PRAGMATIC ELIMINATIVE INDUCTION: PROXIMAL RANGE AND CONTEXT VALIDATION IN APPLIED SOCIAL EXPERIMENTATION[1]

William N. Dunn

1. *Introduction*

Donald T. Campbell's vision of science was a product of his trust in its past achievements and potential, his doubts about its organization and everyday practice, and his conviction that a dialectical conjunction of trust and doubt is essential to the success of the applied social sciences.

This dialectical vision is nowhere more important than in Campbell's contributions to the epistemology and methodology of applied social experimentation. Although these quintessentially philosophical contributions established him as a master methodologist,[2] Campbell is better known for his advocacy of the "true" experiment, which for him represented a methodologically optimal way to make causally interpretable knowledge claims.[3] Yet he also believed that the vast bulk of real-life social experiments—that is, those conducted outside the relatively closed system of the laboratory--are rarely if ever "true" (in the sense of genuine), because they are plagued by uncontrolled and uncontrollable contingencies which originate in the complexity of social systems. Applied social experiments, argued Campbell, are carried out in complex systems in which numerous contingencies lie beyond the control, and often the awareness, of the experimenter (see Campbell 1959, 1988c; Campbell and Stanley 1963, pp. 4-5-5).

It is precisely this systemic character of applied social experimentation which justified early contrasts between (true) "experiments" and "quasi-experiments" conducted in field settings (Campbell 1960, 1974b; Cook and Campbell 1979). Later, in an effort

to emphasize the unalterably systemic character of social experimentation, Campbell proposed the adoption of new terms for his own traditional categories of internal and external validity, which he now called "local molar causal validity" and "proximal similarity" (Campbell 1986a). Although some may see "systemic social experimentation" as a recent turn in Campbell's work, it is important to recognize that systems thinking was an integral part of his scholarship and teaching for four decades. Drawing on those two great traditions of systems thinking, cybernetics and evolutionary biology, the bulk of Campbell's wide ranging contributions to the philosophy, methodology, and sociology of the social sciences (see Brewer and Collins 1981; Campbell 1988a; Overman 1988) employ a natural-systems, or naturalistic, perspective to address problems in the psychology of knowledge processes (1959), evolutionary epistemology (1974a, 1977), the ethics of human survival (1975b, 1976), and what he called the sociology of scientific validity (1986b, 1994, 1996).

One of the most important features of social experiments, then, is that they occur amidst unmanageable contingencies which originate in the complexity of social systems (Campbell 1988c). These unmanageable contingencies represent "threats to validity" or "rival hypotheses" (Campbell and Stanley, 1963; Cook and Campbell (1979), which must be tested and, where possible, eliminated. This process of elimination, or eliminative induction, is the distinctive feature of Campbell's methodology of quasi-experimentation, and typically is not necessary when conducting a fully randomized laboratory experiment, at least in theory.[4] In contrast to enumerative induction, which involves a search for evidence which corroborates an initially favored hypothesis, eliminative induction involves a search for evidence which either corroborates a rival hypothesis, and thus weakens or eliminates the originally favored hypothesis, or which fails to corroborate the rival hypothesis, which is then eliminated.

Eliminative induction, which builds on J. S. Mill's normative logic of causal reasoning (Mill 1843) and Karl Popper's fallibilist theory of induction (Popper 1963, 1968; see Campbell 1974a), permits a critical examination of social systems contingencies which cannot be directly managed by the experimenter. In one of his last published papers, Campbell (1996) reaffirmed that his views on eliminative induction were "pragmatic," as distinguished from "logical" or "analytic" (see also

Campbell 1969a)." He endorsed *pragmatic eliminative induction* (Dunn 1995) as a process in which the "elimination of rival explanations is only 'plausible,' historically dated, and *never* complete."[5] While recognizing the "ubiquity of inductive incompleteness," he concluded that inductive incompleteness, per se, does not preclude efforts to eliminate *some* plausible rival hypotheses (Campbell 1996, Sec. 5.4). Pragmatic eliminative induction, in Campbell's formulation, is thus restricted to testing *some* rival hypotheses.

This paper argues that it is not necessary to restrict pragmatic eliminative induction to the identification and testing of *some* rival hypotheses. Although the process of identifying and testing rival hypotheses is never complete, this does not mean that the process is entirely indeterminate. On the contrary, we can estimate through a process of successive approximation the likely range of plausible rival hypotheses which, in specific socio-temporal contexts, may be invoked to challenge an initially favored hypothesis.[6] Instead of viewing pragmatic eliminative induction as an "all-or-some" affair—with the problem of incompleteness reduced to testing all rival hypotheses, or some indeterminately smaller number of them—we can view eliminative induction as a pragmatic process of *truth-estimation,* with "truth" understood as an optimally plausible estimate in a given problem situation or context.

Thus, while Campbell enjoins us to test and eliminate *some* rival hypotheses, this paper makes the bolder claim that it is possible to test and eliminate an *approximately complete* set of rival hypotheses. My argument is as follows:

● In contrast to the practice of testing *some* rival hypotheses on grounds that these hypotheses are conveniently accessible and backed by already existing scientific evidence, pragmatic eliminative induction begins by estimating the *proximal range* of rival hypotheses present in a disputatious community of knowledgeables. Hence, an estimate of the proximal range of rival hypotheses *precedes* the process of testing and elimination.

● By estimating the proximal range of plausible rival hypotheses we can assess the approximate *context validity* of social experiments, or any other type of social research. *Threats to context validity* represent a fifth class of rival hypotheses which may be eliminated,

or at least mitigated, before testing rival hypotheses of the four remaining types: statistical conclusion validity, internal validity, external validity, and construct validity.

● When pragmatic eliminative induction is used to estimate the context validity of rival hypotheses in a disputatious community of knowledgeables, we may expect significant improvements in the statistical conclusion, internal, external, and construct validity of social experiments. This is demonstrated throughout the paper by an analysis of the failed social experiment, undertaken in Europe and the United States, to reduce traffic deaths by setting maximum speed limits.

2. The proximal range of rival hypotheses

The methodology of applied social experimentation, although it represents one of the most useful methodological innovations in 20th-century social science, does not satisfactorily address an important, long-standing problem associated with the process of eliminative induction. This problem, which appeared in 19th-century debates between John Stuart Mill and his critics,[7] can be stated as a question: In conducting a social experiment—for example, a policy designed to reduce some social ill—is there a way to know if we have identified and tested an approximately complete range of rival hypotheses?

   At first glance, the range of possible rival hypotheses usually appears unmanageably huge or even infinite. For this reason, the process of defining an applied research problem may seem like "a never-ending discourse with reality, to discover yet more facets, more dimensions of action, more opportunities for improvement" (Dery 1984, pp. 6-7). Here, applied social scientists appear as homesteaders clearing a plot of land in the wilderness.[8] The homesteaders are aware that enemies lurk in the wilderness that lies beyond the clearing. To increase their security, they clear a larger and larger area, never feeling confident they have succeeded. Frequently, they must decide whether to clear more land, or remain within the perimeter and attend to regular chores. The homesteaders try their best to push back the wilderness, knowing that the enemies lurking beyond the clearing may one day surprise and even destroy them. Although they have occasional doubts, most of the time the

homesteaders trust that the boundary with the wilderness will protect them.

This problem, which is captured by the notion of *proximal range*,"[9] asks whether there is some way to estimate the number and variety of rival hypotheses which should be tested, and, where possible, eliminated. In the natural sciences, observes Campbell (1969, p. 353), the problem is alleviated by means of a relatively simple practical process. While acknowledging that the pool of rival hypotheses is potentially infinite, natural scientists pay little or no attention to this mere logical possibility. Threats to the validity of scientific claims arise only when actually developed alternative explanations exist.

In the applied social sciences, however, the problem is more complex and acute. Here, procedures are needed for estimating the range of rival hypotheses, both theoretical and methodological, which arise out of the many contingencies which lie beyond the control of experimenters, be they social scientists or policy makers. The elimination of *some* rival hypotheses, while valuable, tends to be *ad hoc* and indeterminate. Indeed, standard eliminative induction resembles the *ad hoc* literature review which examines positive and negative evidence by restricting studies to those which are readily available, consistent with conventional social science wisdom, and/or backed by extant scientific (or political) authority. The typical literature review, much like the process of identifying and testing some indeterminate number of rival hypotheses, tends to be arbitrary, unsystematic, biased, and, by most definitions, unscientific (see Light and Pillemer 1984).

Pragmatic eliminative induction addresses this problem by building on the doctrine of fallibilism, as developed by Popper (1963, 1968), where the theories we provisionally trust are those which have not yet been eliminated. Pragmatic eliminative induction nevertheless departs from Popper's doctrine of falsificationism, because the process of elimination is not seen as a direct confrontation of a dominant single theory with fact, but as a process of testing many competing theoretical and methodological hypotheses against available trusted "facts."[10] This pragmatic point of view accords with Popper's rejection of positivist principles of verification and confirmation, recognizing that theories are never verified or confirmed as true. But it rejects the notion that falsification demands the unequivocal overthrow of single theories by pitting them against accumulated facts. Instead, the process of eliminating

rival hypotheses is a matter of comparative plausibilities, a process which seeks but never fully attains completion (Campbell 1996, Sec. 5.2-5.3).

The problem of proximal range points to a major deficiency of the tradition of eliminative induction represented by Mill and Popper. Mill's claims to have created a certain method for demonstrating causation has been rightly challenged on grounds that, to achieve this, one must have a complete enumeration of rival theoretical and methodological hypotheses. Although Mill's methods (canons) of induction—agreement, difference, agreement-and-difference, concomitant variation, and residues—are important to the methodology of eliminative induction, their application does not produce the certainty Mill seemed to believe they produce. Mill's methods presuppose prior knowledge of causally relevant constructs and hypotheses when, in fact, we usually begin our causal investigations in a state of relative ignorance.[11] Popper (1963), while he enjoins methodologists to offer bold conjectures and refutations, leaves them essentially empty handed when faced with the problem of determining the proper range of rival hypotheses to test. Although the four-fold classification of threats to validity developed by Campbell (1957), Campbell and Stanley (1963), and Cook and Campbell (1979) is a valuable research heuristic, it offers little assurance that we have defined the proper range of rival hypotheses.

The usefulness of the four-fold classification of threats to internal, external, construct, and statistical conclusion validity is that it increases the likelihood that the host of rival hypotheses typically present in social experiments will be incorporated in our analyses. This four-fold classification, although it does not address the problem of proximal range, does reject Mill's ambitious "essentialist" claim of having established a method for discovering and verifying the occurrence of necessary and sufficient causes.[12] This explicit rejection, based on the fallibilist methodology of eliminative induction, is essential to Campbell's postpositivist philosophy and methodology of science (see Dunn, 1998, Table 1). Nevertheless, no procedures are available to estimate the extent to which the several-dozen threats to validity detailed by Cook and Campbell (1979, Ch. 2) might exhaust or approximate the real but unknown range of plausible rival hypotheses. The four-fold framework is not sufficient for discovering this range, because to be sufficient would require prior empirical knowledge of the complete collection of rival hypotheses, or a commitment to some a priori principle such as that of

finite possibilities or truth-tropism (Rescher 1980, p. 218). In short, the four-fold framework assumes prior knowledge of rival hypotheses before the framework itself can be productively used.[13]

Because the range of rival hypotheses is never complete does not mean that the proximal limit of this range cannot be estimated. In this context, pragmatic eliminative induction grounds the search for rival hypotheses by seeking rival explanations within a disputatious community of knowledgeables, rather than in some a priori principle which assumes a fixed range. This enlarges and naturalizes the range of rival explanations available to contest an established scientific theory or preferred policy.[14] Argumentation and debate within and among political and scientific communities can be an important source of rival hypotheses within this proximal range (Campbell 1982; Dunn, 1993b; Fischer and Forester 1993).

3. The context validity of social experiments

So far, this essay has attempted to place Campbell's theory of social experimentation within the larger framework of pragmatic eliminative induction, focusing on the problem of establishing the proximal range of rival hypotheses. I now want to introduce and apply a fifth class of rival hypotheses which complements the Campbell-Stanley-Cook classification of threats to validity. This fifth class of rival hypotheses, which I will call context validity, refers to the validity of inferences that we have estimated the proximal range of rival hypotheses. Context validity, which draws on Rescher's (1980) methodological-pragmatic criteria for making plausibility truth-estimates, has four major requirements: character, coordination, correctness-in-the-limit, and cost effectiveness.[15]

Character
An estimate of the proximal range of rival hypotheses should have the same character as that which it estimates. Just as an estimate of a length should be a length, not a temperature, an estimate of the proximal range of plausible rival hypotheses should meet a character requirement: Plausible rival hypotheses should be subjectively meaningful to, and elicited from, those who support and oppose solutions for social problems. The recognition that all knowledge is "embodied" leads to a

search for rival hypotheses in naturalistic epistemologies (Campbell 1977). Accordingly, estimates of the proximal range of rival hypotheses are based on observations which are subjectively meaningful to members of a disputatious community of knowledgeables. Observations may be obtained by means of interviews, questionnaires, participant observations, ethnographies, content analyses, or transcripts of public debates.

Coordination

An estimate of the proximal range of plausible rival hypotheses should coordinate with the shape and distribution of beliefs known to characterize knowledge systems of many kinds. The closer the coordination, the more accurate the estimate. An estimate of rival hypotheses within a disputatious community should be asymmetric and markedly skewed. Rival hypotheses which occur less frequently indicate trust in taken-for-granted knowledge; those occurring more frequently are symptoms of doubt.[16] This reflects the large trust-doubt ratio (see Campbell 1977) observed in natural systems of many kinds. Because estimates of rival hypotheses depend on those who believe in these hypotheses, the sampling of individuals is necessary. Purposive samples of knowledgeables who are part of the same knowledge system, rather than random samples of unrelated policy makers, scientists, and citizens, are more likely to produce the coordinated distribution of rival hypotheses required for approximate context validity.[17]

Correctness-in-the-Limit

As information on which an estimate is based becomes more complete, the estimate should eventually converge on the presumptively true range of rival hypotheses in a natural knowledge system. A cumulative frequency distribution of unique (non-duplicate) rival hypotheses may be arranged in order of decreasing frequency of occurrence and plotted on a graph. The plot of these unique rival hypotheses should eventually flatten out, indicating that we have achieved approximate context validity by approaching the limit of disputed truths in a natural knowledge system.

Cost-Effectiveness

An estimate of the proximal range of rival hypotheses should be economical, in the sense of "economy of cognition." Although the total

costs of an estimate may vary with the relative complexity of a natural knowledge system, evidence from areas as diverse as marketing, linguistics, bibliometrics, and psychotherapy suggests that the proximal range of a system of rival hypotheses is likely to be reached within a small number (15-30) of probes.[18] The "probative value" of a rival hypothesis--defined as its potential usefulness in challenging a knowledge claim--is inversely related to its frequency of occurrence.[19] The marginal costs of identifying each additional rival hypothesis appear to increase at a diminishing rate, as does the probative value of each new rival hypothesis. The achievement of approximate context validity is cost-effective, reflecting an economy of cognition at the knowledge-system level.[20]

The importance of context validity to social experimentation is evident when we probe the efficacy of a policy such as the National Maximum Speed Limit of 1974, which established a uniform speed limit of 55 mph on all interstate highways in the United States. The 55 mph speed limit, adopted as a means to reduce gasoline consumption during the 1973-74 OPEC oil embargo, was unexpectedly followed by a dramatic decline in traffic fatalities the year after its implementation. Between January 1 and December 31, 1974, there was a decline of 9,100 fatalities, a 32 percent drop. Despite continuing opposition from rural Western states and the trucking industry, there was broad support for the policy among policy makers, policy analysts, and the general public. And, clearly, the problem is not unimportant or trivial--traffic fatalities represent the leading cause of death among persons 35 years of age and younger, while average annual highway deaths since 1974 are equivalent to a fully loaded 767 aircraft crashing with no survivors every third day of the week.

On April 2, 1987, Congress enacted the Surface Transportation and Uniform Relocation Assistance Act of 1987, overriding President Reagan's veto. Provisions of this bill permitted individual states to experiment with speed limits up to 65 mph on rural interstate highways. By July, 1988 forty states had raised the speed limit to 60 or 65 mph on eighty-nine percent of rural interstate roads. Senator John C. Danforth, an influential supporter of the 55 mph speed limit, argued against the new policy. The 65 mph speed limit, said Danforth, would save an average of one minute per day per driver, but result in an annual increase of 600 to 1,000 deaths.[21] *The Washington Post* also expressed unequivocal

opposition to the new policy: "The equation is in minutes versus lives. It's not even close . . . . A hundred miles at 55 mph take about 17 minutes longer than at 65. That's the price of those lives. It ought to be the easiest vote the House takes this year." Several years later, in a November, 1995 press release, U.S. Secretary of Transportation Federico Pena reaffirmed the Clinton Administraion's opposition to higher speed limits, urging the Governors of the states where the 55 mph speed limit would automatically increase upon the 1995 repeal of the 1974 National Maximum Speed Limit to carefully consider the costs and benefits of increasing speed limits, especially health care costs.

The majority of policy analysts who have evaluated the effects of the policy, along with elected officials from the ten northeastern states which retained the 55 mph speed limit, have trusted the conclusion that the 1974 law was responsible for the decline in traffic fatalities. However, the evidence shows that these social science stakeholders failed to consider rival hypotheses which, had they been tested, would have resulted in an altogether different explanation of traffic deaths and, consequently, a different policy recommendation. In effect, research on the 55 mph speed limit was lacking in context validity.

To satisfy the four requirements of context validity--character, coordination, correctness-in-the-limit, cost-effectiveness--an analysis was conducted with documents prepared by thirty-eight state officials responsible for reporting on the effects of the 55 mph and 65 mph speed limits in their states. There is considerable political, administrative, professional, and geographic variation within this set of stakeholders, which includes governors, secretaries of transportation, chief highway engineers, traffic safety analysts, and commanders of state highway patrols in every region of the country.[22] Because this network of stakeholders is complex and relatively research-intensive, it was anticipated that a large number of hypotheses and rival hypotheses would be identified. As expected, there were sharp disagreements among many of the thirty-eight stakeholders, who represent the kind of disputatious community which Campbell (1986) sees as necessary for effective social experimentation and what he called the "experimenting society" (Campbell 1998). For example, some states were tenaciously committed to the hypothesis that speed limits are causally related to fatalities (e.g., Pennsylvania and New Jersey), while others were just as firmly opposed (e.g., Illinois, Washington, Idaho). Of direct importance to context

Table 1. Rival Hypotheses Used by 38 State Policy Makers to Dispute or Support Claims About the Effectiveness of the 55mph Speed Limit in Saving Lives: The Probative Value of a Rival Hypothesis Tends to be Negatively Related to its Frequency of Occurrence

The observed decline in fatalities after 1974 is caused by:

55 mph speed limit (38)	speed adaptation (2)	commuters (1)
dual speed limits (38)	interchange spacing (2)	interchanges (1)
law enforcement (38)	stopping distance (2)	travelers (1)
average speeds (38)	reaction time (2)	driver alertness (1)
dispersion of speeds (38)	citation bias (2)	overly easy driving
driver education (38)	age of drivers (2)	environment (1)
public information (38)	weaving (2)	awareness programs (1)
unreliable data (18)	driver alertness (2)	injury rate (1)
miscoded data (17)	driver awareness (2)	speed motivation (1)
traffic volume (15)	recovery zones (1)	night driving (1)
alcohol (12)	lane restrictions (1)	vehicle design (1)
catastrophic accidents (11)	perceived enforcement (1)	property accidents (1)
public support (10)	number of lanes (1)	risk taking (1)
random fluctuations (9)	length of mainlines (1)	roadway safety (1)
public attitudes (9)	night speed limits (1)	road conditions (1)
sample size (8)	equipment defects (1)	energy crises (1)
"safety switch" (7)	acceleration lanes (1)	nonlinear relation of risk
state contexts (6)	type of highway (1)	and speed (1)
fatal accident rate (6)	drinking age (1)	terrain (1)
trucks (5)	commercial traffic (1)	economic factors (1)
accident histories (5)	driver experience (1)	safety diversion (1)
interstate speed spillover (5)	low-flying aircraft (1)	contiguous roads (1)
improper use fatality rates (5)	truck accident rate (1)	driver endurance (1)
time of accidents (5)	perceived fairness (1)	population density (1)
drugs (5)	primary cause of crash (1)	regression toward the
highway design (4)	police presence (1)	mean (1)
speeding (4)	tactical enforcement (1)	police tolerance (1)
weather (4)	insurance premiums (1)	augmented compliance (1)
exceeding design speed (4)	vehicle occupancy (1)	vehicle proximity (1)
DUI cases (4)	speed adaptation (1)	passing maneuvers (1)
driver error (4)	economic mobility (1)	collision accidents (1)
traffic density (4)	coefficient of variation (1)	highway expansion (1)
sanctions for violators (4)	types of crashes (1)	platooning (1)
energy conservation (4)	misuse of data (1)	accident types (1)
fluctuations in trends (3)	enforcement contacts (1)	accident exposure (1)
state autonomy (3)	patrol hours (1)	energy savings (1)
vehicle safety (3)	acceptable risk (1)	
driver fatigue (3)	availability of emergency	
complexity of quantitative	medical services (1)	**109 unique hypotheses**
variables (3)	social costs (1)	**718 total hypotheses**
construction projects (2)	integrated compliance (1)	
demonstration effect of	declining oil prices (1)	
speeds (2)	traffic controls (1)	
failure to measure	technological breakthroughs (1)	
injury accidents (2)	child restraints (1)	

validity is that 718 plausible[23] rival hypotheses were used by thirty-eight stakeholders to affirm or dispute the effectiveness of the 55 mph speed limit in saving lives. Of this total, 109 hypotheses are unique, in that they do not duplicate hypotheses advanced by any other stakeholder.

Table 1 lists the 109 unique rival hypotheses used to dispute or support claims about the effectiveness of the 55 mph speed limit in saving lives. The number of times a unique hypothesis was mentioned is shown in parentheses. Here, it is important to note that, from the standpoint of communications theory and language, the information-content of a hypothesis tends to be negatively related to its relative frequency, or probability of occurrence.[24] Accordingly, hypotheses which are mentioned more frequently—those on which there is greater agreement or consensus--have less probative value than rarely mentioned hypotheses, because highly probable or predictable hypotheses do not challenge accepted knowledge claims. When inquiry is guided by the strategy of pragmatic eliminative induction, the only process available for determining the plausibility of a knowledge claim, or for corroborating a scientific hypothesis, is one of testing and eliminating rival hypotheses (Campbell 1969b, p. 170).

The rival hypotheses listed in Table 1 were entered into a program designed for ethnographic research[25] and analyzed in two ways. First, the frequency distribution of all hypotheses mentioned by the thirty-eight stakeholders was placed on a rank-frequency graph (Figure 1(a)). This graph shows the entire pool of hypotheses, which are ranked in descending order beginning with hypotheses with the highest frequency of occurrence. A second graph (Figure 1(b)) shows the cumulative frequency distribution of unique (non-duplicative) hypotheses used by each stakeholder, beginning with the stakeholder generating the most hypotheses, which sum to a total of 41 rival hypotheses. As expected, the distribution in Fig.1(a) conforms to the contours of Zipf's (1949) law of least effort, also known as his rank-frequency law. In turn, Fig. 1(b) conforms to the cumulative frequency distribution predicted by Bradford's law of scattering (Bradford 1934). Similar frequency distributions have been found to characterize the structure of languages, cultures, information, and knowledge in a wide array of diverse natural systems.[26] In this respect, evidence accumulated over the past 200 years suggests that the structure of beliefs, ideas, and languages in natural systems of many kinds conforms to the theories of Lotka (1926),
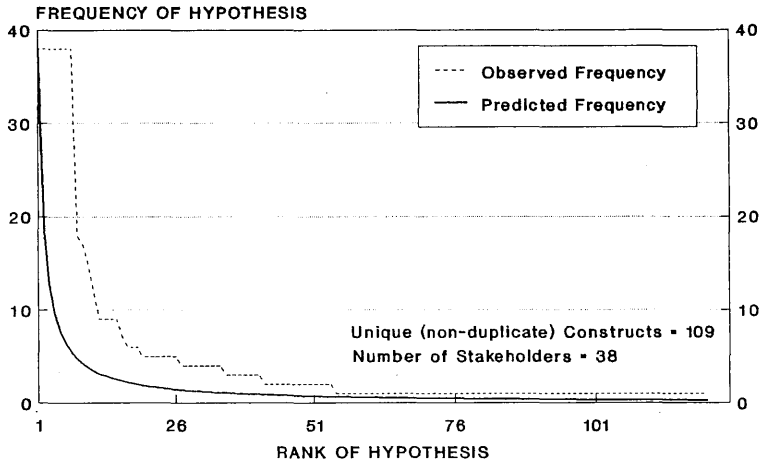
Bradford (1934), Zipf (1949), and Price (1986). The extent of this conformity may be assessed by means of quantitative goodness-of-fit procedures, for example, linear regression analysis with a semi-logarithmic transformation (see, for example, Simon 1972. This conformity can also be evaluated visually, by matching the pattern of observations predicted by the theories to the curves derived from those theories (compare Campbell 1966).[27]

The rank-frequency distribution of observed constructs and hypotheses (Figure 1(a)) is proximally similar to the patterns predicted by Lotka's inverse-square law of scientific productivity, Price's law of cumulative advantage in science, and Zipf's law of least effort in cognitive enterprises. In turn, the cumulative frequency distribution (Figure 1(b)) is proximally similar to Bradford's law of scattering, which has been used to estimate the redundancy of information contained in library holdings. The "empirical laws" of Lotka and Price may be explained by Zipf's "theoretical law" of least effort, which also explains the principle of economy of cognition used to underwrite methodological-pragmatic theories of induction (see Rescher 1980, p. 174, 181). Zipf's law of least effort states that human problem solving is based on the minimization of the total work required to solve immediate and future problems, as perceived by the problem solver.[28]
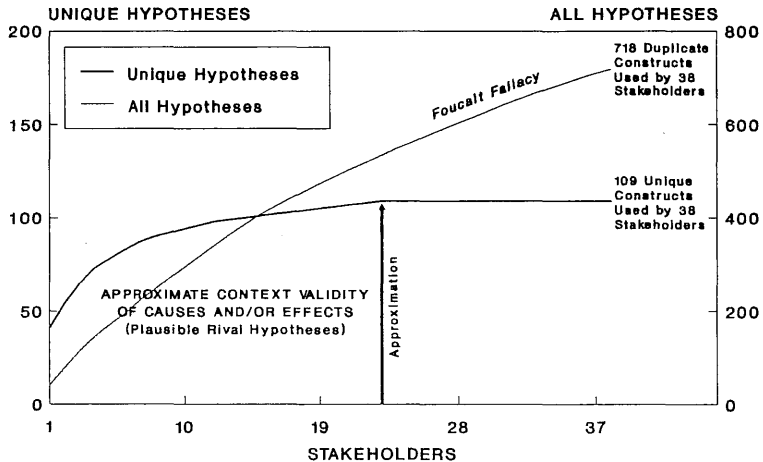

4. *Threats to context validity*

The results of the analysis displayed in the two graphs appear to satisfy the four requirements of context validity: character, coordination, correctness-in-the-limit, and cost-effectiveness. The results also appear to alleviate a critical problem facing applied social scientists—namely, the lack of a rational strategy or logic for discovering the proximal range of rival hypotheses in a complex knowledge system. If we make the reasonable assumption that knowledge of rival hypotheses is possible only in the presence of a knowing subject—that is, all knowledge is personally embodied--then the global structure of rival hypotheses in a given domain of knowers (a knowledge system) is equal to the total possible combinations of causally relevant constructs which form these hypotheses. However, in contrast to those who seek an intuitive (or a priori) warrant to justify what Rescher (1980, p. 217) calls "convenient range structure,"

Fig. 1(a). Zipf's rank-frequency law
predicts the observed frequency of
stakeholder rival hypotheses

FREQUENCY OF HYPOTHESIS


```
- - -  Observed Frequency
____   Predicted Frequency
```

Unique (non-duplicate) Constructs = 109
Number of Stakeholders = 38

RANK OF HYPOTHESIS

NOTE: Hypothesis rank (1=high) defined
by number of times each of 109 unique
hypotheses mentioned by 38 stakeholders.

Fig. 1(b).Cumulative frequency of unique
rival hypotheses used by stakeholders
shows approximate context validity

UNIQUE HYPOTHESES   ALL HYPOTHESES


```
____  Unique Hypotheses
- - -  All Hypotheses
```

Foucalt Fallacy

718 Duplicate
Constructs
Used by 38
Stakeholders

109 Unique
Constructs
Used by 38
Stakeholders

APPROXIMATE CONTEXT VALIDITY
OF CAUSES AND/OR EFFECTS
(Plausible Rival Hypotheses)

Approximation

STAKEHOLDERS

SOURCE: NHTSA, *The Effects of the 65 mph
Speed Limit Through 1988* (October 1989),
Appendix: Responses of State Officials.

pragmatic eliminative induction warrants claims of having discovered a proximal range in the naturalistic epistemology of observed and observable knowledge systems. Accordingly, it is possible to know that the candidates for elimination are located, in all likelihood, within the proximal range of rival hypotheses. Thus, the process of pragmatic eliminative induction yields an approximately true answer to the question: In conducting a social experiment, how many rival hypotheses should be tested and, where possible, eliminated?

Pragmatic eliminative induction seems to represent a rational strategy or logic for discovering, by means of a process of estimation, the proximal range of rival hypotheses. The terms "rational" and "logic" do not refer to the analytic properties of one or another valid logical form, or process of reasoning, but to a process of fallible, contingent, and corrigible truth-estimation guided by criteria of character, coordination, correctness-in-the-limit, and cost-effectiveness.[29] This process would involve an infinite regress in the absence of the proximal observed limit on the number of unique hypotheses in natural knowledge systems, or evidence that these hypotheses increase at a diminishing rate which rapidly approaches that limit.[30]

Philosophers traditionally argue that there is no logic of discovery which is distinct from logics of verification, confirmation, validation, or corroboration. Whether pragmatic eliminative induction constitutes a distinct "logic" in one of these particular technical senses is less important than the fact that the range of plausible rival hypotheses may be estimated.[31] Although these procedures are deductively imperfect, they nevertheless are useful and even necessary, since there does not appear to be an appropriate alternative way to justify the process of context validation.[32] The pragmatic or naturalistic epistemological justification of eliminative induction is related, of course, to evidence that the bulk of our knowledge is and must be vicarious and distal (Campbell 1988a, pp. 412-13). Pragmatic eliminative induction recognizes that we are almost completely dependent on socially vicarious learning to acquire knowledge of rival hypotheses which, once identified, may be placed within the four-fold classification of threats to statistical conclusion, construct, internal, and external validity. The addition of context validity to this four-fold classification recognizes and builds upon the naturalistic epistemology of vicarious social learning (Campbell 1988, Ch. 17), with context validity defined by the proximal range of a natural system of rival

hypotheses.

Any claim about the approximate context validity of a natural system of rival hypotheses is fallible, contingent, and corrigible. Accordingly, several threats to context validity may be invoked to challenge a claim that we have properly estimated the proximal range of rival hypotheses in a given problem situation or context. Three threats to context validity seem particularly relevant.

4.1 Inadequate Responsiveness of Methods

Methods used to elicit causally relevant constructs may be more or less responsive to the character requirement. A method should elicit subjectively meaningful beliefs of stakeholders who are knowledgeable about given policies. Inadequate responsiveness occurs when methods yield surface beliefs, rather than underlying causal assumptions, or when methods yield incorrect inferences about beliefs from overt behavior. The inadequate responsiveness of methods, which is particularly severe in information science and economic policy analysis, is also a problem of standard survey research methods in the social sciences.[33] In the present case, the beliefs of stakeholders in western states which abandoned the 65 mph speed limit displayed a profound skepticism about the hypothesis that speed limits and speeding are causally related to fatalities. By contrast, stakeholders in northeastern states which retained the 55 mph limit displayed unequivocal confidence in the same hypothesis. The open-coding procedure used to identify rival hypotheses appears to be responsive to the subjectively meaningful beliefs of policy stakeholders.

4.2 Discordance

The coordination of the observed distribution of rival hypotheses with the patterns predicted by the Lotka, Bradford, Price, and Zipf distributions may be more or less imperfect. If the scope of stakeholders had been restricted to authors of papers in scientific journals, or to a relatively homogeneous group of policy advocates (or opponents), the cumulative frequency distribution (Figure 1(b)) would have flattened out prematurely. For example, in the extreme case of total consensus on rival hypotheses, the cumulative frequency curve would show zero change after the first probe. The other extreme would be total disagreement, and the

cumulative frequency curve would move in the direction of infinity (this is "Foucalt's fallacy" of Fig. 1(b)). In the present case, discordance does not appear to be a threat to context validity, given the proximal similarity of predicted and observed patterns. An expanded sample of stakeholders from the automobile industry, gasoline retailers, truckers, insurance companies, emergency medical services, and environmental protection groups might provide additional rival hypotheses. Given the substance of rival hypotheses elicited from stakeholders (Table 1), it seems likely that the rival hypotheses of additional stakeholders would be similar to those listed, with few if any significant changes in the pattern of the two distributions (Fig. 1(a) and 1(b)).

4.3 Sub-Optimal Elevation

The elevation of the system of rival hypotheses, measured by the average number of constructs per stakeholder, may be suboptimal. Although judgments about optimal elevation appear to require prior knowledge of the relative complexity of a policy system, it is the discovery of this complexity that is the problem to be solved. Policy research in complex systems begins from a state of relative ignorance, the recognition of which is useful and even essential (Ravetz 1990). In the present case the elevation (mean) is approximately 15, given that 38 stakeholders mentioned a total of 718 hypotheses, of which 109 were unique. This number is far greater than that found in published literature on the 55 mph and 65 mph speed limits, which displays a high degree of trust (consensus) with regard to the hypothesis that the 55 mph speed limit caused most of the observed decline in traffic fatalities after 1974. Disagreements, when these exist, are found mainly in literature on the economic efficiency of the 55 mph speed limit. Elevation and discordance, which would be represented by excessive consensus and a premature flattening of the cumulative frequency curve, do not appear to be significant threats to context validity.

5. *Testing rival hypotheses with pragmatic eliminative induction*

Pragmatic eliminative induction enables estimates of the proximal range of rival hypotheses in specific natural contexts. It recognizes that the

identification, testing, and elimination of rival hypotheses is a pragmatic process, not a logical-analytic or metaphysical one. It is not based on rules of formal logic, or on a priori principles. Instead, it is pragmatic because it employs knowledge present in natural systems—what Campbell called naturalistic epistemologies—to make estimates of the proximal range of rival hypotheses. These estimates, which are governed by pragmatic criteria of character, coordination, correctness-in-the-limit, and cost-effectiveness, are optimally plausible in particular problem situations or contexts. Thus, pragmatic eliminative induction, by establishing the approximate context validity of rival hypotheses, permits us to estimate whether we have a full set of hypotheses before they are tested and, where possible, eliminated. The approximate full set of contextually valid rival hypotheses, already displayed in Table 1 and Figures 1(a) and 1(b), contain a number of threats to internal, external, construct, and statistical conclusion validity, are plausible rivals to the initially favored hypothesis that the 55 mph speed limit caused the decline in traffic fatalities after 1974. The most important of these rival hypotheses are tested below.

5.1 Internal (Local Molar Causal) Validity

Internal validity, later relabelled local molar causal validity (Campbell 1986a), refers to the approximate validity of inferences affirming that two or more variables within a complex system of interdependent factors are causally related or unrelated. The emphasis on molar properties, rather than molecular ones, stresses that detailed theoretical specifications and precise empirical estimates of causal linkages among the smaller parts which form larger wholes, are not essential for making causal inferences about these wholes.

The analysis of traffic fatalities, miles of travel, and employment during periods of economic recession and recovery suggests that economic fluctuations (business cycles) affect changes in miles of travel and fatalities (Fig. 2(a)). In turn, changes in employment, a key indicator of recession and recovery, affect changes in miles of travel (Fig. 2(b)).[34] Three identifiable cyclical declines occur during recessions, and two of those recessions (1974-75 and 1980-82) follow sharp increases in oil prices by the Organization of Petroleum Exporting Countries (OPEC). These conclusions provide a serious challenge to the hypothesis that the 55 mph speed limit was responsible for the observed decline in fatalities.

Parallel conclusions have been reached by investigators examining the effect of unemployment on fatalities per 100,000 population in Europe and Canada. The negative correlation between unemployment and fatalities ranges from 0.68 to 0.88 (Wilde 1994, Ch. 5).

The most important threat to internal validity is what may be called the "political economy hypothesis." Stakeholders who put forth this hypothesis contend that observed increases and decreases in traffic fatalities over the thirty-year period 1966-1996 can be explained by the state of the economies of the United States and European countries, and their vulnerability to external petroleum shocks originating in political conflicts between North Atlantic and Middle Eastern countries. This rival hypothesis, which falls in the formal category of *history* as a threat to validity, challenges the claim that the 55 mph speed limit caused the observed decline in fatalities on grounds that one or more events occurring in the time interval between pre-policy and post-policy measurements affect the decline in fatalities. In this context, several stakeholders (see Table 1) argued that events such as 1973-74 OPEC oil embargo, energy crises, the price of petroleum, and exogenous economic factors affected the decline in fatalities

The line graphs showing changes in recessionary periods, fatalities, miles driven, and employment (Fig. 2(a) and 2(b)) supply a systemic, or molar causal, representation of the effects of the "exogenous economic factors" noted by one stakeholder. Additional analyses of relations among fatalities, mileage, and key economic indicators—including coincident, leading, and lagging economic indicators, indices of consumer confidence, and gasoline prices per gallon—yield simple correlation coefficients with the expected signs and magnitudes. The price of gasoline (in constant dollars) is negatively related to miles driven and fatalities, while indicators of the state of the economy and consumer confidence have positive signs. As the state of the economy improves, miles driven and traffic deaths increase. There is a moderate to strong positive cross-correlation between change in miles driven per cent (an index that combines fuel costs and fuel efficiency) and change in miles driven (Fig. 3).[35]

The hypothesis that the 55 mph speed limit affects fatalities may be challenged on other grounds.[36] Stakeholders in 55 mph northeastern states identified the "demonstration effect" on drivers in 55 mph states of changing the policy to 65 mph in the rest of the country. This threat to

**Fig. 2(a). Economic (business) cycles
affect changes in miles of travel and
changes in fatalities**

MILES OF TRAVEL (billions)   FATALITIES (thousands)
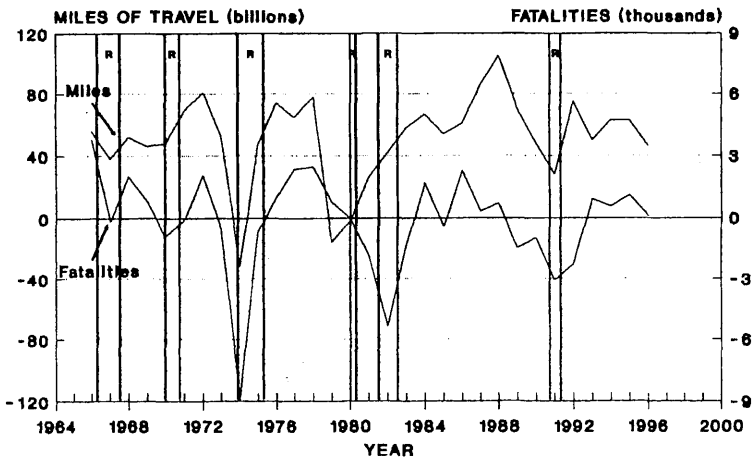
NOTE: Recessionary periods [R] defined
by the National Bureau of Economic
Research. Changes are first differences.

**Fig. 2(b). Changes in employment affect
and are affected by changes in miles
of travel**

MILES OF TRAVEL (billions)   EMPLOYMENT (thousands)
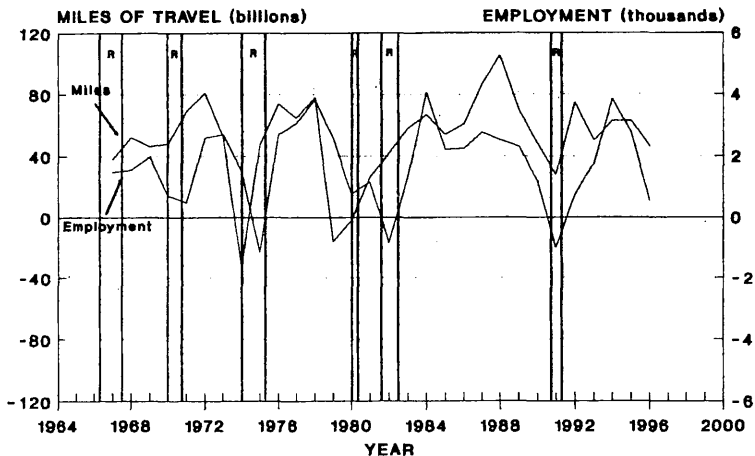
NOTE: Miles of travel estimated on the
basis of average fuel efficiency and
excise taxes paid on gasoline sales.

Fig. 3. Changes in miles of travel
and miles per cent: Effects of fuel
efficiency and fuel costs on driving
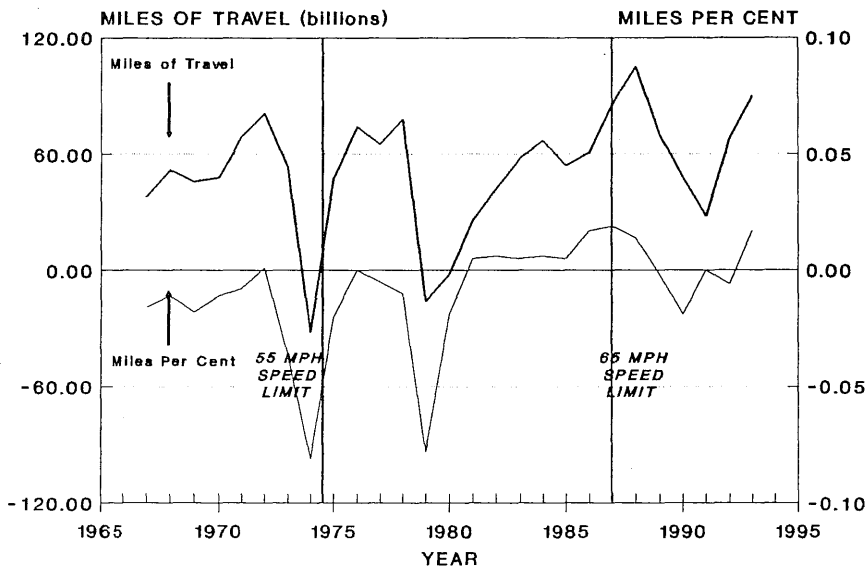
validity, the *imitation of treatments or policies* by controls, was eliminated because fatalities in 55 mph states were not responsive to the small increase in average speeds from 58.7 mph to 58.9 mph between 1986 and 1987 (NHTSA 1989, Table 16, p. 26). Stakeholders in the 55 mph states also identified rival hypotheses of "speed spillover" from 65 mph to 55 mph states and a perceived sense of "unfairness" among drivers in 55 mph states, which represent threats to validity commonly known as *diffusion of treatments or policies, compensatory equalization of treatments or policies,* and *resentful demoralization of controls* who have been deprived of a perceived benefit. Given the relatively small increase in average speeds in 55 mph states, none of these rival hypotheses appears plausible. Indeed, these rival hypotheses may be somewhat beside the point, given that cross-sectional analyses report that there is no statistically significant correlation between average speeds and fatality rates.

So far, the joint effects of employment, miles of travel, and the 55 mph speed limit have not been examined. For this purpose, several MARIMA (Multivariate Autoregressive Integrated Moving Average) models were estimated. These models are useful in representing interactions among these variables and in correcting for serial correlation and random instability (white noise) in the time series. The most economical of these models employs miles of travel, employment, and the 55 mph speed limit (a dummy variable) to explain traffic fatalities. The coefficient which measures the effect of the 55 mph speed limit is statistically significant ($p = 0.01$) only for the year immediately following the implementation of the 55 mph speed limit. The coefficients for miles of travel and employment are statistically significant ($p = 0.05$) over the entire period in which the three recessionary cycles occur. The observed and estimated traffic fatalities per 100,000 population are displayed in Figure 4. The political economy hypothesis provides a plausible rival explanation of the causal nexus among fatalities, miles of travel, and employment. The political economy hypothesis also explains the "counter-intuitive" decline in traffic fatalities which occurred after the maximum speed limit was abandoned.

5.2 Construct Validity

The theoretical constructs forming the propositions in a theory may be

inadequately conceptualized, defined, and measured, creating the impression of an adequate theoretical explanation when none exists. Concepts such as speed or velocity, while they play an essential role in theories and theoretical laws in physics, explain the severity of impact-- and not fatalities and fatal accidents. So-called "ordinary knowledge" based on everyday experiences appears to reinforce the inadequate explication of the concepts speed, accident, and fatality. Average speeds and the standard deviation of these speeds have increased since 1974, but the long-term trend in cyclically fluctuating fatalities is one of continuing decline.

5.3 Statistical Conclusion Validity

Threats to statistical conclusion validity include plausible rival hypotheses which may render the claims about the effects of a policy intervention ambiguous or uninterpretable. In the case of the 55 mph speed limit, there are three potential threats. The instability characteristic of random fluctuations in the time series must be removed by logarithmic transformations of one or more variables and/or the inclusion of a moving average term. If instability is not removed, there is a bias toward Type II errors. Similarly, when observations in the time series are not independent, serial correlation (or auto-correlation) will inflate a significance test, suggesting that the 55 mph has an effect when it does not. The inclusion of an autoregression term in a MARIMA model permits appropriate significance testing and adequate statistical conclusion validity. A MARIMA model requires a minimum of 50-60 time-series observations. Otherwise there is a risk of a Type I error. The three threats to statistical conclusion validity were eliminated by differencing the time series and conducting a sensitivity analysis of the effects of a sample size. Tests for serial correlation and the effect of sample size showed that the results displayed visually in Figure 4 are statistically sound.

5.4 External Validity (Proximal Similarity)

A control series analysis (Fig. 5) was performed to assess the approximate external validity, or proximal similarity, of the political economy hypothesis in other settings. The United States was compared

Fig. 4. ARIMA estimate: Miles driven,
employment rate, and the 55 mph speed
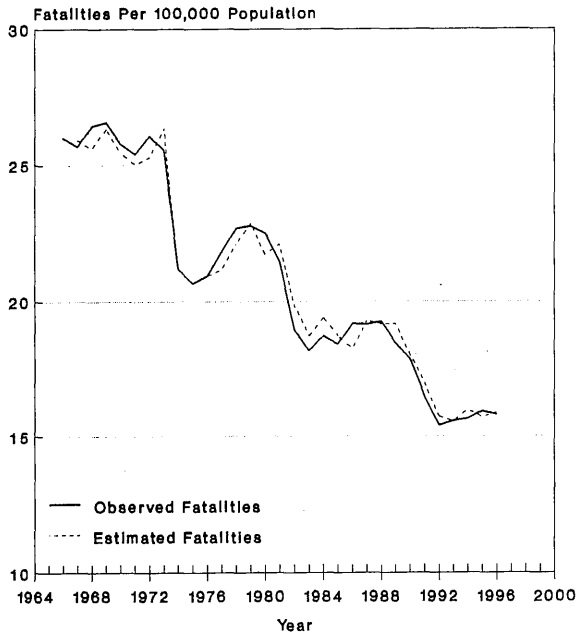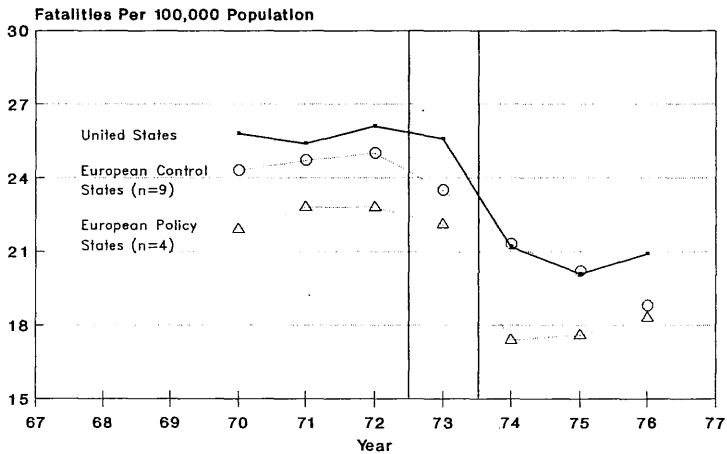limit jointly explain the fatality rate

Fig. 5. Fatality rates in Europe and the
United States: Maximum speed limits have
temporary effect on fatality rates

SOURCE: *Social Indicators III*, Table
5/20.

with European countries which imposed new maximum speed limits of
48-54 mph (80-90 kph) after 1973, as well as countries which did not
impose new limits. Significantly, a highly similar pattern of declining
fatalities per 100,000 population occurred in the United States and in
European countries with and without a new maximum speed limit. This
strongly suggests that the political economy hypothesis applies in Western
Europe as well as the United States. The OPEC oil embargo and its
effects on gasoline prices, and on recessionary cycles, provides a
compelling explanation of traffic fatalities.[37] Although the 55 mph speed
limit had a temporary effect which was eroded after 1975, most of the
variation in fatalities is explained by political and economic factors
associated with the international politics of oil and domestic business
cycles.

6. *Conclusion*

This paper has attempted to show how pragmatic eliminative induction
satisfies four requirements of context validity: character, correctness-in-
the-limit, coordination, and cost-effectiveness. Supplemented by a critical
examination of threats to context validity, pragmatic eliminative induction
is a workable strategy for estimating the proximal range of rival
hypotheses within a disputatious community of knowledgeables. The
broadly favored and intuitively appealing hypothesis that speed limits save
lives is more or less decisively rejected and replaced with what I have
called the political economy hypothesis. This conclusion would have been
unlikely in the absence of context validation and the estimation of a
proximal range of rival hypotheses. Although pragmatic eliminative
induction is always incomplete, historically dated, and fallible, it
nevertheless permits us to estimate this proximal range before we test
and, where possible, eliminate rival hypotheses by means of applied
social experimentation--a methodological innovation for which Campbell
rightly earned the title master methodologist.

University of Pittsburgh

NOTES

1. This chapter draws on my "Campbell's Experimenting Society—Prospect and Retrospect," in W.N. Dunn (ed.), *The Experimenting Society: Essays in Honor of Donald T. Campbell* (New Brunswick, NJ: Transaction Books, 1998) and on an earlier working paper titled "Testing Rival Hypotheses with Pragmatic Eliminative Induction: The Case of National Maximum Speed Limits" (Pittsburgh, PA: University of Pittsburgh, Graduate School of Public and International Affairs, January 1995). Campbell offered historical interpretations, background materials, queries, and suggestions which substantially improved the 1995 working paper and, hence, this article. It was his suggestion to explicitly link the terms "pragmatic" and "eliminative" induction, which I treated separately in the first draft. It is with gratitude and joy that I acknowledge my debt to Don Campbell, a dear friend, mid-life mentor, and master methodologist.

2. Campbell's obituary (*The New York Times*, May 12, 1996) described him as a "master methodologist" who virtually revolutionized the practice of applied social research in this century.

3. The so-called "true" experiment is an experiment which achieves experimenter and statistical controls through a research design (plan, strategy) incorporating pre-treatment and post-treatment outcome measures, an experimental and control group(s), and a three-stage process of randomization--selection of subjects, assignment of subjects to groups, and assignment of the treatment to one of these groups. Campbell believed he had exaggerated the importance of the true experiment (see Dunn 1998, Ch. 1), to the detriment of case studies, hermeneutics, and other qualitative methods, all of which he endorsed and used in his own work (see, e.g., Campbell 1975a, 1978a, 1995b).

4. Ostensibly, eliminative induction is redundant because the effects of rival factors (hypotheses) are eliminated by means of random sampling, or by the standard experimenter controls of scientific laboratories. For arguments against the redundancy thesis see, among others, Knorr-Cetina (1981) and Knorr-Cetina and Mulkay (1983).

5. All page references are to the November, 1996 version of Campbell's paper, edited by Barbara Frankel and Cecilia Heyes. This paper was later published in *Evolution and Cognition* (1997).

6. Nicholas Rescher (1980) develops a methodological-pragmatic theory of truth-estimation by defining truth as a process of satisfying what he calls "parameters of cognitive systematization." He opposes the methodology of falsificationism, including efforts to justify this particular theory of

induction by assuming that rival hypotheses originate in a (metaphysical) "convenient range structure." By way of contrast, what I have called the "proximal range" of rival hypotheses originates in a descriptive epistemology of ongoing knowledge systems. There are no hypotheses—rivals or otherwise—outside these natural knowledge systems.

7. One major critic of Mill's *System of Logic* (1843) was William Whewell, who, in his *The Philosophy of the Inductive Sciences, Founded Upon their History* (1847), advanced his consilience theory of induction. Consilient inductions are inductions involving two or more concordant hypotheses which, based on independently derived observations, are each compatible with the stock of empirical knowledge available at a given moment in history. Mill wanted a methodology for discovering causes by testing hypotheses and rival hypotheses against properties of nature, which he saw as constant and uniform across time. By contrast, Whewell challenged this uniformity of nature assumption, arguing that changes in nature may be catastrophic as well as incremental. Whewell wanted a historical theory of scientific induction. Mill wanted a methodological one.

8. The metaphor is from Kline (1980), who addresses the unsuccessful quest for certainty in mathematics.

9. This notion of proximal range accords with Rescher's (1980) methodological-pragmatic approach to induction, an "erotetic" process of truth-estimation which yields optimally plausible answers to evidence-transcending questions. The term erotetic is related to the belief that approximate truth is a product of thought interacting with external environments, rather than a product of that environment, per se (as if we could know this), or of the knowing subject, per se (as if we could know that).

10. Critics rightly claim that it is just as difficult to use observations to falsify as to corroborate a rival theory or hypothesis. Campbell (1974) and Cook and Campbell (1979), who are fully aware of this problem, cite the Quine-Duhem thesis of the multiple theory-ladenness of observations to argue that single theories and single observations do not in practice falsify hypotheses. Cook and Campbell (1979, pp. 23-4) note that Popper had a similar complex view of conditions for falsification.

11. In the area of risk assessment, a recognition of this state is what Ravetz (1990) describes as "usable ignorance."

12. Contrary to the claims of critics, Cook and Campbell (1979) expressly reject Mill's claims that his canons (methods) of induction yield certain knowledge. Indeed, Mill is offered as an example of "essentialism," a doctrine which does not stand up to the skeptical critique of induction later

developed by Popper, a critique which was implicit in Whewell's consilience theory of induction.

13. This point is similar to standard critiques of Mill which have been available in standard logic texts for many years(see, e.g., Copi 1953, pp. 368-69).

14. Fuller (1992) draws on the pragmatist tradition of Peirce, Dewey, and Mead to argue for the radical naturalization of the study of knowledge, a project which would make epistemology self-referentially normative, social, and experimental (see also Fuller 1988). Fuller's argument leads me to place Campbell among 20th-century pragmatists, given his early and radical naturalization of epistemology (see Campbell 1959) and despite his early doubts about some aspects of pragmatism (Campbell 1974 in 1988, pp. 431-32).

15. The four requirements draw on Rescher (1980) and my later attempts (Dunn 1988, 1993a, 1997) to adapt and incorporate them into a problem structuring methodology for policy research and analysis. Giles Jackson (personal communication) notes that the requirements represent "4Cs."

16. A high ratio of trust to doubt among leaders of scientific and political communities does not promote vigorous debate, creating a condition which Lindblom (1990) describes as cognitive impairment. In this context, Campbell's conclusion (1988, pp. 482-84) that there is a 99:1 trust-doubt ratio in most knowledge systems may seem pessimistic. Interestingly, a study of revolutionary leaders by Suedfeld and Rank (1976) suggests that success in carrying out a revolution is associated with a high trust-doubt ratio accompanied by a simple conceptual structure, which may be interpreted as an adaptive restriction on the number of doubt-inducing constructs. Revolutionaries who succeeded after as well as during a revolution (Cromwell, Washington, Stalin, Mao Tse-tung) seem to have a lower trust-doubt ratio and a more complex conceptual structure than those who were unsuccessful after the revolution (Trotsky, Guevara, Hamilton).

17. The metaphor of a laboratory meat grinder explains why random selection does not contribute to the requirement of coordination. Researchers investigating a particular strain of laboratory rat placed the animals in a meatgrinder, subsequently analyzing a random sample of n elements of ground rat in the foolish hope that subsequent statistical analyses would yield knowledge of the rat's anatomy and physiology.

18. The construction of a cost-effectiveness curve, displayed as an ogive, shows that the point of tangency (optimal efficiency) lies at about the 85th percentile. Thus, an additional $0.15n$ probes, with n equal to the total number of stakeholders in a sample, would be necessary to reach the proximal limit of rival hypotheses.

19. The backing for this claim originates in the application of information theory to theoretical linguistics. See Lyons (1969).

20. Much like the "tragedy of the commons," economies at the individual level do not translate into economies at the collective level. The "principle of least effort" (Zipf 1949) may be applied at both levels, with very different consequences. With Campbell (1986b, 1994), we are interested in economies of cognition—including cost-effective validity improvements—in knowledge systems.

21. See Danforth's op-ed article in *The Atlanta Constitution* (February 12, 1987) and the case materials on the 55 mph speed limit in Guess and Farnham (1989).

22. The twelve states not responding to the Department of Transportation request for a review of their report include three northeastern states which have maintained the 55 mph speed limit (Connecticut, Maryland, Rhode Island). Other non-responding states changed to 65 mph on rural interstate highways in 1987 or 1988 (Arkansas, Indiana, Kansas, New Hampshire, New Mexico, Tennessee, Vermont, and Virginia).

23. Given the experiential and scientific knowledge of the 38 stakeholders, all rival hypotheses were judged to be plausible (i.e., they are taken to have a prior probability greater than zero). When they are generated in the manner described here, all rival hypotheses have a prior probability (plausibility) greater than zero.

24. The declining probative value of hypotheses (and of scientific trust) is explained by Lyons (1969, p. 89): "Information-content varies inversely with probability. The more predictable a unit is, the less meaning it has. This principle is in accord with the commonly-expressed view of writers on style, that cliches (or 'hackneyed expressions' and 'dead metaphors') are less effective than more 'original' turns of phrase."

25. This shareware program is ANTHROPAC 3.2, developed by Stephen P. Borgatti, Department of Sociology, University of South Carolina, Columbia SC, 29208.

26. For a stimulating essay on these distributions, see Simon (1972), who focuses on the work of Harvard socio-linguist George K. Zipf. The classic syntheses are provided in Zipf's studies of psycho-biology and language (1932, 1935) and in his major work, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (1949). Simultaneous and apparently independent work on variants of the same distribution include Lotka (1926) and Bradford (1934). A thorough but somewhat dated review of work based on the Lotka-Bradford-Zipf distribution, along with a

mathematical correction proposed by Benoit Mandelbrot, may be found in Subramanyam (1975). In quantitative social studies of science, Derek de Solla Price (1986) refines some of these distributions.

27. Among the many ways to explore quantitative estimates of similarity and goodness-of-fit are several chi-square tests (e.g., homogeneity). OLS regression (linear and nonlinear) may be used to estimate the probability of a Type III error (testing the wrong rival hypotheses) by means of a coefficient of indetermination equal to $1-R^2$). Although these quantitative estimates (see Dunn 1993) may be more accurate than visual ones, Tufte (1983, 1997) argues persuasively that the two are complementary.

28. I have rephrased Zipf's law, which he states (1949, p. 1) as follows. A person "will strive to solve his problems in such a way as to minimize the *total work* that he must expend in solving *both* his immediate problems *and* his probable future problems...*as estimated by himself*" (emphasis original). The Bradford distribution can be explained by Robert Merton's Mathew Effect (Merton 1973, 221-78), and the principle of cumulative advantage stated by Price (1986, pp. 257-64).

29. Rescher (1980, p. 36) characterizes this process of truth-estimation as cognitive systematization, which he describes as a "Hegelian inversion." This inversion replaces "systematic because true (because truth is a system)" with "true because systematic (since systematicity is the test of truth)" (p. 37).

30. That diverse knowledge systems are proximally similar to the same hyperbolic distribution, with each system governed by diminishing rates and limits, may be attributed to properties of the process that generates the distribution. From the standpoint of naturalistic epistemology, these processes include imitation and association (see Simon 1972, p. 199), as well as honesty and deception, conformity and non-conformity, and the ratio of trust to doubt (Campbell 1988, pp. 482-44). The highly uneven distribution of trust and doubt in knowledge systems (although "subsystems" of dogmatic or self-assured scientists may trust themselves completely) is an enormous advantage rarely exploited by social scientists who seek to estimate the context validity of plausible rival hypotheses.

31. It may be argued that this kind of analysis involves "meta" hypothesis testing at the collective or group level, with members of the group consisting of individual rival hypotheses. This argument distinguishes class (the proximal range of rival hypotheses) and member (individual rival hypotheses within the range).

32. The methodology of pragmatic eliminative induction is not based on logical-analytic argumentation, but on the analysis of comparative plausibilities among many rival hypotheses—theoretical, methodological, and substantive. It therefore differs from the following standard strategy which is designed to avoid infinite regress in the search for rival hypotheses. This strategy, which is of little use in dealing with problems addressed in this paper, "solves" the problem of infinite regress by specifying a hypothesis {X is hypothesized to cause Y} as the major premise of a logically valid deductive argument of the following form: {X is hypothesized to cause Y} {Y was observed to follow X on n occasions when Z was also present. But Y was not observed at all on n occasions when Z alone was present} ∴ {X causes Y}. The warrant for the hypothesis need not be a causal law. It may be a practical action, as in the activity theory of causation (Cook and Campbell 1979, pp. 25-28), or an optimally plausible enthymematic (contingent) premise, as in practical policy reasoning (Dunn 1994, pp. 97-100). Both of these conform to Rescher's methodological-pragmatic approach to inductive argumentation (Rescher 1980, pp. 11-14).

33. The determination of the "value" of information from its reported use, and the notion of "revealed preferences," are examples of the problem of inadequate responsiveness. In this context, the classic critique is Machlup (1980).

34. The analysis of cross-correlations shows that the hypothesized effects of employment on miles driven and fatalities are greater than the effects of either on employment, at lags of order 1-8.

35. Miles per cent, a measure of average fuel cost, is based on *Social Indicators III* (December 1980), Table 4/23, p. 200. Miles per cent includes changes in the average cost of gasoline per gallon (constant 1982 dollars) and average fuel efficiency (EPA estimate).

36. Space constraints do not permit a discussion of other rival hypotheses advanced by stakeholders. Miles of travel are estimated on the basis of state excise taxes on gallons of gasoline sold, with different states keeping more or less reliable records (instrumentation). Western states which abandoned the 55 mph limit have low population density and high fatality rates, while the reverse is true for northeastern states retaining the 55 mph limit. Population density, not speed limits, accounts for most of the difference (selection). Fatalities reached an extreme level (within the three regular cycles) immediately before the speed limit was increased in 1974 and just after it was abandoned in 1986-87 (regression toward the mean). Part of the decline in fatalities is due to the "migration" of motorists from slower (and more dangerous) local roads to faster (and safer) interstate highways

(mortality). Other rival hypotheses include mandatory seat belt laws, anti-drunk driving legislation, and the growth of emergency medical services, all of which account for some part of the long-term secular decline.

37. Campbell (personal communication) noted that the results presented here are equally or more definitive than findings he and others have reported with respect to similar social experiments (see Ross and Campbell, 1968; Ross, Campbell, and Glass, 1970).

REFERENCES

References to papers available in Campbell's selected papers are designated below as **SP**, which refers to Donald T. Campbell, *Methodology and Epistemology for Social Science: Selected Papers*, ed. E. Samuel Overman (Chicago, IL: The University of Chicago Press, 1988).

Alcoff Linda Martin (1996), *Real Knowing: New Versions of the Coherence Theory*. Ithaca, NY: Cornell University Press.

Bradford Samuel C. (1934), "Sources of Information on Specific Subjects." *Engineering* 137: 85-6.

Brewer Marilynn B. and Barry E. Collins (eds.) (1981), *Scientific Inquiry and the Social Sciences: A Volume in Honor of Donald T. Campbell*. San Francisco, CA: Jossey-Bass.

Campbell Donald T. (1957), "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54: 297-312.

Campbell Donald T. (1959), "Methodological Suggestions from a Comparative Psychology of Knowledge Processes." *Inquiry* 2: 152-82.

Campbell Donald T. (1960), "Blind Variation and Selective Retention in Creative hought as in Other Knowledge Processes." *Psychological Review* 67: 380-400.

Campbell Donald T. (1964), "Distinguishing Differences of Perception from Failures of Communication in Cross-Cultural Studies," *Cross Cultural Understanding: Epistemology in Anthropology*, edited by F.S.C. Northrop and H.H. Livingstone. 308-36. New York: Harper and Row.

Campbell Donald T. (1966), "Pattern Matching as an Essential in Distal Knowing." In *The Psychology of Egon Brunswik*, edited by K.R. Hammond, 81-106. New York, NY: Holt, Rinehart, and Winston.

Campbell Donald T. (1967), "Administrative Experimentation, Institutional Records, and Nonreactive Measures." In *Improving Experimental Design and Statistical Analysis*, edited by J.C. Stanley and S.M. Elam. Chicago, IL: Rand McNally. *SP*, 243-60.

Campbell Donald T. (1969a), "Definitional versus Multiple Operationalism." *et al.* 2: 14-17. *SP*, 31-36.

Campbell Donald T. (1969b), "Reforms as Experiments." *American Psychologist* 24: 409-29. *SP*, 261-289.

Campbell Donald T. (1974a), "Evolutionary Epistemology." In *The Philosophy of Karl Popper*, edited by P.A. Schilpp, 413-63. Lasalle, IL: Open Court Press. *SP*, 393-434.

Campbell Donald T. (1974b), "Quasi-Experimental Designs. In *SP*, 191-221.

Campbell Donald T. (1975a), "'Degrees of Freedom' and the Case Study." *Comparative Political Studies* 8: 178-93. *SP*, 377-388.

Campbell Donald T. (1975b), "On the Conflicts Between Biological and Social Evolution and Between Psychology and Moral Tradition." *American Psychologist* (December): 1103-1125.

Campbell Donald T. "Reprise." (1976), *American Psychologist* (May): 381-84.

Campbell Donald T. (1977), "Descriptive Epistemology: Psychological, Sociological, and Evolutionary." William James Lectures, Harvard University. *SP*, 435-86.

Campbell Donald T. (1978a), "Qualitative Knowing in Action Research." In The Social Context of Method, edited by M. Brenner, P. Marsh, and M. Brenner, 184-209. London: Croom Helm. *SP*, 360-76.

Campbell Donald T. (1978b), "Response to Dyer." *American Psychologist* 33: 771-72.

Campbell Donald T. (1979a), "Assessing the Impact of Planned Social Change." *Evaluation and Program Planning* 2: 67-90.

Campbell Donald T. (1979b), "Comments on the Sociobiology of Ethics and Moralizing." *Behavioral Science* 24: 37-45.

Campbell Donald T. (1979c), "A Tribal Model of the Social System Vehicle Carrying Scientific Knowledge." *Knowledge: Creation, Diffusion, Utilization* 2: 181-201. *SP*, 489-503.

Campbell Donald T. (1981), "ERISS Conference." *Society for Social Studies of Science Newsletter* 6, 3 (Summer): 24-5.

Campbell Donald T. (1982), "Experiments as Arguments." *Knowledge: Creation, Diffusion, Utilization* 3: 327-37.

Campbell Donald T. (1984), "Can We Be Scientific in Applied Social Science?" Evaluation Studies Review Annual 9, edited by R. Connor, D.G. Altman, and C. Jackson, 26-48. *SP*, 315-334.

Campbell Donald T. (1986a), "Relabelling Internal and External Validity for Applied Social Scientists." In *Advances in Quasi-Experimental Design and Analysis*, edited by W.M.K. Trochim, 67-77. San Francisco, CA: Jossey-Bass.

Campbell Donald T. (1986b), "Science's Social System of Validity-Enhancing Collective Belief Change and the Problems of the Social Sciences." In *Metatheory and Social Science: Pluralisms and Subjectivities*, edited by

D.W. Fiske and R.A. Schweder, 108-35. *SP*, 504-25.

Campbell Donald T. (1987), "Guidelines for Monitoring the Scientific Competence of Preventive Intervention Research Centers: An Exercise in the Sociology of Scientific Validity." In "National Policies for Optimizing Validity in Applied Social Science," edited by W.N. Dunn, *Knowledge: Creation, Diffusion, Utilization* 8, 4 (March 1987): 389-430.

Campbell Donald T. (1988a), *Methodology and Epistemology for Social Science: Selected Papers*, edited by E. Samuel Overman. Chicago, IL: The University of Chicago Press.

Campbell Donald T. (1988b), "Perspective on a Scholarly Career." *SP*, 1-30.

Campbell Donald T. (1988c), "Systems Theory and Social Experimentation." Paper prepared for the U.S.S.R-U.S. Conference on Systems Theory and Management, May, 1988. U.S. coordinator: Prof. S.A. Umpleby (Management Science, George Washington University).

Campbell Donald T. (1994), "Foreword: Toward a Sociology of Scientific Validity." In Kyung-Man Kim, *Explaining Scientific Consensus: The Case of Mendelian Genetics*, ix-xviii. New York, NY: The Guilford Press.

Campbell Donald T. (1995a), "In What Sense Does a Selectionist Model Provide Epistemological 'Justification' for *Scientific* Beliefs." Paper prepared for John Ziman's "The Epistemology Group: The Evolution of Knowledge and Invention." Royal Society of Arts, London. May 24.

Campbell Donald T. (1995b), "The Postpositivist, Nonfoundational, Hermeneutic Epistemology Exemplified in the Works of Donald W. Fiske," edited by Patrick Shrout and Susan Fiske, 13-27. *Personality, Research Methods, and Theory*. Hillsdale, NJ: Lawrence Erlebaum.

Campbell Donald T. (1996), "From Evolutionary Epistemology Via Selection Theory to a Sociology of Scientific Validity." Revised and edited by Barbara Frankel and Cecilia Heyes. Also in *Evolution and Cognition* (1997).

Campbell Donald T. (1998), "The Experimenting Society." In W.N. Dunn, *The Experimenting Society: Essays in Honor of Donald T. Campbell*. New Brunswick, NJ: Transaction Books. Vol. 12 of *Policy Studies Review Annual.*

Campbell Donald T. and Donald W. Fiske (1959b), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56 (1959): 81-105. *SP*, 37-61.

Campbell Donald T. and Robert F. Boruch (1975), "Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects." In *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*, C.A. Bennett and A.

Lumsdaine, eds. 195-296. New York: Academic Press.

Campbell Donald T. and H. Lawrence Ross (1968), "The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis." *Law and Society Review* 3: 33-53. *SP*, 222-37.

Campbell Donald T. and Bonnie T. Paller (1989), "Extending Evolutionary Epistemology to 'Justifying' Scientific Beliefs (A Sociological Rapprochement with a Fallibilist Perceptual Foundationalism?)." In *Issues in Evolutionary Epistemology*, ed. Kai Hahlweg and C.A. Hooker, 231-57. Albany, NY: State University of New York Press.

Campbell Donald T. and Julian C. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.

Cook Thomas D. (1985), "Postpositivist Critical Multiplism." In *Social Science and Social Policy*, edited by R.L. Shotland and M.M. Mark, 21-62. Beverly Hills, CA: Sage Publications.

Cook Thomas D. and Donald T. Campbell (1979), *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.

Copi Irving M. (1953), *Introduction to Logic*. New York: Macmillan.

Dery David (1985), *Problem Definition in Policy Analysis*. Lawrence, KS: University of Kansas Press.

Dunn William N. (1982), "Reforms as Arguments." *Knowledge: Creation, Diffusion, Utilization* 3: 293-326.

Dunn William N. (1988), "Methods of the Second Type: Coping with the Wilderness of Conventional Policy Analysis." *Policy Studies Review* 7, 4: 720-37.

Dunn William N. (1993a), "Discovering the Boundaries of Ignorance in Policy Research: A Continuous Measure of Errors of the Third Type." In *Between Sociology and Sociological Practice: Essays on Social Policy Research-- Liber Amicorum Dedicated to Mark van de Vall*, Edited by B.M.J. Wolters, K.M. Schultz, J.T.A. Koster, and F.L. Leeuw, 36-62. Nijmegen: Institute for Applied Social Sciences.

Dunn William N. (1993b), "Policy Reforms as Arguments." In *The Argumentative Turn in Policy Analysis and Planning*, edited by F. Fischer and J. Forester, 254-290. Durham, NC: Duke University Press.

Dunn William N. (1995), "Testing Rival Hypotheses with Pragmatic Eliminative Induction: The Case of National Maximum Speed imits." Pittsburgh, PA: Graduate School of Public and International Affairs, Working Paper. January.

Dunn William N. (1997), "Probing the Boundaries of Ignorance in Policy Analysis." *American Behavioral Scientist* 40, 3: 277-98.

Fischer Frank and John Forester (eds.) (1993), *The Argumentative Turn in*

Policy Analysis and Planning. Durham, NC: Duke University Press.

Fuller Steve (1988), *Social Epistemology*. Bloomington, IN: Indiana University Press.

Fuller Steve (1992), "Epistemology Radically Naturalized: Recovering the Normative, the Experimental, and the Social." In *Cognitive Models of Science,* edited by R. Giere, 427-59. Minneapolis, MN: University of Minnesota Press.

Guess George M. and Paul G. Farnham (1989), *Cases in Public Policy Analysis*. New York: Longman.

Kline Morris (1980), *Mathematics: The Loss of Certainty*. New York, NY: Oxford University Press.

Knorr-Cetina Karin D. (1981), *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.

Knorr-Cetina Karin D. and Michael Mulkay (1983), "Introduction: Emerging Principles in Social Studies of Science." In *Science Observed*, edited by K.D. Knorr-Cetina and M. Mulkay, 1-17. Beverly Hills, CA: Sage Publications.

Kuhn Thomas (1962), *The Structure of Scientific Revolutions*. Chicago, IL: The University of Chicago Press.

Larson Richard C. and Amadeo R. Odoni (1981), *Urban Operations Research*. Englewood Cliffs, NJ: Prentice Hall.

Laudan Larry (1984), *Science and Value*. Berkeley, CA: University of California Press.

Light Richard J. and David B. Pillemer (1984), *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press.

Lindblom Charles E. (1990), *Inquiry and Change: The Troubled Attempt to Understand and Shape Society*. New Haven, CN: Yale University Press.

Lotka A.J. (1926), "The Frequency Distribution of Scientific Productivity." *Journal of the Washington Academy of Sciences* 16,2: 317-23.

Lyons P. (1969), *Theoretical Linguistics*. New York: Oxford University Press.

Machlup Fritz (1980), *Knowledge: Its Creation, Distribution, and Economic Significance*. Princeton, NJ: Princeton University Press.

MacRae Duncan Jr. (1976), *The Social Function of Social Science*. New Haven: Yale University Press.

MacRae Duncan Jr. (1985), *Policy Indicators: Links Between Social Science and Public Debate*. Chapel Hill, NC: University of North Carolina Press.

Merton Robert K. (1973), *The Sociology of Science*, edited by N.W. Storer. Chicago, IL: The University of Chicago Press.

Mill John Stuart. (1843), *A System of Logic*. London.

Mitroff Ian I. (1974), *The Subjective Side of Science: A Philosophical Inquiry into the Psychology of the Apollo Moon Scientists*. New York, NY:

Elsevier.

Mulkay Michael (1979), *Science and the Sociology of Knowledge*. London: Allen & Unwin.

Overman E. Samuel (1988), "Introduction: Social Science and Donald T. Campbell." In *SP*, vii-xix.

Popper Karl R. (1963), *Conjectures and Refutations*. New York, NY: Basic Books.

Popper Karl R. (1968), *The Logic of Scientific Discovery*. New York, NY: Basic Books.

Price Derek de Solla (1986), *Little Science, Big Science . . . and Beyond*. New York: Columbia University Press.

Quine W.V. (1960) *Word and Object*. Cambridge, MA: MIT Press.

Ravetz Jerome R. (1990), *The Merger of Knowledge with Power: Essays in Critical Science*. London: Mansell Publishing Ltd.

Rescher Nicholas (1980), *Induction*. Pittsburgh, PA: University of Pittsburgh Press.

Ross H. Lawrence and Donald T. Campbell (1968), "The Connecticut Speed Crackdown: A Study of the Effects of Legal Change." Pp. 33-53 in H.L. Ross, ed., *Perspectives on the Social Order: Readings in Sociology*. New York: McGraw-Hill.

Ross H. Lawrence, Campbell, Donald T., and Gene V. Glass (1970), "Determining the Social Effects of a Legal Reform: The British "Breathalyser" Crackdown of 1967." *American Behavioral Scientist* 13: 493-509.

Segall M.H., Campbell Donald T., and Melville J. Herskovitz (1966), *The Influence of Culture on Visual Perception*. Indianapolis, IN: Bobbs-Merrill.

Simon Herbert A. (1972), "The Sizes of Things." In Judith Tanur et al., eds., *Statistics: A Guide to the Unknown*. New York: Holden Day.

Subramanyam K. (1975), "Laws of Scattering." Pp. 336-54 in Alan Kent, ed., *Encyclopedia of Information Science*. New York: Marcel Dekker.

Suedefeld P. and A. Rank (1976), "Revolutionary Leaders: Long-Term Success as a Function of Changes in Conceptual Complexity." *Journal of Personality and Social Psychology* 34: 169-78.

Tufte Edward R. (1974), *Data Analysis for Politics and Policy*. Englewood Cliffs, NJ: Prentice Hall.

Tufte Edward R. (1983), *The Visual Display of Quantitative Information*. Cheshire, CN: Graphics Press.

Tufte Edward R. (1997), *Visual and Statistical Thinking: Displays of Evidence for Making Decisions*. Cheshire, CT: Graphics Press, 1997.

Wilde Gerald J.S. (1994), *Target Risk: Dealing with the Danger of Death, Disease and Damage in Everyday Decisions*. Toronto: PDE Publications.

Wilde Gerald J.S. and Sebastian L. Simonet (1996), *Economic Fluctuations and the Traffic Accident Rate in Switzerland: A Longitudinal Perspective*. Berne: Swiss Council for Accident Prevention.

Zipf George K. (1932), *Selected Studies in the Principle of Relative Frequency of Language*. Cambridge, MA: Harvard University Press.

Zipf George K. (1935), *Psycho-Biology of Language*. Boston, MA: Houghton Mifflin.

Zipf George K. (1949), *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Chicago, IL: Addison-Wesley.