RUNNING HEAD: DEEP AUDIT

The Deep Audit as an Epistemology for the Watchdog:

Computer-assisted Reporting and Investigative Journalism

John E. Newhagen

Associate Professor

Philip Merrill College of Journalism

University of Maryland

College Park, MD 20742

jnewhagen@jmail.umd.edu

A paper presented to the Theory and Methodology Division of the Association for Education in Journalism and Mass Media, Miami, FL. August 7-10, 2002

Abstract

Computer-assisted data analysis, sometimes called computer-assisted reporting, has emerged as an innovative form of investigative journalism during the past 35 years, especially in the 1980s and 1990s. Its legitimacy has been validated by prizes and awards, but it has failed to become fully integrated into the journalistic routine. That fact is evidenced by the ambiguous location that practitioners of computer analysis occupy on newsroom organizational charts, frequently classified as members of "special projects" teams. The uncertain place of computer analysis may be due in part to journalism's plodding embrace of the computer as a reporting tool. On the other hand, it may reflect a deeper problem, that the practice has not been fully understood in journalism. This paper looks at the idea that the use of computers to clean, merge, and scrutinize large government databases in a quest for newsworthy patterns represents an unique methodology -- referred to here as the *deep audit* -- that is especially useful in alerting communities and their leaders to the need for action to correct misaligned social policy. This discussion suggests the methodology may have much broader application in public policy management and deserves consideration as more than simply a high technology journalistic tool. Time is used to situate and compare the deep audit along a continuum of other research methodologies. An examination of homicide is used as an example of how asking the same question various levels of analysis yields different conclusions, suggesting the special contribution that the deep audit has to offer. Finally the risks of mistakes within various levels of analysis will be discussed to further situate the deep audit as a unique and important research methodology in its own right.

The Deep Audit as an Epistemology for the Watchdog: Computer-assisted Reporting and Investigative Journalism

Scholarly researchers either implicitly or explicitly constrain the questions they ask within a level of analysis. Traditional labels such as “individual” or “aggregate” come to mind and are usually associated with an academic discipline, such as psychology or sociology respectively. Similarly, levels of analysis also suggest epistemological frameworks and methodologies. Psychology, for instance might employ experiments while sociology is more apt to apply surveys to generate empirical data. Journalists, however, usually do not think of themselves as part of this continuum, even though they may be engaged in a research function. This paper argues that journalists who use computers to analyze large databases are practicing an epistemologically unique methodology that stands between the individual and aggregate levels of analysis.

Another way to mark the boundaries of levels of analysis is by the time needed for a problem solving cycle to take place. Newell (1990) elaborates a scale of human actions according to a set of time bands that roughly correspond to theoretical levels of analysis. The bands range from evolution, which takes a very long time, to biology, which takes a very short time. One important feature of this framework is that it relies heavily on the information processing component of human action. There are certainly behavioral and physical constraints on these processes. However, human problem solving can best be seen as a description of the processes used to acquire and organized enough information to arrive at workable solutions.

Working within the information-processing framework suggests discrete epistemological “contexts” or temporal domains worth considering for their comparative value. The most fundamental level shows how individuals apply psychological heuristics, a kind of pseudo-statistical process, to solve problems immediate to their existence. Above that, the social band informs us about aggregations of humans, such as cultural, political, or social groups.

Newell (1990) also describes a rational band that fits between the social and psychological levels. It has to do with problems that may have a direct bearing on individuals, but whose scope is beyond their capacity to independently solve. A range of local public policy issues, such as law enforcement, school quality, or zoning, can have direct and important effects on the course of individuals' lives but are too remote and arcane to be monitored closely by them. This paper will argue that problems in this rational level can be dealt with through the application of a *deep audit*, accomplished by the merging, cleaning, and close scrutiny of public databases by investigative journalists, represents a unique epistemological solution to this midrange set of problems.

<u>Matching the Problem Spaces to Epistemologies</u>

These epistemological approaches can also be understood in terms of the implications of their outcomes. While discussion of the societal and psychological levels is fairly well developed, the rational level has not achieved the same status as a discrete way of knowing. Each of these three epistemological levels of analysis will be described as unique approaches engineered to address discrete sets of problems.

<u>The Societal Level</u> Things taking place at the level of social aggregation regulate how we organize and control ourselves as groups. Beniger (1986), for instance, details how the rationalization of production during the Industrial Revolution led to the growth of information processing bureaucracies intended to control increasingly complex human interactions. Giddens (1984) describes how humans deal with the problem of aggregating political power into

institutional structures. This social level is where norms, customs, and public policy are made, and is best suited to the epistemology of hypothesis testing and the application of the stochastic inference suggested by Bayesian probability. While the efficacy of such decisions may have profound effects on the course of human life, they are generally not immediate or personal. Discourse about methodology at this level of research is generally confined within a small academic elite, where a refereed scientific journal may have a circulation of only a few thousand.

Efforts to popularize social science as journalism have proved to be somewhat awkward, with scientists frequently claiming such attempts fail to capture the nuance of their work. One important exception has been the application of survey methodology to public opinion polling. Meyer (1991) spearheaded the movement to bring the rigor of social science survey techniques to the mainstream of the newsroom nearly 30 years ago. However, while public opinion polls have come to be commonplace in public affairs reporting, their status is still somewhat ambiguous. Some critics argue that polls tend to focus only on the "horse race" dimension of who will win an election at the cost of deeper analysis of issues. Other critics note that when polls are outsourced to professional consultants, journalists who report on the results may not fully understand the methods and statistics employed. Further, news reports may reduce data to percentages gleaned from simple descriptive statistics or the two-way cross tabulation of data, burying or distorting the inferential power of statistics such as analysis of variance or linear regression that this method has to offer. Even where journalists versed in sophisticated statistical techniques do work at the caliber of peer-reviewed research, translating inferential statistics to an audience of news consumers in a convincing fashion poses special challenges.

The Individual Level How individuals make sense out of their immediate surroundings is described by the psychological end of the scale of levels of analysis. Here the problem is to confront a complex information ecology with limited cognitive resources and make survival-contingent decisions in real time (see Geiger & Newhagen, 1993; Lang, 2000). Time stands out as the constraint that dominates this level because, from a purely functional view, the stakes in an archetypical scenario may be very high -- sometimes even life itself. Because this is so, the rules of the road within the individual level are generally either hardwired into the human brain or so highly over learned that they behave as if they were hardwired (Ekman, 1992). The human information processing apparatus has come to be the result of millennia of evolutionary change, and is by no means specifically designed to deal with the pressures of modern life. Nevertheless, individuals frequently show substantial durability, where, for instance, the processes designed to control attention in a walk through a hostile forest on the daily hunt come into play in guiding an automobile along a six-lane freeway every morning.

The Rational Level One ironic similarity between the abstraction of social science and the concrete reality of everyday problem solving is that the processes that shape them are largely outside of our direct control. Societal level change is simply too grand to be addressed by individual action, while the urgency of daily life dictates that our responses to many situations are largely generated automatically and without much deliberate thought.

This paradox leads to the consideration of the *deep audit* as one solution to a unique epistemological problem: how to study the gray area between individual and social structures that may be under direct human control. This range of problems open to deep audit is situated within a manageable time frame – say months or weeks. It involves rational or deliberate information processing strategies. Those processes are not, however, raised to the abstraction of inferential statistics or social science. Neither are they reduced to the application of the automatic heuristic processes frequently used by individuals to deal with the concrete reality of daily experience.

The key to this problem-solving method has to do with bringing the two adjacent levels of analysis into conformity rather than changing one to suit the other. It is accomplished by manipulating complex data universes in the hope of reducing their dimensionality to the point where discrete patterns become apparent. While journalism may be the highest profile social institution practicing deep audit, it is by no means the only one. A full range of institutions dedicated to the implementation and monitoring of social policy, such as public interest groups and government regulatory agencies, also employ this practice. When the deep audit of public data is employed as an epistemology in a democracy, it implies advocacy – that is, it calls for leaders and policy makers to correct specific problems – and helps serve the “watchdog function” of journalism.

This paper will review the action bands for the epistemologies directly above and below the deep audit in the context of Bayesian probability to better understand the deep audit’s unique function. The study of homicide will be used as an example of how inquiry across levels of analysis yields qualitatively unique insights. Finally, the implications of such a scheme are discussed in terms of what happens when the mechanisms of such belief systems fail.

Time as a Constraint on Empirical Knowledge

Time is one of the most important -- and perhaps most neglected -- constraints on the ways humans understand the world around them. Functionalism in cognitive psychology offers one exception, were the role of time is central to the theories generated during the last half century (Greenwald, 1992). That paradigm proposes that organisms must make life and death decisions about rapid changes in their environment in the context of real time if they hope to survive. Real time allows little margin for error and seldom gives the participant in life's drama a chance to go back for a second try. At this level the individuals cannot afford the luxury of deliberation afforded social scientists or journalists when they decide how to respond to imminent danger. It is further important to point out that questions asked at this level are not arbitrary. They may be subjective, but they are guided by the introspective imperative that survival is better than extinction.

At the levels of analysis above the individual the onus of the real time constraint can be relaxed, allowing for qualitatively different problem solving strategies. The relaxation of the time constraint is further reflected by the normative or subjective nature of the questions asked at higher levels of analysis, where the mandate for survival is more distant. This last point is important in bringing the postmodern critique (see Derrida, 1983) into line with the epistemology of empirical science.

Two things, the complexity of the information ecology that surrounds us and the limited mental resources we bring to it, dictate that the most functional way of solving problems involves making the best guess possible given the time allowed by a given problem space. This process usually takes place within a stochastic framework, based on some form of statistical probability. "Problem space" here refers to the range of all possible ways of configuring samples in the arena of inquiry. The problem space serves as the context for any specific sample being used to detect patterns and solve a problem. Because of this stochastic framework, the best place to begin the discussion its at the highest level of statistical certainty, the abstraction of Bayesian probability.

<u>Bayesian Probability.</u> Blalock (1979), in a seminal text on basic social science statistics, demonstrates how "if repeated random samples of size N are drawn from any population having

a mean μ and a variance σ^2, then as N becomes large, the sampling distribution of the sample means approaches normality" (p183). This means that the more times a perfect roulette wheel, with 36 numbers, 0, and 00, is spun, the more likely each of the 38 outcomes will appear every 1 out of every 38 tries. This powerful idea, called the Central Limit Theorem is the motor driving most the inferential statistics used to validate hypotheses in social science. The theorem makes the practice of social science possible because it allows for inference to be drawn from a reasonably small sample taken from a very large population. Otherwise questions asked at this level would require canvassing the entire population, a feat attempted by the U.S. government only once a decade.

There is, however, an important constraint on this process usually not taken into consideration. To truly reach the perfect normal distribution, the wheel of probability will have to be spun an infinite number of times, which implies an infinite period of time in which the task can take place. The importance of unlimited time to abstract probability, and its implications on the inferential statistics employed by social science, has gone without much consideration until some social scientists gave the tenants of postmodernism a closer look. Cohen (1994), for instance, points out that probability, or random sampling, does not guarantee a perfect reflection of the underlying population, even when its N is very large. Increasing the size of the sample may make inferential statistics more sensitive to systematic differences, but it may also be misleading in the sense that random outcomes will more likely return Type I errors, that is, saying something is going on when it is not.

Thus, the epistemology best suited to the abstract consideration of an unbounded system is the thought experiment and relies heavily on the logic of the predicate calculus. One of the best known of such mental exercises is the challenge of 19th century Scottish physicist James

Clerk Maxwell to the Second Law of Thermodynamics, or entropy. Maxwell proposed that a demon might sit on the valve between two glass jars filled with gases of different temperature. Entropy holds that over time the temperatures in the two jars will come to be equal once the valve is opened. Maxwell proposed that from time to time his demon might let a large enough number of hot molecules go from the cold jar to the hot jar and upset the certainty of entropy. While the idea was refuted, it still provokes debate even today.

One final important caveat to add to the discussion of stochastic logic is that even at this high level of abstraction, the questions that get asked are not value free. Smith (1998) discusses how the Victorian ethic was indeed evident in exactly the kind of thought experiments represented by the Maxwell demon.

Figure 1 details the epistemological context of Bayesian probability as taking place in an infinite time space, allowing for an infinitely large population, and being best suited for study by the thought experiment.

<u>Social Science.</u> As Western society became increasingly complex, the need for an epistemology capable of detecting changes and patterns not apparent to the naked eye became more and more important. The first step was fostered by the need for statistics to regulate quality control in the mass production process and came to serve social science as tool to detect very small differences in very large populations. But as the Industrial Revolution matured into modernism, the need to be able to describe complexity became even more profound. Gergen (1991), for instance, discusses this complexity as a problem at the very root of concepts as basic as our sense of self.

For instance, social science techniques can inform us about the effects of viewing violence on television (Gerbner, Gross, Morgan, & Signorielli, 1985). Questions like this are tested with inferential statistics based on the Central Limit Theorem. Here the null hypothesis, or hypothesis of no difference, is set up as a straw man, and its rejection implies that there is "not nothing" associating the variables that make it up. Hence, Gerbner and his associates began by stating the proposition *viewing television violence makes no difference to how the viewer perceives the world*, and then went show its statistical improbability.

The time problem surfaces at this level of analysis because, while the population of images on television is very large, it is not infinite. This means that the even the most carefully drawn sample still contains a small, but nontrivial, margin for error. This forces social scientists to make a leap of faith. They are forced to draw a finite sample and study it during a finite time span simply as a function of the limited resources available to them. Such decisions are justified on the grounds that the populations they study are so large, or what Dennett (1995) calls *vast*, they can be practically conceptualized as infinite. Chances are that assumption is correct most the time, or at least frequently enough to trust social science as more than alchemy. These methods are indeed sound enough that a close look at the radical critique of empirical science proposed by some critical theorists can be made to look foolish (Sokal, 1996).

Consider the application of social science methods to the problem of murder in our society. Wellford and Cronin (1999) used a field experiment to examine 798 homicides in four cities with different combinations of success rates in solving homicides and other crimes. The study employed the use of inferential statistics, such as linear regression, to show that the city adhering most closely to basic standards of crime investigation solved the most homicides. It is important to note that the study does not provide insight into the course of a particular investigation, but comments only on police departments as aggregated units. The statistics employed generate a standard error that can serve as the basis for inference, or as the ability to predict future events. (In this case, the authors argue that if standard police practices are fully employed, the number of unsolved homicide cases can be dramatically reduced.)

It is, however, important to understand that even though social science is cloaked in empirical methods driven by statistical certainty, the basic question-asking process is still normative. Krueger (2001) points out that “the selection of hypotheses, their number, their location on the continuum of possible hypotheses, and their prior probabilities depend on the researchers’ experience, their theoretical frame of mind, and the state of the field at the time of study” (p. 19). Why, for instance, wasn’t the resolution of cases of petty theft studied? Couldn’t it be argued that such crime would be much more likely to affect the lives of most people, and thus be more relevant to their perception of public safety than homicides, which take place much less frequently?

Figure 1 summarizes the epistemological context of social science as being constrained by time frames of years, within a vast (but not infinite) population space, and employing the methods of hypothesis testing. At a practical level, the process of peer review within the canons of the Academy employed by scholarly journals may take years to complete before publication of results becomes a reality. Of course, this makes the plodding process of social science awkward to the application of human problems that need immediate attention.

<u>Everyday life</u> Solving problems in the real time of everyday life is not a new human enterprise. Our most distant ancestors dealt with the daily ordeal of figuring out what caused the rustling in the bushes fast enough to make survival-contingent decisions to flight or flee. Evolution has endowed humans with their own set of pseudo-statistics called psychological heuristics. These heuristics are shortcuts to help make decisions in complex environments very rapidly based only on limited information. Tversky and Kahneman (1974), for instance, detailed the

representativeness heuristic, where people expect a process such as a coin toss to be represented locally as well a globally. This leads to the gamblers fallacy and the insensitivity to information about predictability. It can also be, however, a conservative strategy that offers the security in classification that overestimates risk. If a person who believes he or she will be more likely to die in a plane crash than by cancer decides to quit flying, that person is much less likely to die in a plane crash – notwithstanding the fact that the chance of death by cancer was really much higher statistically in the first place. This strategy of "it is better to run away, to come back and fight another day" is the direct result of having to make high-risk decision under the pressure of real time. In a sense life is an ordeal of constant struggle to stay in the present, that is, to survive. In a contemporary setting, Newhagen and Reeves (1992) have shown it takes around 1 to 3 seconds to execute memory retrieval tasks for material that people see on television. The weakness of the heuristic approach is that it tends to over classify things as negative (Newhagen, 1998). This stands to reason because the cost of being wrong about any novelty in a benign environment is arguably higher for negative changes than for positive ones. This makes plausible Gerbner's contention that exposure to violence on television gives viewers a sense that the world is a scarier place that it really is.

Returning to the example of homicide: The range of direct human experience is very short in this area, and the population of interest is very small. Here one homicide down the street will likely *cause* neighbors to inflate the actual chances of their falling victim to such a crime simply because of its proximity. This is efficient to the degree that *any* odd shadow in the back yard at midnight can be classified as a possible murderer on the prowl. Even if it is the case that the vast majority of such shadows turn out to be neighbors' pets, it does prompt the resident to slam the double bolt lock shut and call 911, thus reducing the possibility of falling victim to such a violent crime. True, this heuristic may promote neighborhood crime watch programs, but it also may result in the kind of social isolation described by Putman (2000). While it can be argued that the application of ancient psychological heuristics to contemporary life may be non-adaptive, the questions that get asked at this level are neither normative nor arbitrary. However distorted the cognitive or behavioral responses to complex stimuli generated by our contemporary social milieu might be, the force driving decisions at this level is the imperative to survive.

Figure 1 describes the epistemological context of everyday life as taking place from minutes to milliseconds, the population space being very small, and the method of choice to solve problems to be the psychological heuristic.

Pattern Identification The process of computer-assisted reporting in epistemologically interesting because of the level of analysis it represents. On one hand, the patterns under scrutiny are too subtle to be detected by the psychological heuristic. On the other hand, the patterns are not nested within a sufficiently large population to warrant the application of inferential statistics. It stands to reason that humans would have figured out an epistemology designed to deal with midrange problems between the adreneline-driven present and vastness of the nearly infinite future. As reasonable as that idea seems, it is odd that more attention has not been paid to medium-term problems -- those too large to be managed by internal heuristic mechanisms but still small enough to make a real difference in peoples lives. Common problems within this range exist in time spans ranging between months and weeks. Problems that arise at this level might have to do with public policy. Issues about zoning and growth control may be out of the hands of individuals, but they may still have important medium-term implications for an individual's security and livelihood. This range of issues can be the subject of overt advocacy in a democratic society, where different entities take on the task of figuring out what the "best" level of human action is or at least calling attention to the fact that existing systems are not functioning at an optimal or utilitarian level. Journalism may be viewed as one of those entities.

Journalism's proper role in this regard – and whether journalism's place is merely to transmit information or to unearth patterns not visible to others – has certainly been a subject of debate within and outside journalism. Less subject to discussion is a theory of knowledge appropriate to detecting patterns of interest to the public at this level of analysis. Meyer (1973) proposed the application of social science methods to journalism, including computer analysis of survey data, significance testing and inferential statistics. He called this practice "precision journalism." It was seen as a way to improve the collection and interpretation of information, to get beyond the practice of merely transmitting untested information, and to detect important truths that might otherwise be missed, all in service if a systematic attempt to understand matters of relevance to readers. Though precision journalism has gained a following among some investigative journalists, the use of such statistical methods has been adopted by no more than a tiny fraction of reporters (Maier, 2000; DeFleur, 1997). On the other hand other forms of computer analysis in reporting have taken hold in journalism, following the model of demonstrated by Jaspin (1989, 1993) in the late 1980s and early 1990s. Jaspin used computers to analyze large government databases by such basic techniques as searching, sorting and counting, and cross-indexing, or looking for records in one database that also appeared in a second, unrelated database. These databases may suffer from what is sometimes described as a "dirty data" problem (Houston, 1999), ranging from typing errors and inconsistent spellings to the disconnect between the purposes for which a government database was created and the way in which a reporter plans to use it. The cross-indexing or merging of databases may require extensive examination of the underlying records – first to determine how best to perform the match and then to confirm that the links suggested by computer analysis are valid. These practices in the service of pattern identification, often in connection with a newly configured data universe that is at once an enhanced and constrained version of the original, encompass the concept of "deep audit" as used in this paper.

Ettema and Glasser (1998) argue that such investigative reporting is qualitatively unique because the journalist takes an active role in the formulation of the value-laden questions they ask, thereby making "moral judgments." They state that standard beat reporters, for instance, simply transmit "bureaucratically sanctioned accounts" which reinforce conventional definitions of social reality, upholding the normative order (Ettema & Glasser, 1985, p.188). This view depicts the difference between investigative journalism and standard daily reporting as having to do with the way questions are asked, rather than how they are answered. But does this imply that investigative journalism reaches into some value system independent of the established "normative order," or is the threshold for what constitutes a significant outcome set either implicitly or explicitly by law, policy, or custom? Sante (1991) argues that 19th century muckrakers such as Jacob Riis simply looked at extant social standards as the foundation for their campaigns. Sante contends that the standard for housing could be seen in the New York tenement, an architectural urban form originally intended to be a single-family dwelling. Riis' campaign against overcrowding, where each room of a tenement frequently housed an entire family, was simply based on a comparison of reality to the implicit single family standard. It should further be emphasized that from a postmodern viewpoint investigative reporting is not unique in asking value-laden questions, virtually any question asked above the level of the individual is -- at its heart -- normative.

It might be more epistemologically useful to work from how journalists answer their questions rather than how they ask them. The substantive difference between computer-assisted reporting and standard public affairs reporting has as much to do with who is doing the primary analysis of data as it does the questions asked of those data. Standard reporting technique does not allow the time or resources to engage in primary data analysis. It is, then, at the mercy of those that do the analysis with regards to the kinds of questions that get asked. Ettema and Glasser (1985) may be doing a disservice to beat reporters when they depict them as mindless drones simply moving information from public officials to mass publics without taking the time to evaluate the validity of their claims. For instance, Levy (1981) details a number of techniques journalists use to "disdain" reports from officials they find questionable by inserting words such as "so-called" in their stories. What constrains journalists may have more to do with their inability to analyze primary data than it does with their will to challenge the conclusions of those who can.

Investigative journalism distinguishes itself in its ability to look at raw data, and in doing so having more control over the kinds of questions that get asked. Computer-assisted reporting takes a giant step forward in that process in its ability to harness the power of the digital computer to process vast amounts of data under the time constraints imposed by journalism.

In doing so there are two very important features of the process that should be kept in mind. First, while computer-assisted reporting teams may have more flexibility in the kinds of questions they ask than the beat reporter, those questions are grounded in the norms, cannons, and codes of the social milieu in which they work. Second, while the patterns they uncover may not be visible to the naked eye, they are not the product of social science technique. Quite the contrary, the judgment and values the computer-assisted reporting team employs are founded very much in journalist tradition. What distinguishes computer-assisted reporting teams is the way they creatively construct previously unavailable data sets through the use of the digital computer. Such data sets are by definition censuses and do not require the application of inferential statistics or formal hypotheses. Data are simply compared to social or cultural expectations. In this sense the questions computer-assisted reporting teams ask may be normative but their conclusions are not independent of the extant value system in which they live. Thus the journalist-researcher uses public standards as the threshold when he/she decides if a pattern is significantly aberrant to justify publication.

The art of this craft is in "seeing" the pattern has as much to do with the way data sets are constructed as it does with the way they are subsequently analyzed. The term used here to identify this epistemology, the deep audit, suggests an analytical process of gauging actual performance against social standards that goes below the obvious or the available. A classic example of such journalism, which followed a series of deaths and injuries to Rhode Island children in school bus accidents, involved matching the names in a database of school bus drivers to the names in state databases of traffic infractions and criminal cases (Jaspin, 1989, 1993; Jaspin and Johnson, 1986). This work detected a pattern of school bus drivers with problematic driving histories and led to the exposure of loopholes in the system for hiring and monitoring the drivers. Notice that Jaspin did not decide independently that school bus drivers ought to have good driving records. That conclusion is implicit in the privileged status our society gives children and the importance it gives to their safety. It manifests itself in everything from the bright colors buses are painted to the exceptionally strict laws concerning obedience to their flashing stoplights.

Jaspin's advocacy of what he termed "computer-assisted reporting" and his establishment of a training center at the University of Missouri coincided with the increasing popularity and availability of the personal computer, the increasing computerization of government records, and a system he developed for transferring data from government mainframe computers into reporters' desktop machines. For more than a decade, a steady stream of Pulitzer Prizes and other awards have gone to news stories based in reporting that relied, in part, on computer analysis using the techniques of "precision journalism" and "computer-assisted reporting." Computer analysis is practiced in some form at many newspapers, however, within those newsrooms even the more accessible techniques advocated by Jaspin do not appear to be widely diffused among reporters (Maier, 2000).

The challenges in the deep audit process could be enumerated in various ways. One, for example, would be a consideration of the problem of legal access to data. Though laws and court decisions have gradually expanded the definition of public records to include a greater variety of electronic data, some sorts of information are not available to the public and press in electronic form because the law restricts their release in any form. This includes such things as police records in ongoing criminal investigations, applications for welfare, or most income tax returns. Yet another useful way to look at the epistemology of the deep audit, in terms of constraints, is to think about the questions of time and problem space and draw comparisons to what is going on at the levels of everyday life and social science.

The deep audit in journalism typically takes place over a much longer period than the "everyday life" level of analysis and the work typically associated with meeting the ever-present deadlines around which news organizations are typically organized. In addition, the time required to detect and then verify newsworthy patterns cannot be reliably predicted, and the expenditure of such time carries the risk of yielding no story at all. This feature of the deep audit stands in contrast to the more anecdotal forms of journalism centered on tips, news events, or disputes between public figures who are more than happy to tell their side of the story to a reporter. The differential in expenditure of time – including the time needed to learn to use analytical software and haggle with public officials over access to data – is certainly a potential source of friction in the newsroom and a disincentive for many to engage in the deep audit process. Yet the deep audit may be concluded in a time frame shorter than that required for a rigorous and extended research effort in the social sciences, in part because the journalist's work may be confined to a single community and limited to what can be learned from available data. In addition, rather than being limited in research design to testing a hypothesis conceived in advance and then either accepting or rejecting it, as is the custom in certain traditional forms of social science research, the process of computer-assisted analysis leaves the reporter free to drop the initial line of inquiry when it bogs down and head off on a tangential path that seems more likely to reveal something achievable and newsworthy. When successful, the deep audit may result in blockbuster revelations of patterns not detectable or credible at the mere level of anecdote, and such revelations may be followed by prizes and enhanced stature in the community and the publishing world. Senior editors have found computer analysis to be an irresistible capability to have in their newsroom, but few have made an all-out effort to expand the percentage of staff capable of doing that work.

The extent of the problem space for the deep audit is also a useful way to compare this approach to what is going on at levels of analysis above and below it. By making use of data that may be detailed and extensive, the deep audit can operate at a level beyond what is accessible in everyday life. Other aspects of the process, however, act to reduce the problem space below the "vast" level claimed by social science, which endeavors to develop findings applicable to society at large. For example, computer-assisted analysis in journalism, with the exception of polling, tends not to want to apply its findings to anything other than the "sample" studied, and that "sample" is typically seen as 100 percent of the population at interest. The previously mentioned study of school bus drivers looked at all school bus drivers in Rhode Island, not a sample of them, and made no effort to extrapolate its findings either to other states or to school bus drivers in a continuum including the past and the future.

This process of cleaning and merging databases reduces the dimensionality of the data universe in such a way that otherwise hidden patterns emerge. Inferential statistics are not automatically excluded from the deep audit process, and may certainly be used in the background from time to time as one of a number of mechanisms for validating results, but they are not intrinsically part of the process. As in social science methodology, the deep audit does ask the question "compared to what?" But a key difference is that the "watchdog" role is paramount in the deep audit and its focus is often local rather than societal in scope. Also, the deep audit may be exploring new territory where no other relevant data has been gathered by social scientists or others. As a result, the standards of comparison for the deep audit may be more flexible than those for social science. For example, traditional social science seeks to compare results from a sample to the hypothetical picture of a vast population of potential samples. In the deep audit, it may be adequate to compare the results of research to what would be expected from requirements established by law or regulation, or even to make comparisons to what would be expected from prevailing standards of justice, decency, or common sense. The school bus driver story alerted the public and policy-makers to a problem, and even without reference to patterns from other jurisdictions, the need for action was clear. The licensing system and administration were revamped following the story, and the licenses of dozens of drivers were revoked after the state looked into their records (Jaspin, 1993).

Some of the features enumerated here for the deep audit as an epistemology operating in a level of analysis between everyday life and social science can be seen in its application to the issue of homicide investigation. As the 1990s came to an end, reporters and editors at *The Washington Post* became interested in the striking number homicide cases during the decade that had gone unsolved in the District of Columbia – about 1,500 of them.[1] Although the city's murder rate had declined sharply during the 1990s, the police department's success rate in solving homicide cases had declined even faster. A reporting plan was eventually put in place that involved three reporters, a researcher, and an editor, and it included three general lines of inquiry – human sources, examination of paper documents, and computer analysis of homicide and court records. A four-day series resulting from this effort was published at the end of 2000 (Thompson, Chinoy & Vobejda, 2000). One of the original goals of this reporting plan was to understand why the rate of solving homicide cases had declined and to evaluate the claims of some members of the law enforcement community that problems within the police department were contributing to the troublesome state of affairs. Reporting based on documents, human sources, and prior news coverage did suggest problems with the deployment and supervision of homicide detectives, shoddy handling of case files, and a repeated failure to make adequate and lasting improvements in the wake of prior revelations and studies.

Electronic data were collected during the early phases of this project, including a database of homicides from the D.C. police and another database of murder and manslaughter charges from D.C. Superior Court. It soon became apparent, however, that these databases,

[1] Much of the description of *The Washington Post's* series on homicide in Washington D.C. was based on personal conversations with Ira Chinoy, one of the members of the team who worked on the project.

which were primarily designed for tracking homicides or court cases but not for analyzing them, did not contain enough information about each one to be able to detect meaningful correlations between the status of an investigation and the way it was conducted. By comparison, a detailed and thorough social science study released at about the same time (Wellford and Cronin, 1999) examined variables affecting homicide clearance rates in four unnamed cities. In that study, researchers were granted access to nearly 800 homicide case files – including the files of open cases that are normally not accessible by journalists – and extracted 215 variables from each file. The resulting report concluded that a variety of factors related to police performance were associated with case closure rates, including response time, follow-up, and the actions of the first officer on the scene. The databases that *The Washington Post* obtained from police and the courts did not lend themselves to that sort of analysis, and the details of a requisite number open cases would only have been accessible through extraordinarily difficult and time-consuming reporting efforts. But the deep audit process did help reveal other patterns related to the city's problem with unsolved homicides, and the bulk of work was completed in just under a year of steady effort, the sort of time-frame acceptable to a large news operation for a limited number of projects.

Three key patterns emerged from this deep audit approach. One was that there had been an increasing reliance by the D.C. police on something other than an arrest to declare a case closed; these so-called "administrative closures" often involved citing a dead suspect as the most likely perpetrator, and the examination of case files raised questions about the adequacy of the evidence in some of these cases. The second pattern involved repeat killers. In one decade, at least 225 slayings – 1 out of every 10 closed cases – were allegedly committed by perpetrators whom police deemed responsible for a prior killing. The third pattern came to be seen as evidence of a culture of "street justice." At least 150 alleged killers had themselves been killed, and about half of those alleged killers had already gone through the criminal justice system before they were slain. These patterns, not evident or even expected when the reporting process began, emerged through a laborious deep audit process. Separate databases of victims and alleged killers were assembled. The spelling of names was routinized enough to correct for rampant variations and look for matches: alleged killers who appeared in connection with more than one date, alleged killers who also appeared in the list of victims, and homicides that were declared solved without a corresponding murder or manslaughter case being files in D.C. or federal court.

Ironically, the patterns that emerged in this deep audit would probably not have been accessible to a social science study that looked at cases in the aggregate without looking for specific connections between the individuals involved. At the time of *The Washington Post's* reporting project, in fact, a check of leading criminologists specializing in the study of homicide suggested that such patterns had largely escaped notice in the field. On the other hand, though great caution was employed to ensure the veracity of information on which the reporting was based, the problem of limited and initially "dirty" data made it difficult to draw distinct statistical inferences. That, however, did not keep the project from fulfilling the "watchdog" function – alerting the public and policy makers to a serious problem in their midst. What's more, the ameliorative efforts by public officials began even before the series ran. In early 2000, *The Washington Post* requested access, under the city's public records law, to dozens of closed homicide case files maintained by the D.C. police. These included cases identified through the deep audit process as potential "administrative" closures -- cases listed as closed without corresponding murder or manslaughter charges filed in court. Following the newspaper's request, police conducted an internal review of the cases and learned of a number of problems with the way they had been handled, and the police chief announced in a radio interview that his department's process for investigating homicides was "dysfunctional" and that he was going to act to fix it (Thompson & Dvorak, 2000).

Several months later, after the series appeared, the police chief announced additional reforms, members of the public and city officials issued calls for action, and the city council held hearings and issued its own report on the homicide closure problem (Thompson & Pierre, 2000; Patterson, 2001; Ramsey, 2001). Interestingly, after being alerted to the problem through the watchdog efforts of the newspaper, the police and city council both turned to Wellford and his research on homicide closure rates in four cities for guidance in devising reforms that would lead to improved performance on homicide cases in the District of Columbia (Ramsey, 2001; Patterson, 2001).

The allocation of resources and time required for the deep audit, when compared to other reporting practices, are likely factors in journalism's reluctance to do more investigative reporting or to employ computer analysis of data as a standard reporting tool. The perpetual controversy over objectivity – and whether it is compromised when journalism discovers new truths and then takes a de facto advocacy stand in calling for action – may also play into the lack of a wholehearted appetite for the deep audit and what it reveals. The irony here is two-fold. First, even the supposed objective sciences – with their hypotheses tested in a process from which bias has supposedly been removed – operate within a subjective framework when it comes to the choice of subject matter and the questions to be asked. Second, trained journalists using computers to integrate complex databases may be in one of the best positions to comment on how well a community's systems are doing in the medium term. This is simply an extension of the "watchdog" function. Ironically, Mayo J. and Leshner (2001) found that newspaper readers deemed stories based on computer analysis to be no more credible than stories attributed to standard sources and less appealing. That, however, may not be too important to the efficacy of the process because the true discourse may be taking place between journalism and members of political elites, regardless of the fact such scrutiny takes place in the public sphere (Habermas, 1998). Even if such discourse does not enter into the general debate, this does not dilute the fact that it works to benefit the public.

Figure 1 shows that the epistemological context of pattern identification takes from weeks to months, the population space is large (but not vast), and suggests the deep audit as an appropriate methodological tool.

Conclusion

This paper argues for the inclusion of the pattern matching executed by use of the *deep audit* as a discrete and important epistemological response to the kinds of mid-range problems addressed by computer intensive investigative journalism. The discussion of the deep audit has situated it within a continuum of other, more familiar, empirical methods employed to understand causal relationships in the environment. The importance of induction as a way to link effects to causes is that it allows us to make predictions about the future in the hope that it can be improved, or at least better anticipated. The discussion has not dwelled much on the mechanics of the deep audit as methodology, frequently known as computer-assisted reporting. The fact is that overemphasis on the use of computers has likely drawn attention away from the deep audit as a discrete empirical methodology.

The Cost of Error

Thinking about pattern matching within the context of the epistemological systems that bracket it helps emphasis the conceptual component of the process. This paper has attempted to do that in terms of the different levels of analysis of human action and in the context of how time constrains different problem solving methodologies. Finally the implications of making a mistake will be discussed. After all, one of the best ways to understand a process is to see what happens when it doesn't work.

Social science thinks about Type I and Type II error in the interpretation of hypothesis (Babbie, 1989). Type II error takes place when the null hypothesis of no difference is not rejected when it should be. The convention is that the null hypothesis cannot be rejected unless the p value of the inferential statistic employed is greater than or equal to .05. That means the researcher has to be at least 95 percent certain that the results under analysis are systematic. Type II error is generally considered the least damaging because it is an error of omission. Type I error, on the other hand, takes place when the null hypothesis is rejected, but when there is no real relationship between the variables of interest. This is held to be especially dangerous because it leads the researcher to say something is going on when it is not. While the discussion of Type I and Type II errors are usually confined to the interpretation of inferential statistics, the concepts on which they are based can just as easily be applied to the other epistemological systems described here.

Bayesian Probability This is the domain of stochastic probability, where theory is proposed and tested in the abstraction. That, however, does not mean things cannot go wrong, theoreticians make mistakes too. The implications of error at this level are probably less dire to the workings of society than mistakes at other levels. Figure 2 shows that a Type I error, such assuming that the set of all arithmetic theorems is finite, leads to bad theory. Such errors can be detected and corrected through logic-driven proof, which is what Kurt Gödel did when he showed there is an infinite set of such proofs (Rose, 1997). Type II error, on the other hand, results in theoretical ambiguity, or not being able to demonstrate the proof for a theorem. The impact of Gödel's proof was profound because it demonstrates that regardless of how many times attempts to demonstrate a proof fail, because the set of possible proofs is infinite, the chance it can be validated always exists. Mathematicians find this prospect particularly unsettling.

<u>Social Science</u> The consequences of a Type I error to social science can be profound. Figure 3 shows such an error results in bad science, such as the 19th century theory of Phrenology. Phrenology came from the theories of the idiosyncratic Viennese physician Franz Joseph Gall (van Wyhe, 2002). One of basic tenets of Gall's system was that the human skull takes its shape from the brain, and thus could be read as an accurate index of psychological aptitudes and tendencies. Implications of Gall's system included, among other things, a scientific justification for false conclusions about the relationship between race and intelligence. Figure 2 shows how phrenology was used to map the mind based on the shape and size of the skull.

Type II error in social science can lead to dogma. Kuhn (1970) discusses the process of scientific progress as revolution rather than evolution. The period between the time when the dominant paradigm in a field runs out of gas and a replacement becomes recognized could be thought of as Type II error. The example frequently used is the period between the exhaustion of "flat world theory," and the Copernican model of a round world.

<u>Pattern Identification</u> Figure 2 shows the implications of a Type I error to the pattern identification process generated by the deep audit can be inaccurate journalism. A news organization would be guilty of a Type I error if an investigative series claimed that an

automobile's poorly designed gas tank could explode in a crash, when it did not. On the other hand, if a news organization fails to aggressively monitor the performance of government agencies, and overlooks a practice such as racial profiling by police, it commits a Type II error, opening the door to benign neglect.

Everyday Life Figure 2 shows that making a Type I error in everyday life can result in death or suffering. While many routine decisions might escape such dire consequences, a pedestrian only has to misjudge the distance of an oncoming car once to wind up in the hospital. Type II errors in everyday life are not really unusual and represent the kind of conservative over classification of risk discussed earlier.

The key to understanding empirical ways of knowing is not to pit one against the other in search for a "best method." Rather, the point is that questions important to human life, such as the occurrence of violent crime can be asked at *all* levels of analysis and employ the full range of epistemological methods they suggest. If such an approach is employed the same question will not only yield information to the level of analysis where it is asked, but inform us at other adjacent levels as well. This become particularly obvious when the consequences of mistakes are considered across levels.

Reference List

Babbie, E. (1989). The Practice of Social Research. Belmont, CA: Wadsworth Publishing.

Beniger, J. R. (1986). The control revolution: Technological and economic origins of the information society. Cambridge, MA: Harvard University Press.

Blalock, H. M. Jr. (1979). Social Statistics: Revised second edition. New York: McGraw-Hill Company.

Cohen, J. (1994). The earth is round ($p<.05$). American Psychologist, 49(12), 997-1003.

DeFleur, Margaret H. (1997). Computer-assisted Investigative Reporting: Development and Methodology. Mahwah, N.J.: Lawrence Erlbaum Associates.

Dennett, D. C. (1995). Darwin's dangerous idea. New York: Simon and Schuster.

Derrida, J. (1983). Literary Theory: An introduction. Minneapolis, MI: University of Minnesota Press.

Ekman, P. (1992). Facial expression of emotion: New findings, new questions. Psychological Science, 3(1), 34-38.

Ettema, J. E., & Glasser, T. L. (1985). On the epistemology of investigative journalism. Communication, 8 , 183-206.

Ettema, J. S., & Glasser, T. L. (1998). Custodian of conscience: Investigative journalism and public virtue. New York: Columbia University Press.

Geiger, S., & Newhagen, J. E. (1993). Revealing the black box: An information processing approach to understanding media effects. Journal of Communication, 43(4), 42-50.

Gerbner, G., Gross, L., Morgan, M., & Signorielli, N. (1985). Living with television: The dynamics of the cultivation process. J. Bryant, & D. Zillmann (Eds.), Perspectives on media effects (pp. 17-40). Englewood Cliffs, NJ: Lawrence Erlbaum Associates.

Gergen, K. J. (1991). The Saturated Self. New York: Basic Books.

Giddens, A. (1984). The constitution of Society: Outline of the theory of stucturation. Berkeley, CA: University of California Press.

Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed. American Psychologist, 47(6), 766-779.

Habermas, J. (1998). The structural transformation of the public sphere: An inquiry into a category of bourgeois society. Cambridge, MA: The MIT Press.

Houston, Brant (1999). Computer-assisted Reporting: A Practical Guide. 2nd ed. Boston: Bedford/St. Martin's.

Jaspin, Elliot (1989). Computer = Reporting Tool. The Computer Connection, 18-24. Syracuse: S.I. Newhouse School of Public Communications, Syracuse University.

Jaspin, Elliot (1993). The New Investigative Journalism: Exploring Public Records by Computer. Demystifying Media Technology: Readings From the Freedom Forum Center Editors John V. Pavlik and Everette E. Dennis, 142-49. Mountain View, Calif.: Mayfield Publishing Company.

Krueger, J. (2001). Null hypothesis significance testing. American Psychologist, 56(1), 16-26.

Kuhn, T. (1970). The structure of scientific revolutions. Chicago, IL: The University of Chicago Press.

Lang, A. (2000). The limited capacity model of mediated message processing. Journal of Communication, 50(1), 46-70.

Levy, M. (1981). Disdaining the news. Journal of Communication , 31, 24-31.

Mayo J., & Leshner, G. (2001). Analytical journalism: Credibility of computer-assisted reporting. Newspaper Research Journal, 21(4), 68-82.

Maier, Scott R. (2000). Digital Diffusion in Newsrooms: The Uneven Advance of Computer-assisted Reporting. Newspaper Research Journal 21(2), 95-110.

Meyer, P. (1973). Precision journalism. Bloomington, IN: Indiana University Press.

Meyer, P. (1991). The new precision journalism. Bloomington, IN: Indiana University Press.

Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

Newhagen, J. E. (1998). TV images that induce anger, fear, and disgust: Effects on appoach-avoidance responses and memory. Journal of Broadcasting and Electronic Media, 42(2), 265-276.

Newhagen, J. E., & Reeves, B. (1992). This evening's bad news: Effects of compelling negative television news images on memory. Journal of Communication, 42(2), 25-41.

Patterson, Kathy, Chairperson (2001). Oversight Report on the Metropolitan Police Department's Homicide Investigative Practices and Case Closure Rate. Committee on the Judiciary, Council of the District of Columbia. February 27.

Putman, R. D. (2000). Bowling alone: The collapse and revival of American Community. New York: Simon and Schuster.

Ramsey, Charles H. (2001). Testimony. Public Oversight Roundtable on Metropolitan Police Department Homicide Closure Rate. Committee on the Judiciary, Council of the District of Columbia. January 25.

Rose, J. N. (1997). Maxwell's demon and the Turing machine: Their processes/dynamics are identical, making them archetypal guides in evaluating the fundamental characteristics of "consciousness." Presented to A.S.S.C. Conference "Demon, Machine & Mind." Claremont College, Pomona, CA. (June 13-16, 1997). [http://www.ceptualinstitute.com/uiu_plus/turingdemon.htm] Accessed January, 2002.

Sante, L. (1991). Low life: Lures and snares of old New York. New York: Vintage Books.

Smith, C. (1998). The science of energy: A cultural history of energy physics in Victorian Britain. Chicago, IL: The University of Chicago Press.

Sokal, A. D. (1996). Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity. Social Text, 14(1-2), 217-252.

Thompson, Cheryl W., Ira Chinoy, and Barbara Vobejda (2000). Fatal Flaws: The District's Homicide Crisis. The Washington Post, 3-6 December.

Thompson, Cheryl W., and Petula Dvorak (2000). D.C. Homicide Cases to Be Reviewed; Some Files Incomplete, Some Investigations Closed Prematurely, Ramsey Says. The Washington Post, 7 July.

Thompson, Cheryl W., and Robert Pierre (2000). New Team to Review District Homicides; Unclear Case Closures Surprise Ramsey. The Washington Post, 7 December.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Van Wyhe, J. (2002). *The History of Phrenology on the* W*eb* http://pages.britishlibrary.net/phrenology/. [Accessed March 30, 2202].

Wellford, C., & Cronin, J. (1999). An analysis of variables affecting the clearance of homicides: A multistate study. Justice Research and Statistics Association.

Abstract

Time Constraint	Epistemological Context	Population Space	Method
Infinite	Bayesian Probability	Infinite	Thought Experiments
Years	Social Science	Vast	Hypothesis Testing
Months-Weeks	Pattern Identification	Large	Deep Audit
Minutes-Milliseconds	Everyday Life	Small	Psychological Heuristics

Concrete

Figure 1

Research Method Context and Consequences within Error Type

Error Type

Method Context

	Type I	Type II
Bayesian Probability	Bad Theory	Ambiguity
Social Science	Bad Science	Dogma
Pattern Identification	Bad Journalism	Benign Neglect
Everyday Life	Death and Suffering	Missed Opportunity

Figure 2

A 19th Century Representation of the Mind Based on the Theory of Phrenology

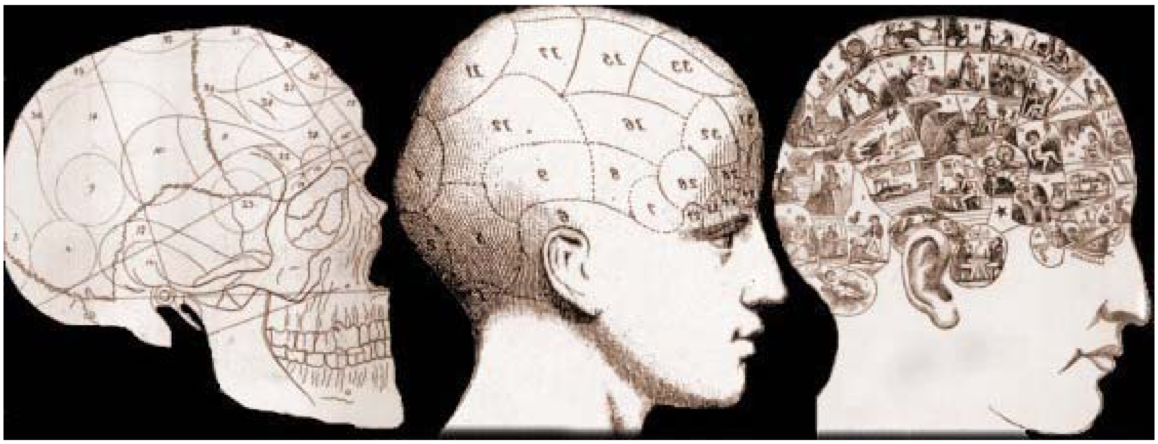

Figure 3