Bishop Berkeley's subjectivism by kicking an objective stone; Hume depresses himself with extreme scientific skepticism, then reassures himself by eating a good dinner. At other times we unobtrusively shift validation and meaning for rhetorical effect in the game of language. Explicitly or implicitly we assign relatively strong scientific validity to some theory, research, or experiences, weaker scientific or political validity to others. Some specific findings of psychological laboratory experiments and survey research, for example, may be used to call into question the general results of those and other lines of research.

Further, as we saw in the earlier discussion of textual and contextual variance, possible underlying definitions and interpretations wobble almost imperceptibly in the dense seas of words and circumstances that surround them. Countless configurations of validity exist with other language and social forms. We can easily lose ourselves in the possible combinations. The simplest case is probably plain bivariate nesting – the scientific validity of political validity, the political validity of scientific validity. In a more formal vein, we might think of a possible modal person with a belief structure V_{sw}/V_{pw} . This person might have views on persons with other belief patterns which we might specify as V_{sw}/V_{pw} (V_{ss}/V_{pw}), V_{sw}/V_{pw} (V_{ss}/V_{pw}), V_{sw}/V_{pw} (V_{sw}/V_{ps}), V_{sw}/V_{pw} (V_{ss}/V_{ps}). The partial expansion in even this very limited case only begins to suggest the complexity of verbal interactions as we approach the limits of ordinary language and understanding.

Linguistic modifiers and specific subjects of discourse are contexts that geometrically expand permutations of meanings. Some modifiers produce compounds – population validity, ecological validity, construct validity – that may make more sense in a V_s framework. They imply a perspective that is more scientific, mechanistic, deterministic. Other word combinations – cultural validity, situational validity, interpretive validity – imply a V_p interpretation, more political, intentional, voluntaristic. The full factorial expansion of the formal possibilities of each modifier, however, would reveal a wealth of plausible combinations.

It has been suggested that alternative meta-models may marry mechanistic and teleological forms of validity in different possible proportions and circumstances. Higher order models might systematically relate research subjects and objects, observers and actors alike, specifying their degrees of freedom and determination in different contexts and interpretations, as well as the dynamics of their interaction. This task would be complicated by the fact that validity constantly

changes shape as we search for workable synthesis and integration at a higher level of abstraction in multiple different contexts (cf. Washington and McLoyd, 1982).

Borrowing from Niels Bohr's complementarity principle in physics, it may be possible to relate discontinuous and mutually exclusive perspectives of "observation and definition" as one slips "from one plane of objectivity to another" and between "different conscious experiences" including "consciousness as the object of description and the subject consciousness which experiences it." This depends on "different placing of the section line between the observing subject and the object on which attention is focused." One possible path is development of "the analogy of the situation to multivalued functions of a complex variable." In this analogy, "each complex number can be represented unambiguously as a point on a two-dimensional plane, but the multivalued function may have potentially an infinite number of values, each a complex number, for each value of the independent complex variable." (Folse, 1985: 113, 175- 177, 180. See also Bell, 1987a; Rasmussen, 1987)

Distance and Proximity

Variants of validity earlier appeared similar to different combinations of shadings of the faces of Janus. One formulation suggests that "certainty about natural phenomena...tends to vary inversely with proximity to the scientific work...." Thus "proximity to experimental work...makes visible the skilful, inexplicable and therefore potentially fallible aspects of experimentation, it lends salience to the web of assumptions that underlie what counts as an experimental outcome....Distance from the cutting edge is the source of what certainty we have."

The co-variation of certainty and proximity suggests that we view validity through different types of glass with varying focal lengths. Longer scientific foci presents a window into the distant outside environment. A closer, medium range political focus mirrors the scientists as actors. It is important that the investigator distinguish between these, that he or she "avoid mistaking a mirror of his (or her) own mind for a window into the minds of others (Schuman, 1986: 438)." The shortest focal lengths, the most minute microscopy, displays the glass itself – the character, idiosyncrasies, and effects of our language, our psychology.

A sophisticated political science should have control over these optics, consciously and systematically choosing the section – adjusting, comparing, and synthesizing mutually exclusive views to achieve a more pluralistic, exhaustive density of description. Perplexing as they are, we can not escape from the paradoxes of validity. The implications of validity suggest not a single mirror, but a whole corridor of them that produce images whose ultimate source is hard to trace and whose forms appear inconsistent. The political scientist and the broader political community which he or she observes appear to be not in separate rooms but in the same one. We are surrounded top and bottom, forward and backward, right and left, by a floor-to-ceiling glass corridor at which we look and which we have created. In different lights, the glass seems either to allow a view into other spaces, reflect back our own image, or provide a picture of itself. We can not know for sure which glass or which light currently exists; we do not have an independent measure of the glass's fidelity or its identity – window onto politics, mirror of ourselves, or kaleidoscope. There is glass ahead of us and behind, on all sides, above and below. It creates illusions and delusions with varying extension and distension. Recursive, regressive images stretch into infinity. Relaxing customary constraints makes us uncomfortable and uncertain at first. The images change with each step that we take. When we first see this, we are dizzy and

disoriented. And yet, as we walk, we do not fall through the glass. We can take steps with our feet that our eyes tell us are impossible. Once we proceed some way down this corridor with our eyes opened, we discover rich new areas for political science attention and investigation. The glass is invisible to the naive observer; but a sophisticated political science – reflective and reflexive – must increasingly take account of it. It must recognize and include its existence and importance, its limits and its power, its validities.

A clearer appreciation of our reflected, refracted situation opens new dimensions of political reality. Eventually in a Bayesian way, we can expand our comprehension and expertise so that we can make more educated guesses, more informed wagers, in our scientific game with political nature. We can not leap the final ontological barriers, and resolve all the paradoxes. Yet, as our scientific discourse, our epistemology, evolves, we can appreciate and use multiple methods in a more precise and powerful manner. We can switch and adjust our foci to specific purposes. We can develop a broader and more consistent, higher order lattice of belief. We can reorient our perception of science and achieve a more comprehensive and complex understanding, differentiating levels of analysis for scientific observers as well as scientific phenomena. In this conception, science can be simultaneously and alternatively an objective product and a subjective process, a material artifact and a human action, a language of nature and a topic of human rhetoric and discourse.

REFERENCES

- Abramson, Paul R., Brian D. Silver, and Barbara A. Anderson.** 1987. 'The effects of question order in attitude surveys, pp.the case of the SRC/CPS citizen duty items.' *American Journal of Political Science* 31, pp. 900-908.
- Arrow, Kenneth J.** 1988. "Workshop on the economy as an evolving complex system: summary," pp. 275-282 in Philip W. Anderson, Kenneth J. Arrow, and David Pines *The Economy as an Evolving Complex System* (Redwood City CA: Addison Wesley).
- Andrews, Frank M.** 1984. 'Construct validity and error components of survey measures, pp.a structural modeling approach.' *Public Opinion Quarterly* 48, pp. 409-442.
- Arlin, Patricia K.** 1982. 'A multitrait-multimethod validity study of a test of formal reasoning.' *Educational and Psychological Measurement* 42, 4 (Winter), pp. 1077-1088.
- Beer, Francis A.** 1986. 'Games and metaphors.' *Journal of Conflict Resolution* 30, 1 (March), pp. 171-191.
- Beer, Francis A.** 1974. *How Much War in History, pp. Definitions, Estimates, Extrapolations, and Trends.* Beverly Hills CA: Sage.
- Beer, Francis A., Alice F. Healy, Grant P. Sinclair, and Lyle E. Bourne, Jr.** 1987. 'War cues and foreign policy acts.' *American Political Science Review* 81, 3 (September), pp. 701-715.
- Beer, Francis A. and Thomas F. Mayer.** 1986. 'Why wars end: some hypotheses.' *Review of International Studies* 12, pp. 95-106.
- Bell, J. S.** 1987. *Speakable and Unspeakable in Quantum Mechanics.* Cambridge: Cambridge University Press.
- Benveniste, Emile.** 1971. *Problems in General Linguistics* (Mary Elizabeth Meek, tr.) Miami Linguistic Series Number 8. Coral Gables, Florida: University of Miami Press.
- Bernstein, Richard J.** 1985. *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis.* Philadelphia: University of Pennsylvania Press.

- Bishop, George F., Robert. W. Oldendick, and Alfred J. Tuchfarber.** 1984. 'What must my interest in politics be if I just told you 'I don't know?'' *Public Opinion Quarterly* 48, pp. 510-519.
- Bishop, George F., Robert. W. Oldendick, and Alfred J. Tuchfarber.** 1982. 'Political information processing: question order and context effects.' *Political Behavior* 4, pp. 177-200.
- Blume, Stuart S.** 1974. *Toward a Political Sociology of Science*. New York: Free Press.
- Boorstin, Daniel. J.** 1975. *The Image: A Guide to PseudoEvents in America*. New York: Atheneum.
- Bradburn, Norman H. and William H. Mason.** 1964. 'The effect of question order on responses.' *Journal of Marketing Research* 1, pp. 57-61.
- Brody, Charles J.** 1986. 'Things are rarely black and white: admitting gray into the Converse model of attitude stability.' *American Journal of Sociology* 92, pp. 657-677.
- Brown, James R.** 1984. *Scientific Rationality: The Sociological Turn*. Dordrecht: D. Reidel.
- Calder, Bobby J. and Lynn W. Phillips.** 1983. 'Beyond external validity.' *Journal of Consumer Research* 10 (June), pp. 112-114.
- Calder, Bobby, J., Lynn W. Phillips, and Alice M. Tybout.** 1982. 'The concept of external validity.' *Journal of Consumer Research* 9, 3 (December), pp. 240-244.
- Campbell, Donald T.** 1987. 'Evolutionary Epistemology,' pp. 47-90 in Gerard Radnitzky and W. W. Bartley, III, *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. La Salle IL: Open Court.
- Carmines, Edward G. and Richard A. Zeller,** 1979. *Reliability and Validity Assessment*. Beverly Hills: Sage.
- Casti. John L.** 1989 *Alternate Realities: Mathematical Models of Nature and Man*. New York: John Wiley.
- Cook, Thomas D.** 1985. 'Postpositivist critical multiplism,' pp. 21-62 in R. L. Shotland and R. R. Mark (eds.) *Social Science and Social Policy*. Beverly Hills: Sage.
- Cook, Thomas D. and Donald T. Campbell.** 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- Devereux, George.** 1967. *From Anxiety to Method in the Behavioral Sciences*. The Hague: Mouton.
- Duncan, Otis D. and Magnus Stenbeck.** 1988. 'No opinion or not sure.' *Public Opinion Quarterly* 52, 4 (Winter), pp. 513-525.
- Eco, Umberto and Thomas A. Sebeok.** 1988. *The Sign of Three: Dupin, Holmes, and Peirce*. Bloomington: Indiana University Press.
- Edelman, Murray.** 1988a. *Constructing the Political Spectacle*. Chicago: University of Chicago Press.
- Edelman, Murray.** 1988b. 'Skeptical studies of language, the media, and mass culture.' *American Political Science Review* 82, 4 (December): 1333-1339.
- Fiske, Donald W. and Richard A. Shweder,** eds. 1986. *Metatheory in Social Science: Pluralisms and Subjectivities*. Chicago: University of Chicago Press.
- Fodor, Jerry A.** 1975. *The Language of Thought*. Cambridge MA: Harvard University Press.
- Freund, Tallie, Kruglanski, Arie W., and Shpitzajzen, Avivit.** 1985. 'The freezing and unfreezing of impressional primacy: effects of the need for structure and the fear of invalidity.' *Personality and Social Psychology Bulletin* 11, 4 (December), pp. 479-487.
- Fauconnier, Gilles.** 1985. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge MA: Bradford.
- Fienberg, Stephen E.** 1989. *The Evolving Role of Statistical Assessments as Evidence in the*

Courts. New York: SpringerVerlag.

Flathman, Richard E. 1987. *The Philosophy and Politics of Freedom*. Chicago: University of Chicago Press.

Fleishman, John A. 1988. 'The effect of decision framing and others' behavior on cooperation in a social dilemma.' *Journal of Conflict Resolution* 32, 1 (March), pp. 162-180.

Folse, Henry J. 1985. *The Philosophy of Niels Bohr: The Framework of Complementarity*. Amsterdam: North-Holland.

Gibbons, Michael T., ed. 1987. *Interpreting Politics*. New York: New York University Press.

Glenn, Norval. 1989. 'What we know, what we say we know: discrepancies between warranted and stated conclusions' in Heinz Eulau, ed., *Crossroads of Social Science: The ICPSR 25th Anniversary Volume*. New York: Agathon Press.

Gordon, Michael E., L. Allen Slade, and Neal Schmitt. 1987. 'Student guinea pigs: porcine predictors and particularistic phenomena.' *Academy of Management Review* 12,1, pp. 160-163.

Gordon, Michael E., L. Allen Slade,, and Neal Schmitt. 1986. 'The 'Science of the Sophomore' revisited: from conjecture to empiricism.' *Academy of Management Review* 11, pp. 191-207

Greenberg, Jeff. 1987. 'The college sophomore as guinea pig; setting the record straight.' *Academy of Management Review*, 12, pp. 157-159.

Greiner, Russell 1988. 'Learning by understanding analogies.' *Artificial Intelligence* 35, 1 (May), pp. 81-125.

Hallett, Garth L. 1988. *Language and Truth*. New Haven: Yale University Press.

Harris, Francis C. and Benjamin B. Lahey. 1982. 'Subject reactivity in direct observational assessment: a review and critical analysis.' *Clinical Psychology Review* 2, 4:523-538.

Haynes, Stephen N. and Wade F. Horn. 1982. 'Reactivity in behavioral observation: a review.' *Behavioral Assessment* 4, 4 (Fall), pp. 369-385.

Herring, Philip F. 1987. *Joyce's Uncertainty Principle*. Princeton: Princeton University Press.

Herzog, A. R. and J. G. Bachman. 1981. 'Effects of questionnaire length on response quality.' *Public Opinion Quarterly* 45, pp. 549-559.

Hobbes, Thomas. 1958. *Leviathan*. Indianapolis: Bobbs Merrill.

Hoffmann, Stanley, ed. 1987. *Janus and Minerva: Essays in the Theory and Practice of International Politics*. Boulder: Westview.

Jackson, Sally and Scott Jacobs. 1983. 'Generalizing about messages: suggestions for design and analysis of experiments.' *Human Communication Research* 9, 2 (Winter), pp. 169-181.

Jagtenberg, Tom. 1983. *The Social Construction of Science: A Comparative Study of Goal Direction, Research Evaluation and Legitimation*. Dordrecht: D. Reidel.

Illich, Ivan, and Barry Sanders. 1988. *The Alphabetization of the Popular Mind*. San Francisco: North Point Press.

Kahneman, Daniel, Paul Slovic, and Amos Tversky, editors. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

King, Gary 1989. 'Event count models for international relations: generalizations and applications.' *International Studies Quarterly* 33, 2 (June):123-148.

Krosnick, Jon A. and Duane F. Alwin. 1988. 'A test of the form resistant correlation hypothesis.' *Public Opinion Quarterly* 52, 4 (Winter), pp. 526-538.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge MA: Harvard University Press.

- Lee, Jo. A., Kathleen E. Moreno, and J. B. Sympson.** 1986. 'The effects of test administration on test performance.' *Educational and Psychological Measurement* 46, 2:467-474.
- Leichsenring, Falk.** 1985. 'Die probleme der 'externen' validit t in der psychotherapieforschung.' *Zeitschrift fur Klinische Psychologie: Forschung und Praxis* 14,3, pp. 214-227.
- Lewis, David** 1973. *Counterfactuals*. Cambridge MA: Harvard University Press.
- Lodge, Milton, Kathleen M. McGraw, and Patrick Stroh.** 1989. An impression-driven model of candidate evaluation. *American Political Science Review* 83, 2 (June), pp. 399-420.
- MacKenzie, Donald.** 1990. *Inventing Accuracy: A Historical Sociology of Missile Guidance*. Cambridge MA: MIT Press.
- Mook, Douglas G.** 1983. 'In defense of external invalidity.' *American Psychologist* 38, 4 (April), pp. 379-387.
- Morey, Leslie C., Roger K. Blashfield, Warren W. Webb, and Jennifer Jewell.** 1988. 'MMPI scales for DSM-III personality disorders: a preliminary validation study.' *Journal of Clinical Psychology* 44, 1 (January), pp. 47-50.
- Mukerji, Chandra.** 1990. *A Fragile Power: Scientists and the State*. Princeton: Princeton University Press.
- Neisser, Ulric.** 1976, *Cognition and Reality: Principles and Implications of Cognitive Psychology*. New York: W. H. Freeman.
- Norpoth, Helmut, and Milton Lodge,** 1985. 'The difference between attitudes and nonattitudes in mass publics: just measurement?' *American Journal of Political Science* 29 291-307.
- Norris, Christopher.** 1987. *Derrida*. Cambridge MA: Harvard University Press.
- Quine, Willard van O. Word and Object.** Cambridge MA: MIT Press.
- Park, Aran.** 1988. *War Cues and Foreign Policy Acts in South Korea*. Boulder CO: Center for International Relations, University of Colorado.
- Pollock, John C.** 1981. *The Politics of Crisis Reporting*. New York: Praeger.
- Radnitzky, Gerard and W. W. Bartley III,** eds. 1987. *Evolutionary Epistemology Rationality, and the Sociology of Knowledge*. 1987. La Salle IL: Open Court.
- Rasmussen, Erik.** 1987. *Complementarity and Political Science: An Essay on Fundamentals of Political Science Theory and Research Strategy*. Aarhus: Odense University Press.
- Redner, Harry.** 1987. *The Ends of Science: An Essay in Scientific Authority*. Boulder: Westview.
- Rorty, Richard.** 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rorty, Richard.** 1982. *The Consequences of Pragmatism*. Minneapolis MN: University of Minnesota Press.
- Rosnow, Ralph L. and Marianthi Georgoudi,** editors. 1986. *Contextualism and Understanding in Behavioral Science: Implications for Research and Theory*. New York: Praeger.
- Ross, Dorothy.** 1991. *The Origins of American Social Science*. Cambridge: Cambridge University Press.
- Ruhl, Charles.** 1989. *On Monosemy: A Study in Linguistic Semantics*. Albany: State University of New York Press.
- Scheffler, Israel** 1979. *Beyond the Letter: A Philosophical Inquiry into Ambiguity, Vagueness, and Metaphor in Language*. London: Routledge and Kegan Paul.
- Schuman, Howard.** 1986. 'Ordinary questions, survey questions, and policy questions,' *Public Opinion Quarterly* 50, 3 (Fall), pp. 432-442.
- Schuman, Howard and Jacob Ludwig.** 1983. 'The norm of evenhandedness in surveys as in

- life.' *American Sociological Review* 48 (February), pp. 112-120.
- Schuman, Howard, Jacob Ludwig, and Jon A. Krosnick.** 1986. 'The perceived threat of nuclear war, salience, and open questions.' *Public Opinion Quarterly* 50, pp. 519-536.
- Schuman, Howard and Stanley Presser.** 1981. *Questions and Answers in Attitude Surveys: Experiments in Question Form, Wording, Context.* New York: Academic Press.
- Schuman, Howard and Jacqueline Scott.** 1987. 'Problems in the use of survey questions to measure public opinion.' *Science* 236 (May 22), pp. 957-959.
- Sears, David O.** 1986. 'College sophomores in the laboratory: influences of a narrow data base on social psychology's view of human nature,' *Journal of Personality and Social Psychology* 51, 3, pp. 515-530.
- Sigelman, Lee** 1981. 'Question Order effects on Presidential popularity.' *Public Opinion Quarterly* 45, pp. 199-207.
- Silverman, Kaja** 1983. *The Subject of Semiotics.* New York: Oxford University Press.
- Singer, Eleanor.** 1988. 'Pushing back the limits to surveys,' *Public Opinion Quarterly* 52, pp. 416-426.
- Small, Steven, I, Garrison W. Cottrell, and Michael K. Tanenhaus,** 1988. *Lexical Ambiguity Resolution.* San Mateo CA: Morgan Kaufmann.
- Sudman, Seymour, and Norman M. Bradburn.** 1974. *Response Effects in Surveys.* Chicago: Aldine.
- Sussmann, Mario and Donald U. Robertson.** 1986. 'The Validity of validity: an analysis of validation study designs.' *Journal of Applied Psychology* 71, 3, pp. 461-468.
- Taylor, Marylee C.** 1983. 'The black-and-white model of attitude stability: a latent class examination of opinion and nonopinion in the American public.' *American Journal of Sociology* 89, pp. 373-401.
- Turner, Charles F. and Elizabeth Martin,** eds. 1984. *Surveying Subjective Phenomena.* New York: Russell Sage.
- Van Dijk, Teun A.** 1984. *Structures of International News: A Case Study of the World's Press.* Amsterdam: Department of General Literary Studies, University of Amsterdam.
- Washington, Ernest D. and Vonnie C. McLoyd.** 1982. 'The external validity of research involving American minorities. *Human Development* 25, 5 (September-October) 324-339.
- West, Stephen G.** 1986. 'Methodological developments in personality research: an introduction.' *Journal of Personality* 54, 1 (March), pp. 1-17.
- Wilson, Richard. K. and Roberta Q. Herzberg.** 1988. 'Of machines and men: a cautionary note on the use of robots in decision-making experiments,' *Simulation and Games* 19, 2 (June), pp. 157-172.
- Wolf, Fred A.** 1988. *Parallel Universes: The Search for Other Worlds.* New York: Simon and Schuster.
-

MISCELLANEOUS

- Critical skepticism toward experimental research is thus well founded. A vast array of experimental variables radically explodes uncontrolled variance and degrees of freedom. Standard experimental strategies identify only a few possible determinants of results.

- Yet non-experimental methods of research are similarly flawed, glass houses equally vulnerable to stones.
- This belief rests on intuitive common sense, the default values of our culture, rather than our science as it is usually defined.
- most political scientists are probably reluctant to move too far along the path of reflexivity. This road is not only unpaved, but also uncharted. Nevertheless, contemporary research suggests
- This is common sense, the default value of our culture, the way things ordinarily look to the person in the street.
- the scope of universal commensurability.
- What the political scientist observes is, as Rorty (1979) suggests, at least partly a mirror of the political community and of political science within it, reflecting their characteristics.
- The viewer, from this angle, resembles the experimental subject, whose perception is deeply affected by context. The prior discussion strongly suggests that the wider political context strongly influences the political science that is located within it. These political presuppositions and assumptions may be stronger for being unremarked and unstated. For example, political scientists within contexts where political experience has been relatively continuous, for example Britain and the United States, may believe that political science can come close to offering stable general knowledge, a sure thing. Their experience suggests that past patterns will be good future predictors, that parameters will be stable over long periods of time, that context-free global invariants probably exist and can be discovered. On the other hand, the contemporary political experience of the major continental European nations – France, Germany, Austria-Hungary, the Soviet Union – or Asian ones – China, India – may produce a greater scientific caution, more of an inclination to proceed slowly and hedge scientific claims with greater sensitivity to parametric instability and validation paradoxes. Political turbulence and discontinuity, revolution and dismemberment, teach the terrible lightness of being. Their lesson is that the future is not necessarily like the past. Reality is ephemeral and tenuous; nothing lasts. We may believe we have found local regularities. When we search for global constants, our reach surely exceeds our grasp.
- Similarly, we must learn to see and integrate different glasses of political science and politics.
- Political science is like a corridor made of glass, down which we walk. We do not know the exact type of glass – window, mirror, kaleidoscope – its fidelity or distortion. In some lights, the glass shows us pictures that we take to be external political behavior; in other lights the glass seems to reflect back an endless series of our own refracted images. A complex, sophisticated political science must provide a way to understand these optics. We can move toward this understanding by systematically adjusting and comparing the views provided by multiple focal lengths. Longer foci show our present view of the world

and reflections of ourselves. They allow us to recognize and articulate inconsistencies, resolve paradoxes in meta- and sub-systems. A shorter focus pays more attention to our view of the glass itself, the character and idiosyncrasies of our language – words, numbers, pictures – and their relationship to political action and representation.

- Politically defined and charged words (validity, text, context, analysis), numbers (statistics, mathematical formulae), and pictures (figures, graphs, models) carry distinct presuppositions, assumptions, and implications. They are our collective tools for conveying the meanings and intentions that are the basis of political communication and action. Under different conditions we let them correspond to, or represent, external referents. We must be careful the words, numbers, and pictures do not, metonymically or metaphorically, simply substitute for what they supposedly represent.
- They allow us to recognize and articulate inconsistencies, to resolve paradoxes in meta- and sub-systems, rather than simply assuming them away.
- Yet language is much more imprecise and unstable than even this structured research might suggest.
- Within the zone of semantic stability, marginal changes in language may make very important differences in what is understood.
- Regardless of the experimenter's care and precision,
- The picture theory of language is seductive.
- The meaning of text varies for a number of reasons.
- Small lexical differences can produce dramatic semantic effects.
- Lexical determinacy, stability. Words have no definite meanings.
- Objective language. This is hardly the case. Practice is, however, far more complex.
- The correspondence theory of language.

Words do not have definite meanings.

but there is a tremendous loss of information.

- The import of words changes to some extent with circumstance, context, intention, and perception.
- 'The more deeply one looks inside the black box, the more one realizes that 'the technical' is no clear-cut and simple world of facts insulated from politics' (MacKenzie, 1990: 381).
- Language can be used with the precision of a surgical instrument
- Lexical-semantic congruence, multiplicity

- Each single word does not necessarily have a single determinate meaning.
 - The meanings of words are defined in many places, in many ways.
 - Moore, Robert C. 1985. 'Semantical considerations on nonmonotonic logic.' *Artificial Intelligence* 25, 1 (January), pp. 75-94.
 - Konolige, K. 1988. 'On the relation between default and autoepistemic logic.' *Artificial Intelligence* 35, 3 (July), pp. 343-382.
-