Routinizing the Discovery of Secrets

Computers as Informants

GARY T. MARX
NANCY REICHMAN
Massachusetts Institute of Technology

The king has note of all they intend
By interception which they dream not of.
—William Shakespeare, *The Life of Henry V*

- A computer cross-check resulted in the investigation of a California woman suspected of bilking the welfare department out of more than $4,000,000. Using a variety of aliases over a seven-year period she successfully filed fraudulent assistance claims for 38 nonexistent children.
- The Commerce Department, concerned over illegal exports, has distributed a list of 12 "red flag" signals that may suggest an illegal transfer of goods. A 24-hour-a-day telephone hotline has been established. Persons working in high-technology industries are encouraged to report any suspicions.
- The FBI and IBM jointly run a fake consulting firm in the Silicon Valley in San Jose, California. The sting operation involves selling IBM trade secrets to Hitachi and Mitsubishi.

These diverse examples are typical of recent efforts to solve a traditional problem faced by any enforcement agency: the need to locate infractions.

Police in the United States traditionally have relied heavily on unsolicited information from citizens to direct their efforts (Black, 1980, Reiss, 1971.[1] In a democratic society there is much to be said for this means of mobilization. It can offer a degree of citizen control over police discretion. This, along with other limitations on the autonomy of police to initiate investigations, is surely a necessary feature of liberty.

The traditional citizen-reporting approach may work well where there are clear victims or observers who are aware that infractions have occurred and who are willing to report what they know. It is less

AMERICAN BEHAVIORAL SCIENTIST, Vol. 27 No. 4, March/April 423-452

effective when those with information are intimidated or otherwise not forthcoming. When witnesses are not even present, when there is no clear individual victim, when the offense is hidden or highly technical, or where a well-organized conspiracy is present the traditional approach is irrelevant.

Reliance on citizens for information can have two major drawbacks: (1) the ratio of offenses citizens choose to report, relative to those they actually know about, may be too low or may be systematically biased in an undesirable direction; (2) there are many offenses of which citizens are unaware. These drawbacks have become more apparent in recent decades. In response, an important area of criminal justice reform has sought to improve the ability of social control agents to discover offenses and offenders systematically.

REFORMS INTENDED TO IMPROVE THE DISCOVERY PROCESS

Systematizing or routinizing discovery has taken two broad forms. One form responds to the problem of underreporting. It seeks to structure the environment so that citizens will be more likely to come forward with information. Toll-free hotlines where citizens may anonymously call in tips, televised police appeals for information, neighborhood crime watches, and citizen patrols seek to make reporting easier and more accessible and to increase the flow of information to police.[2] Protections for those who report have also been enhanced.[3]

The second form of enhancing information discovery involves police taking initiatives to discover infractions on their own, without being dependent on what citizens may choose to report. Undercover work is an example. Police increasingly have sought to discover crimes by becoming a party to them, whether as fellow conspirators, observers, or victims (Marx, 1982). Another form of police initiative we have chosen to call "systematic data searching." As illustrated by the discovery of the California woman who fraudulently received welfare aid for 38 non-existent children, systematic data searching involves gleaning data, usually in computerized form, for direct or indirect evidence of infractions.

While it would be worthwhile to devote equivalent attention to each attempt at enhancing the discovery process, we have chosen,

instead, to use this limited space to explore systematic data searching in greater detail [4] We do this because of its relative newness, its rapid expansion, and its having received little research attention. While considerable attention has been devoted to the vast new crime opportunities computers offer (Parker, 1976; Whiteside, 1978), less attention has been given to the role of computers in discovering crimes.

Systematic data searching involves more than just the application of computer technology to existing law enforcement process.[5] It is in some ways a new tool. It permits the joining of heretofore independent pieces of information in order to expose offenses and offenders that would remain hidden unless such links could be drawn. Systematic data searches do not merely expedite existing discovery processes. They offer an entirely new means of exposing rule breaking. They offer a "value-added" or inductive method that differs from traditional, deductive methods. Rather than drawing inferences from a "crime scene" that has natural, seemingly self-evident boundaries, systematic data searching permits investigators to construct criminal scenarios from disparate data and events. They may also permit a form of statistical surveillance.

This article draws on 8 interviews with specialists in computer detection and over 100 interviews carried out in the course of our research on undercover tactics and insurance fraud investigations. Information from these interviews is not presented quantitatively, nor is it used to test hypotheses. Systematic research is premature until issues have been framed and questions raised. It is hoped that our discussion can contribute to the type of systematic research required to answer the questions to be suggested.

A MORE DETAILED LOOK AT SYSTEMATIC DATA SEARCHING

Systematic data searching has been facilitated by new computer developments. These developments have occurred concurrently with the increased prominence and attention given to what can be called "low-visibility" offenses. Much white-collar crime, such as price fixing, corruption, and trade violations, can be so characterized. The significant expansion of benefits provided by the modern welfare state has also generated new opportunities for fraud. The implications of this for exploitation have rarely been noted.[6]

Factors that inhibit the discovery of such offenses go far beyond the physical barriers and the right to privacy noted in the literature as factors that limit the discovery of offenses by routine patrols of public areas (Stinchcombe, 1963; Mawby, 1981). The impersonal and routine settings in which these offenses occur and the very large numbers of potential offenses/offenders means that control agents usually cannot rely on prior reputation as a means of suspicion, as they can with more traditional offenses.

Many crimes by or against organizations are deceptively masked as legitimate organizational transactions. Applying for and receiving welfare benefits, for example, is legal unless the fact of employment is concealed. Similarly, filing a property insurance claim is legitimate unless there was no loss. Since the infractions occur in the context of many similar, legitimate transactions, they do not stand out immediately as instances of wrongdoing. Organization members and routine organizational process also may shield illegal action from exposure.[7]

In such cases the legitimate and routine appearance of the violations is in sharp contrast to predatory crimes (such as robbery, assault, or rape) or even victimless crimes (narcotics, prostitution) where the apparent act is illegal and traces of the activity (the injured victim, the smashed window) are instantly obvious if seen. No similar "on-site" clues alert social control agents that low-visibility offenses have occurred. There is no "smoking gun."

Beyond their entrenchment in routine organizational process, low-visibility offenses often are difficult to discover because they occur over time and information about them is dispersed across institutional settings. The discovery of low-visibility violations that occur over time, or across agencies or cases, is enhanced by the pooling of information. Death records are a good example. Although they have major bearing on many federal entitlement programs, death records are maintained locally. Historically, there has been no systematized way for federal agencies to obtain these records automatically to confirm program eligibility. In addition, technical advances such as automatic check writing and depositing may further mask discovery. The system grinds along on its own initial momentum, absent an order to decease.

Systematic data searches appear well suited for the exposure of these types of low-visibility offenses. In their simplest form searches may be applied to a single body of data. Before computerization, records such as applications were checked for internal consistency,

errors, and missing information. But this was often done superficially, with little cross-checking and in an inconsistent and nonsystematic fashion. The individual clerk or auditor usually had vast discretion over whether or not, and what, to check.

With computerization screening can become routinized, broadened, and deepened. Computers permit forms of investigation that previously were impractical. In contrast to traditional techniques that could assess static demographic data, computers permit analysis of more complex transactional data, such as number of visits to a doctor, phone calls to particular individuals, travel patterns, bank deposits, and the timing and interrelations of events (Burnham, 1983). A much more textured or dimensional picture is possible.

An internal computer data search may reveal discrepancies, contradictions, and irregularities that would be missed by a clerk reviewing the form. Equity may be increased as all forms are checked, not just those that happen to catch the fancy of an auditor. The IRS, for example, now is able to screen the over 90 million tax returns it receives for missing information and mathematical errors. Cross-referencing distinct data bases (as with social security numbers and death records) may expand and qualitatively change the nature of the search. Data analysis may yield profiles of likely offenders. Patterns of offense may be discovered through aggregation not possible if one follows a Sherlock Holmes logic of deduction and looks at only a few cases. Indicators may be created that suggest that a violation is likely. The investigator may then follow or track these cases over time.

Two increasingly prominent types of computerized data searching are *matching* and *profiling*. These certainly do not exhaust all forms of searching, but they are among the most important.[8] While they may overlap or appear sequentially, they are analytically distinct and offer one way of organizing the empirical material.

MATCHING

Matching involves the comparison of information from two or more distinct data sources. It may be used for cross-checking and verification or to discover inconsistencies and multiple listings suggesting violations. According to one estimate, approximately 500 computer matching programs are being carried out routinely at the state and federal levels (U.S. Senate, 1982: 20).

Among the most dramatic examples of the violations matching may discover are impersonation and false representation. For exam-

ple, a cross-check of social security rolls and medicare records resulted in the arrest of 29 people for cashing checks made out to dead friends and relatives. One woman had been forging the name of a deceased friend for 14 years. Officials reported uncovering losses of over $30 million (New York Times, May 20, 1983).[9] In what a prosecutor called "the most concerted effort yet not simply to respond to complaints but to affirmatively go out and detect fraud," the U.S. Office of Education has used computer searches to flag suspicious applications in federal student loan programs. The rate at which fraud has been uncovered as a result has more than tripled (Boston Globe, June 27, 1983).

Third parties may exploit what once was a valid claim. For example, matching black lung program payments with social security records revealed that the program was continuing to provide compensation to 1200 individuals listed as deceased (U.S. Department of Health and Human Services, 1981: 24).

A second type of violation commonly discovered is "double dipping." A person may be legitimately entitled to the benefit in question, but, through seeking the same benefit in different jurisdictions, or using different names, or (where payment legitimately terminates) reapplying after an extended period of time, he or she may fraudulently obtain additional benefits. For example, a match of the welfare rolls of 34 jurisdictions involving 5 million records turned up 3500 cases where persons appeared to be receiving public assistance in more than one state (U.S. Department of Health and Human Services, 1981: 30). Some welfare systems will automatically cross-check birth records whenever a person claims to have twins, since false claims regarding twins are a well-known means of seeking increased benefits (New York Times, August 3, 1982).

Computer matching may also be used to discover false claims that would render an applicant ineligible for the benefit in question. For example, in Massachusetts computer matching has been used to find welfare recipients with bank deposits in excess of the amount permitted. The welfare department supplied banks with the names and social security numbers of all welfare recipients. Matching these numbers with their customer information, the bank officials gave the state a list of welfare recipients holding cash assets in their banks. The inquiry discovered over 1600 instances in which assets in excess of the $5000 limit appear to have been held (U.S. Senate, 1982: 240).

The fraudulent claim may involve an event rather than some aspect of a person's biography. A common form of insurance fraud

involves purchasing the title certificate for a wrecked car sold as salvage. The car is insured and subsequently reported as stolen. Theft insurance would then be collected on a nonexistent car. However, with computer matching this has become more difficult to do. The National Auto Theft Bureau now maintains records of all vehicles sold as salvage and/or reported stolen.[10] By marrying theft reports with salvage records, the computer matching program permits instantaneous discovery of a type of fraud that previously lay hidden in two rarely connected bodies of data.

Matching may be used to identify persons who fail to meet an obligation. For example, in an effort to discover income tax evasion, particularly by the self-employed, the IRS is testing a system that matches tax records to estimates of income based on the type of neighborhood an individual lives in and the type of car he or she drives. The data are to be purchased from private marketing firms that sell computerized lists to direct-mail companies. The IRS is also matching data from county recorders of deeds with tax returns, to find individuals who fail to pay capital gains taxes owed from the sale of real estate (New York Times, August 29, 1983).

Matching can also be used in a preventive way, for example, linking the failure to meet an obligation with a new request. In rules announced by the Office of Management and Budget in 1983, federal agencies are now prohibited from making loans, procurements, contracts, or major grants until they have prescreened applicants through credit bureau inquiries to be sure that they are not delinquent in repaying prior government loans and other overdue obligations (New York Times, September 24, 1983).

PROFILING

Matching may be used to construct profiles of violations or violators. But the logic of profiling is more indirect than that of matching. It follows an inductive logic in seeking clues that will increase the probability of discovering infractions relative to random searches.

Profiling permits investigators to correlate a number of distinct data items in order to assess how close a person or event comes to a predetermined characterization or model of infraction. The modal characteristics and behavior patterns of known violations or violators are determined relative to the characteristics of others presumed to be nonviolators.[11] Indicators of possible violations are developed from

this comparison. Where the behavior is complex and evolves, a model may be developed of the interrelations among the relevant factors. But most common is a simple laundry list of "red flag" characteristics. As more and more of these occur the case in question becomes more suspect. A second, more in-depth, investigation is then carried out to determine if a case that has been flagged as suspicious actually involves the violation.[12]

Profiling is indirect because the indicators used are not in themselves indicative of illegality. However, their joint appearance is thought to be associated with an increased probability that a violation will occur or has occurred. Profiling may be *singular* or *aggregative*. The former consists of a model of distinct attributes. The latter consists of the reappearance of factors that, appearing only once, in and of themselves would not trigger suspicion. Their appearance across cases, such as a single person's being the owner of several inner-city buildings that burn down, would lead to further investigation.

Let us consider singular profiling first. It focuses on discrete characteristics or events. There is nothing illegal or exceptional about being a male, purchasing a one-way airline ticket, paying for it with cash, and obtaining the ticket at the last minute at the airport. But analysis suggests that when these factors occur together, the chances of a skyjacking attempt are increased. The same thing applies to a drug courier profile used to stop suspicious persons at airports.

The IRS was an early user of profiles in efforts to identify tax violators. Persons claiming deductions beyond a certain percentage of their income and certain configurations of deductions are likely to trigger more detailed inquiry. One way to get on the IRS's "tax gap hit list" appears to be to purchase audit insurance (Wall Street Journal, June 29, 1983). The logic here is that people who purchase audit insurance are likely to have something to hide and are gambling that it's cheaper to purchase the insurance than to pay the tax.

Profiles also can be used in a preventive way. The development of arson early warning detection systems in Seattle, Boston, New Haven, and other cities illustrates this (National Legislative Conference on Arson, 1982). Computer-based arson prediction models are used to identify buildings thought to be at risk of being burned. This opens up the possibility that preventive action will be taken. In another form of prevention, the profile may result in interdiction before the act can be fully carried out. Airline skyjacking profiles are one example, for

instance, refusing to issue tickets to passengers matching the profile may prevent the skyjacking (Time, July 26, 1976). Interrogations and searches resulting from drug courier profiles are another example.

Profiles developed for identifying welfare fraud can be used to prevent ineligible cases from entering the system. For example, in Sacramento County (California) a profile for identifying suspicous cases has been developed around the number and age of children, health care, and school records. This model is based on an assumption of at least occasional childhood illness and treatment. If a recipient claims children and there are no school records and no medical claims for the children, further investigation results (U.S. Senate, 1982).

Profiles of auto theft and bodily injury fraud increasingly are used in insurance cases. Profiles are based on factors that often accompany fraud, such as losses occurring close to the inception date of a policy or claimants avoiding the U.S. mail in correspondence regarding the claim. A series of questions, a checklist of responses, and associated point system have been developed that allow adjusters quantitatively to rate the degree to which a particular claim is consistent with ideal fraud types (Reichman, 1983).

The Educational Testing Service uses profiling to help in the discovery of cheating. In 1982 the service sent out about 2000 form letters alleging "copying" to takers of its scholastic apptitude tests. The letters note that a statistical review "found close agreement of your answers with those on another answer sheet from the same test center. Such agreement is unusual and suggests that copying occurred." Students are told that in two weeks their scores will be cancelled and colleges notified, unless they provide "additional information" to prove they had not cheated. An important factor in the sending of such letters is the "K-index" which compares incorrect answers among suspect test-takers. (N.Y.T. July 2, 1983)

another form of profiling, aggregation

is based not on the distinctive charcteristics of any one case, but on the frequency with which certain factors appear across cases. The profile emerges from the aggregation of similar incidents or configurations. There is an implicit threshold. Once this is reached, red flags appear. Aggregative profiling often is directed against systematic and repetitive violations rather than the one-time violation.

Such profiling has been used extensively in efforts to find insurance fraud. For example, the State of Florida's Division of Insurance Fraud maintains an index of all bodily injury insurance claims. The index is used to ferret out violations that cut across seemingly unrelated claims. Thus when the same doctor-lawyer combination reappears on a significant number of personal injury claims, investigators have reason to look further for a fake accident ring. This pooling of information may give the analyst reason for suspicion that would not appear to an insurance company office paying a single claim.

Similar logic underlies the Property Insurance Loss Registry (PILR), a not-for-profit discovery organization sponsored by the insurance industry. Among other information, it records the location of fires, insurees, mortgagees, and contractors. A current fire prompts a

search through the PILR index for other similar fires involving the same persons or organizations. While the discovery of other fires is not directly discrediting, it suggests that further inquiry into the fire loss is appropriate.

Profiling is also used in some parts of the private sector to identify drug users. For example, one drug consultant goes through computerized company personnel records looking for employees under 35 who show higher-than-average rates of (1) absenteeism, (2) requests for early dismissal or time off, (3) lateness, (4) sick leave, (5) accidents, and (6) Worker's Compensation claims. An employee showing sufficient elements of this profile may be asked to undergo a blood or urine test to determine the presence of drugs (Newsweek, August 22, 1983).

USES OF THE RESULTS

In the data analyst's language, the results of an initial computer search are referred to as "raw hits." Depending on search type, these include indications of direct infractions or a sufficient number of red flags alerting agents to possible violations. A name on both the welfare and city employment rolls, the repetition of an event or characteristic beyond some identifiable threshold (such as four consumer complaints against the same company), or a person or event that matches a profile associated with previous violations are illustrative. These raw hits include the total universe of hits. This universe in turn is made up of "solid hits," "misses," and "inconclusives."

"Solid" or "true" hits are instances in which conclusive evidence of violation is found.[13] But what happens when additional investigation yields conclusions that negate the initial finding of a hit? In most

cases what appeared to be hits will simply be considered misses and it will be possible to explain away the initial suspicion. Misses appear as a result of errors, situational factors that lead to a different interpretation of the facts, or, in the case of profiling, a necessary casualty of probabilistic reasoning.[14] In other cases, while sufficient evidence of infraction is not available, neither is the conclusion of a miss. No evidence is found to cast doubt on the original reasons for suspicion, and evidence to strengthen it may even have been found. The term "inconclusive" is appropriate here. Where there is reason to think that a violation will eventually appear, one response is to monitor or track a case over time.[15]

The goals of a data search may change with its repeated use. When a system is first applied to an existing data base, its goal is likely to be the discovery of current or past offenses. It may seek to "weed out" bad apples. It searches for illegitimate cases. For example, recipients of the black lung benefits are provided with payments for children up to the age of eighteen. When the U.S. Department of Health and Human Services screened its records, it found 3000 offspring whose ages exceeded the eligibility standard, though not all of these were continuing to receive payments (U.S. Department of Health and Human Services, 1981: 25). The statistical technique of discriminant analysis is used by the Farmer's Home Administration to identify problem loans. Based on patterns identified in previous cases of default and foreclosure, the technique permits investigators to screen out current loans exhibiting those characteristics associated with a high probability of default (President's Council, 1983).

Once a data base has been purged initially of such cases, however, the goal may shift to deterrence and prevention. In fact, preventing fraud and abuse before they occur is the new objective of the President's Council on Integrity and Efficiency (PCIE), established in March 1981 to promote and coordinate the activities of inspectors general, many of whom pioneered the use of computer matching. Program administrators hope that the publicity about data searching will deter potential offenders.[16] Public relations efforts may seek to create the impression that the computer's awesome power is all knowing. This may build upon the mystique surrounding technology in general and computers in particular. Fear and trembling may be engendered among the naive, as they impute unrealistic powers to the computer. There is a parallel to the unwarranted power some persons impute to the lie detector. This is reminiscent of President

Nixon's immortalized words on the Watergate tapes, "Listen, I don't know anything about polygraphs, and I don't know how accurate they are, but I do know that they'll scare the hell out of people."

Where such deterrence is not present, applying the search before people are officially entered on the rolls or, in the case of the black lung example above, assuming that they are removed at the appropriate time may anticipate violations and allow for preventive measures. In a private sector example, major credit card companies may soon be confirming the personal identity of credit card holders through signature verification technology. A technique has been developed for analyzing the pressure and direction of a signature as it is being signed. This could then be compared to data stored from previous signatures (Wall Street Journal, June 9, 1983).

SOME POLICY AND RESEARCH ISSUES

> I hope you do not assume yourselves infallibilitie of judgment when the most learned of the apostles confesseth that he knew but in parts and saw but darkly through a glass.
>
> — Sir Richard Saltonstall

It is clear that data searching techniques such as matching and profiling can significantly enhance discovery. As we noted earlier, systematic data searching seems particularly well suited to ferreting out certain low-visibility offenses that involve organizational processing. As with undercover sting operations, their dramatic results make for good media treatment. These techniques generally have been positively received. Their use is expanding rapidly. But, as with any means, they have a cost. The lunch is never free, whatever other attractions it may have. Two of the most important costs are the consequences of error and the implications for civil liberties.

ERRORS

Important factors in the assessment of data searching are the cause, frequency, and consequences of various types of error. At least five sources of error can be identified: (1) erroneously reported or incorrectly entered data, (2) time lags, (3) computer hardware and

software problems, (4) the acontextual nature of the decision process, and (5) the probabilistic nature of profiling.

The extent of erroneously reported, or incorrectly entered, data will vary greatly across programs and data types. We know little about its frequency. A study of the social security numbers of over 2 million food stamp and AFDC recipients found 5100 instances in which nonissued numbers were in use. Approximately one-third of these cases were a result of data input errors — the numbers were transposed by the applicant or by program officials (U.S. Senate, 1982: 5). In the first computer run of the Massachusetts bank records match, 24% of the social security numbers used in the matches were incorrect (U.S. Senate, 1982: 224). A procedure adopted later, which coupled the first letter of the surname with the social security number, helped reduce errors based on incorrect matches to 7%. Although this is a significant reduction in the error rate, the ease and magnitude of such errors gives on pause.

The process used to create the data base must be seen to reflect human judgment and not be seen as a perfect reflection of reality. It must be approached tentatively. Were the data gathered under coercion or periods of great stress? Are data collectors and processors aware of proper data collection procedures and motivated to follow them?[17] Do program staff have incentives for falsifying data? If matters of judgment are involved, how high is reliability across judges? Even when the agency that initially gathers the data discovers an error, the ease of access to computerized information on the part of other agencies may limit its ability to control the flow of erroneous information. The automatic interfacing of computer systems may mean that the original processors of the data are unaware of the ultimate users and uses of such information.

The time lag between events, the reporting of events, and input into computerized data banks and analysis offers another source of error. For example, in New York State a match of work records with a list of persons receiving assistance in the last quarter of 1978 revealed that 10% of welfare recipients were actually working. A second review disclosed that at least half these persons were on both lists legitimately. Some recipients had been on welfare during the beginning of the quarter and only subsequently found work. Because the data were not updated in a timely fashion, some innocent individuals were initially suspect (Boston Globe, July 23, 1979).

Computer hardware problems may lead to data errors. Among problems that can be caused by faulty hardware is the "doubling up of

records" so that the value of a variable is recorded twice. This can wreak havoc with quantitative eligibility requirements such as a minimum amount in the bank, age, or number of children. Such hardware problems are easy to correct technically once they are located. But this requires vigilance in looking for errors and the incentive to make corrections. In the interim, persons may experience loss of benefits or receive benefits to which they are not entitled.

Another not uncommon technical problem lies with software errors. In using large data bases formatting errors can easily occur. If a command has been formatted incorrectly, the wrong variable will be pulled out for analysis. For example, when applicants provide income data for several years, a formatting error could abstract a previous year's income for current income.[18]

The error sources considered thus far are largely technical. With sufficient experience, resources, cross-checks, updating, and incentives, they can probably be reduced to an acceptable minimum. But this may not be the case with errors that are related to substituting technical for human judgment and profiles based on samples for which the true parameters are unknown. The most serious questions raised by systematic data searching lie here.

When a machine recommends a decision, the recommendation is only as good as the data and programs that have gone into it. One measure of goodness has been considered above—whether the data are erroneous in some technical sense. But a more subtle meaning involves completeness and sensitivity to unique parameters. When used as a decisive guide, rather than as an aid, systematic data searching is misused. The machine should not be a substitute for human discretion and judgment.

Errors in interpretation may arise because of the acontextual nature of the data analysis. Only a fraction of reality's richness is abstracted out and put into machine-analyzable form. There is a bias toward general features characterizing many cases, rather than the atypical, idiosyncratic, or extenuating circumstance.

As we move from the formal and general categories used to develop aggregate patterns basic to the actuarial method, to inferences about particular persons in specific situations, problems may appear. An example of this can be seen in the case of a nursing home resident who lost her Medicaid eligibility as a result of the Massachusetts bank matching program described above. The data that resulted in her being dropped were technically correct as far as the

search program was concerned. Yet it was a wrong decision. The woman's bank account included a certificate of deposit held in trust for a local funeral director to be used for her funeral expenses. Although federal regulations exempt burial contracts from asset calculation, the trust was included in the determination of her assets and she was excluded from the program (U.S. Senate, 1982: 106-107).

In another case a Washington, D.C., welfare recipient obtained a job at the Department of Health, Education and Welfare. Although she properly notified the welfare department of her changed status, word never reached those responsible for mailing the checks. The checks kept coming despite her repeated attempts to inform the welfare department of her new status. She eventually cashed the checks to pay off doctor bills incurred as a result of her serious illness. Subsequently, she was indicted on a felony charge and her name (along with 15 others) was listed in local newspaper stories describing the results of HEW's computer matching of its own employee records. Many of the others indicted also had informed the welfare department that they were currently working. When the judge learned the details, a majority of the cases, including that of the woman described above, were dismissed or reduced to misdmeanors. Yet the damage to these people's reputations and six months of uncertainty before their cases came to trial cannot be undone.[19]

A final source of error inheres in the very idea of profiling. It stems from statistical reasoning and group comparisons. With aggregative profiling some hits composed of repetitive events will appear as a result of chance. For example, sometimes persons showing roughly equivalent error patterns at a test will represent random factors rather than cheating. Some persons may simply have the bad luck to have a series of fires on properties they own without arson as the cause.

The data base used for constructing a profile may be reasonably accurate as far as it goes, but may simply not be representative of the larger universe of events. Important data may never enter the system. Thus it is sometimes argued that our knowledge of criminals is distorted because it is based primarily on those who get caught and they may be less competent than those who manage to avoid apprehension.

When data gathering on controversial and confidential topics is separated from data analysis, users may not be in a position to know much about the representativeness of the data they are given. Prose-

cutors, for example, usually have no choice but to accept the selectively reported information police bring them on gambling (Reuter and Rubenstein, 1978).

Even in the unlikely event that a profile was to be developed that described characteristics of all true violators, it would also likely characterize many nonviolators. In the case of skyjackings, for example, a majority of skyjackers may fit the profile, but so too do a large number of nonskyjackers. Given the extreme rarity of skyjackings per airline passenger there are no doubt many more misses than true hits. This also may be true for airport drug courier profiles that include such criteria as arriving from a city noted as a drug source, casual dress, scanning the concourse, making a telephone call on arrival, and appearing nervous (U.S. v. Harrison, 1982). While the profile does turn up solid hits, it may also cause much embarrassment and inconvenience to those wrongly interrogated. Procedures for taking reparative action, to the extent that this is possible, are clearly appropriate.

Whatever the source, errors will occur. In considering their costs it is useful to separate errors involving false accusations from those involving the failure to identify violations. The common distinction used in the analysis of statistical data between Type 1 and Type 2 errors can be usefully applied here. Type 1 errors involve identifying an infraction when in fact none exists. Type 2 errors involve failing to recognize an infraction when one does exist.

Type 1 errors involve false accusations. Like the dolphins who are inadvertently trapped in nets put out for tuna, innocent persons are caught in the net thrown out for offenders. Loss of benefits, defamation of character, alienation, and a more general delegitimation can result from such errors. In the case of false accusation, the state has a moral, and often a legal, obligation to provide a means of review. Although Type 1 errors have an individualized impact, they may incur high societal costs as they challenge democratic ideals of fair process.

Type 2 errors reflect an inefficient discovery mechanism (that is, not netting the universe of offenders). Their consequences vary according to whether one seeks to discover infractions that have already occurred or those that are planned. Not identifying a direct violation (for example, that a person is obtaining public assistance while working) may be inefficient, but it does not produce a clear direct cost since the behavior would have remained hidden whether or not a weak search process were in place. On the other hand, as the case of arson or skyjacking suggests, when the goal is prevention, the failure to

recognize a set of behaviors or events as consistent with a profile of wrongdoing can have more serious consequences.

Type 1 errors almost always become manifest because the investigation reveals a miss or a falsely accused person protests. But whether or not Type 2 errors are identified varies across offense types. Such errors are likely to be discovered only if a victim reports the offense or if it of necessity becomes public. For example, skyjacking offers a great contrast to drug smuggling. With a profile in place every skyjacking attempt represents a Type 2 error. But completed drug smuggling violations are far more difficult to identify. The extent of Type 2 errors involving the former can be checked continuously, but with drugs this is almost impossible. Where profiles can be checked they are subject to more frequent revision and, presumably, improvement. Where the size of Type 2 errors cannot be determined the profile remains a captive of its assumptions, which must remain unvalidated. The IRS, with the power to carry out in-depth investigations of random sample of taxpayers, illustrates one method of assessing the extent of Type 2 errors that would not otherwise be visible.

The assessment of errors also must consider the rate of error relative to the rate of true hits. If you increase the capacity to get true hits, do you proportionately increase the rate of errors or does the error rate grow exponentially? Or are there instances in which they might even be inversely linked?

In his novel *Nineteen Eighty-Four*, Orwell imagined a social control system that was both highly efficient and repressive. Perfect control over information was the key element (whether the ability to discover infraction or to manage beliefs). While not explicitly mentioned, computer technology was implied. Our review certainly does not question the repressive potential of such technology. But the sources of error we have noted clearly call into question limits on the efficiency and accuracy of computer control technology and illustrate the high cost of mistakes.

CIVIL LIBERTIES

Computer data searching involves the same civil liberties issues raised by the use of computer files in general.[20] Visions of the central all-knowing computer and Kafkaesque nightmares lurk on the horizon. Important concerns are privacy, Fourth and Fifth Amendment protections, and due process of law.

Critics argue that these searches are more intrusive than other forms because those subject to them are likely to be unaware that any search is going on. They may have given direct or willing consent for neither the search nor the disclosure of personal information to others. In cases in which consent has been given, this may be a result of duress and coercion rather than a real choice, since one may believe that one may have to forgo a badly needed benefit if one does not give consent.[21]

Privacy may also be violated by the improper disclosure of data to third parties without the consent of the subject. Or the data may be improperly obtained by them. The sharing of data across agencies heightens the risk of unrestricted or improper access to confidential information. Even without such exchanges, the fact that security around these kinds of data sets is generally weak invites abuse.[22]

The use of computerized records for purposes unrelated to their initial collection has also been questioned. At the federal level such use is prohibited normally by privacy legislation. However, the Privacy Act of 1974 exempts computer matching programs when they are classified as "routine use" procedures, meaning when they are used for purposes compatible with the reasons for which the data were collected originally.[23] Broad interpretations of "compatible purpose" have made it possible to include nearly any government-initiated venture. The "routine use" classification can thus be used to circumvent protections against invasions of privacy the legislation was designed to prevent.

The programs may be questioned on Fourth and Fifth Amendment grounds. Searches can be viewed as "fishing expeditions," absent any substantial evidence of wrongdoing by the person in question. As such, they may be seen to violate the Fourth Amendment's protection against unreasonable searches and seizures. When data voluntarily given for one purpose are used for another, a person's right to protection against self-incrimination may be violated.

To the extent that one is not provided with proper notice that an individual is subject to a search, timely notice that one is a "hit," and an opportunity to contest the results of a search, due process questions also emerge.[24]

In contrast to conventional criminal accusations, data searching may transform the presumption of innocence into an assumption of guilt. It can lead to imperious behavior as an agency cuts off benefits or cancels test scores without even a hearing. Accusations become

equivalent to convictions without a trial. The burden of proof may be on the target of the hit to show that the violation did not occur, rather than on the agency to show that it did. Officials may abdicate responsibility for their accusations to computer programs or models. In such cases suspects effectively relinquish their rights to face their accusers, at least directly.[25] Even then, challenges may be possible only after punitive action has been taken on, and publicity generated with with respect to, the presumed guilt.

Supporters, however, argue that a balance must be struck between the rights of the individual and the needs of the state, and do not view matching programs as undue intrusions. Properly conducted computer searches are seen to be less intrusive than other forms of search, such as rummaging through a person's bank records. Data searches abstract specific variables from records, with total disregard of other variables. In contrast, an individual searcher can scan entire records picking and choosing among items. Furthermore, consent for computer searches is often given, or implied when one voluntarily provides the data. Advocates claim that with proper guidelines and administration, problems are minimal.[26]

Thus far most of the debate between opponents and supporters has reflected competing value positions. It also has been at a very general level and has not made distinctions between types of search or error. Disagreements are now based primarily on value positions, with neither side able to examine adequately the empirical premises that bear upon the arguments. Given the absence of adequate data on most of the issues in question, it could hardly be otherwise. We have only minimal data on the extent of falsely accused people and the ratio of hits to misses for various kinds of searches. Little is publicly known about the validity of different profiles. Data on the frequency of the concerns raised by civil libertarians (or the counterclaims regarding the effectiveness of guidelines offered by supporters) are also missing. Nor do we have studies showing whether the discovery benefits continue over time or become neutralized with regular use.[27]

We do not have the detailed case studies of the actual operation of matching and profiling programs that are requisite for sound policy recommendations. There has been little discussion of how risks can best be minimized and errors corrected or of how competing values should be weighed. How do matching and profiling differ from each other with respect to the costs of error? What are the relative costs of Type 1 (false accusations) and Type 2 errors (failing to identify an

infraction)? Should there be a presumption against using such techniques, or certain forms of them just as there is with the use of weapons or Fourth Amendment searches, except under special circumstances and when no other practical means are available? How does systematic data searching compare to other means of obtaining information on low-visibility offenses such as undercover tactics and efforts to increase citizen reporting?[28]

As in so many other areas of contemporary life, rapid technological development has outpaced the establishment of ethical and legal standards for their use. The important Federal Privacy Act of 1974 does not address many of the issues raised by recent computer developments. Less than one-fifth of the states have laws requiring written standards for the collection, maintenance, and dissemination of person information, though this number is growing. Of course, as time passes and problems are identified the quality of computer use in the areas considered above will no doubt improve. But this is likely to be offset by problems associated with the continuing expansion of computers to new untested areas.

SOME THEORETICAL IMPLICATIONS

The significance of systematic data searching goes beyond the public policy questions considered above. It also has implications for understanding society and the nature of social control. The use of computers as informants is but a small part of a broad social process of rationalization.

The recent growth of matching and profiling is part of a more general process of rationalization that began in the nineteenth century. The same broad social forces affecting the economy touch criminal justice (Spitzer, 1979). In a rational effort to control the environment, policy has become more systematic and routinized. Social control has sought greater effectiveness, efficiency, certainty, and predictability.

Rather than having to rely on what citizens happen to report or police accidentally discover, control agents are taking greater initiative. This may bring greater equity as police seek independence from the biases a citizen-based reporting system may entail. With a skeptical and scientific ethos and a broad data base that can be inexpensively screened, it becomes prudent to consider everyone a possible suspect initially. Analysis rather than tradition becomes the basis for action.

Eliminating the traditional temporal distinction between locating an offense and searching for an offender may yield greater efficiency. Some systematic data searches collapse these processes as offense and offender are discovered simultaneously.

Yet, just as Mark Twain observed that claims of his death were greatly exaggerated, so too many claims about the efficacy of a rationalized criminal justice system be overoptimistic. In the case of systematic data searching, for example, if it does not contain within it the seeds of its own destruction, it at least contains an ironic vulnerability to its own neutralization (Marx, 1981). In any setting of strategic conflict, efforts at systematization (unless kept secret) can be exploited by skilled adversaries.[29]

The certainty such techniques seem to offer may be illusive. Their advantages may be temporary or may result in a skewed population of apprehended offenders. Routinizing discovery procedures usually involves focusing attention on a limited number of indicators. These may be invested with far more predictive power than they warrant. Focusing attention on specific indicators implicitly diverts attention from other indicators and can result in tunnel vision.[30] The indicators chosen can easily come to be treated in a ritualized way. Enforcement agents may be held accountable for following correct procedures, rather than for the results of following those procedures. Only superficial concern may be given to whether or not indicators are valid or have been obtained or presented properly.

While deterring or discovering some offenders, routinization can offer an almost guaranteed means of unauthorized access to others, who gain knowledge of the system and take actions to neutralize it. Altheide (1975) has illustrated how security operations designed to restrict territorial access also can serve as a means for facilitating unofficial entry. The same holds for access to the benefits that systematic data searching is designed to control.

By learning what prompts a hit or a red flag, knowledgeable violators may take steps to avoid them. Some variables used in matching and profiling can be manipulated or avoided easily. For example, the well-publicized match of welfare and bank records in Massachusetts no doubt led some persons to hide money in banks outside the state, to entrust it to others, or to convert assets to a different form.

A different type of neutralization lies in the use of false names and identification numbers. Basic to some contemporary matching is discovering the same name, identification number, address, and the

like on lists that should be mutually exclusive. This can be avoided through the use of false identification.[31] The name, identification, or record presented may be valid but may simply not belong to the person presenting it. A record check may attest to the validity of the record, but it is unlikely to discover that it does not legitimately belong to the person presenting it.

Publication of the characteristics used to profile arsonists or skyjackers may offer such persons a way to avoid detection. The likelihood of the discovery of an arson pattern through the Property Insurance Loss Registry described above is reduced if each property is in a different and unrelated name. In response to five skyjackings to Cuba in a two-month period, the Federal Aviation Administration is considering changing its behavior profile (New York Times, July 7, 1983).

Awareness of this neutralization potential raises questions about who is likely to get caught in a routinized discovery system. Clearly, not all potential offenders can acquire the knowledge, or have the skill, sufficient to neutralize the system. However, over time it seems likely that these systems will disproportionately net the marginal, amateur, occasional, or opportunistic violator, rather than those who are more systematic, repetitive, skilled, or professional in their rule breaking. The latter ironically may be granted a kind of license to steal, even while headlines hail the effectiveness of control agents using new techniques.[32] To be sure, where costly violation of the public trust or serious crimes are involved, any apprehension may be desirable. But the routinization of discovery does raise a type of equity issue rarely heard. The question is not the familiar one of how authorities use their discretion in deciding what laws to enforce or who to go after, but, given the means they use, what kinds of cases they are likely to discover.[33]

Beyond questions of equity, efficiency, and the cyclic and dynamic nature of rule enforcement and violation, there is a broader question about the reach of social control. Observers such as Foucault (1977) view an irreversible continuing historical process of more intensive and extensive social control. The capacity of the modern state to gather information and to punish is seen to extend ever deeper into the social fabric. Control is based on "observation, surveillance, and inspection" rather than primarily on physical coercion. Conformity is

thought to emerge out of fear of a pervasive and omnipresent panoptic eye. The net has widened and the mesh thinned (Cohen, 1979). While computer matching and profiling may seem to be relatively pale and benign variants of this, variants they are.

How far do we want those in authority to go in their power to discover infraction? In a time of strong citizen concern over crime and the increased prevalence of low-visibility offenses, there is a great deal to be said for enhancing this ability. The proportion of offenses discovered by police relative to those reported by citizens is increasing.

Yet there is another side as well. A different version of the equity problem may appear when there is a gap between the knowledge of violation and the ability to sanction. While ignorance is not bliss, there is a certain wisdom to the inability of the three monkeys to see evil when action cannot be taken with respect to it. Powerful new discovery means may overload the system. Authorities may discover far more violations than they can prosecute or process. This overabundance can lead to the misuse of discretion and demoralization. Charges of corruption and favoritism may appear and the system may be perceived as unfair.

If this were all that was at stake, awareness of the potential problems and well-conceived policy for structuring choices might suffice. But there is a more onimous side. Paradoxically, *both* repression and equal law enforcement may be inhibited when authorities lack information. As Selznick (1948: 84) observes:

> Do we need or want agencies of control so efficient and so impartial that every actual offense has an equal chance of being known and processed? . . . I am concerned that we do not respond too eagerly and too well to the apparent need for more effective mechanisms of social control. In the administration of justice, if anywhere, we need to guard human values and forestall the creation of mindless machines for handling cases according to set routines. Here vigilance consists in careful study of actual operations so that we may know what will be lost or gained.

Systematic data searching, along with the new citizen reporting programs, undercover police practices, electronic surveillance, and other technical means, offers compelling and little understood arenas for such study.

NOTES

1. We are defining "police" very broadly. By "police" we refer to those charged with the policing function, regardless of what the formal title is. All persons who enforce rules must confront issues around the discovery of their violation.

2. Such programs do generate information. For example, a Baltimore call-in radio program, "Report a Pusher," led to 91 arrests on drug charges. During the 4-hour program, police appealed to citizens for information on drug trafficking. Off the air, detectives took calls and recorded names, license numbers, and other information about persons callers suspected of being involved in narcotics transactions (New York Times, November 7, 1982). In Michigan, $1000 is offered for information leading to the arrest and conviction of arsonists. From the inception of this reward system in 1975 to 1981, 26 payments were made (Arson News, 1981). What is not usually considered is how much of the information provided would have been forthcoming even in the absence of such programs.

3. Thus federal and in many places state legislation and judicial decisions have offered new protections for whistle-blowers. The Federal Witness Protection Program provides relocation and a new identity to informants (see Montanino, this issue). Legislation has also introduced negative sanctions for *not* reporting things such as child abuse and certain hazardous working or environmental conditions.

4. The methods are not mutually exclusive. For example, a lead generated by a hotline or a computer search may lead to an undercover operation. Computers, of course, are part of a broader family of rapidly developing technological means, including electronic surveillance and forensic science, also used to enhance discovery.

5. For example, it contrasts with a New York City program called "CATCH" (computer-assisted terminal criminal hunt) designed to streamline the identification of suspects. CATCH is a computerized "mug book" permitting quick retrieval of names of suspects who fit the description fed into the system. Computers have simply improved upon a traditional tactic (Computerworld, April 7, 1980).

In focusing on the discovery of offenses we are also referring to something beyond merely checking a second data base to find a person's address (such as the Selective Service's use of IRS data to locate people suspected of failing to register for the draft) or using that data base for sanctioning purposes, as with the state's garnishment of income tax refunds due to fathers who default on child support payments.

6. In 1959 entitlement programs accounted for 15% of the federal budget; in 1970 such expenditures had increased to one-third of the $62 billion budget; by 1981 they were $300 billion—almost half the budget.

7. See Katz (1978), Vaughan (1980), and Altheide and Johnson (1980) for discussions of the way task differentiation and bureaucratic organization can shield deviance and neutralize control.

8. Matching across data bases, which is one of our concerns here, shares much with the more traditional and common searching of a single data base. At an abstract level the correlation of distinct information involves the same logic of inquiry. But the former raises questions of privacy and data compatibility (which may have implications for errors and misinterpretations) not found when a single data source

belonging to the agency in question is used. Profiling, the second technique we consider, may draw upon single or multiple data sources.

While similar privacy issues are raised, matching is also distinct from simply looking at another agency's data for cases. For example, Skolnick and Woodworth (1967) have noted how police in Westville located cases of statutory rape from the files of other public agencies. In a British example, Mawby (1981) reports on police identifying drug users by monitoring hospital emergency room activities for drug overdoses.

9. It is well to note that all accounts of the dramatic success of such programs have come from advocates who carried them out. Whether an external audit and a careful figuring of costs and benefits would yield equivalent support is another matter. For example, the New York Civil Liberties Union (1982) argues that the unreported costs of New York State's wage-reporting system, a match of public assistance, unemployment records, and reported earnings, may add up to three or four times those officially stated, while savings may be far less than assumed.

10. This is a not-for-profit clearinghouse supported by insurance companies to provide information and assistance to the insurance industry and law enforcement.

11. Of course the profile is only as good as this assumption. Some in this group are undiscovered violators, though designers of profiles usually assume that this constitutes a small proportion.

12. Depending on whether the data offer direct or only indirect evidence of violation, matching may also trigger a more in-depth investigation. But the more in-depth investigation is always found with profiling.

13. The discovery of infractions, of course, is only the first stage in the enforcement process. How the information is used, and whether it is even used at all, are distinct questions that we will not consider here. Among actions that may result from discovery are prosecution, restitution, denial of a claim or benefit, public relations, blackmail or bribery, and entering into some form of exchange relationship with the violator, such as turning the person into an informer or witness. An overabundance of cases and disinterest, or bias on the part of the enforcement agent, may result in no action being taken. Or, in Silbey and Bittner's (1982) term, the "reservoir of unenforced law" may be directed toward enforcement ends far from those intended by drafters of the original legislation.

14. Raw hits are less meaningful for profiling than for matching on average. Since profiles are based on statistical reasoning rather than the often binary and mutually exclusive categories (at least with respect to an agency's rules) of matching, far fewer solid hits are to be expected.

Efforts to make insurance rates and benefits "gender blind" involve some equivalent issues. While perhaps rational and fair in the aggregate, for any given case the prediction on which they are based can be wrong and unfair. The size of the comparative standard deviations can permit some estimate of the frequency of this.

Similar issues are raised by proposals to base sentencing on "career criminal profiles." A controversial Rand study (Greenwood, 1982), for example, proposed that courts use a profile of the career criminal in deciding the length of sentencing for convicted criminals. A person is presumed to be a high risk for a career in crime if he or she shows at least four of seven variables (for example, in jail for more than half of the preceding two years, previous conviction for the same crime, a record before the age of 16, or unemployed for more than half the preceding two years).

15. Interesting civil liberties and policy questions are raised about the intensiveness and length of time of such monitoring. The monitoring of a targeted person because of an inconclusive search can be separated from the routine monitoring that may occur when computers are part of the system being searched/monitored, rather than merely an instrument of the search. Discovery may be built into the work process. For example, an economic forecaster was arrested after it was discovered that he illegally tapped into a Federal Reserve computer in an attempt to obtain secret information about money supply. The computer recognized the tapping. The man was identified through a trace on his phone line (New York Times, January 5, 1983).

Social security field offices use a specialized "intelligence terminal" that records the author of all computer entries. This is used to monitor the work performance of data entry clerks, and can also be used as an audit trail (Wall Street Journal, July 7, 1982).

The completed input of records and the time they take to process can be logged, as can things such as the number of keystrokes for a given worker. In Massachusetts Blue Cross/Blue Shield claims offices a computer keeps track of worker productivity. Wages are adjusted every two months to reflect the output of data clerks (Kuttner, 1983).

The monitoring of a targeted person is also separate from the use of "computer software time bombs" that may automatically go off when a particular data configuration appears. For example, where personal biography intersects organizational rules in a predictable way, computers can be programmed to respond to changes in a person's status that affect eligibility for a benefit. Changes in age are a clear example.

16. For example, former Inspector General of the Department of Agriculture Thomas McBride, who was instrumental in establishing federal matching programs, reports that the publicity generated about a food stamp matching program resulted in a number of persons asking to be dropped from the program (U.S. Senate, 1982: 20). Whether all of these persons were ineligible or would have been discovered from the match is another question.

17. Criminal records, for example, offer an area where data quality leaves much to be desired. Laudon's (forthcoming) analysis of the FBI's automated criminal history file found that 54% of the records disseminated had data quality problems.

18. In a slightly different context, computer program errors may lead to erroneous medical diagnoses. The General Accounting Office reported that improperly programmed medical instruments have led to wrong diagnoses and at least one death (New York Times, August 22, 1983).

19. The failure to cut off a check once a recipient has reported a change in status represents a type of government-sponsored random integrity test of citizens (though this is not intended). This shows some parallel to indiscriminately applied undercover temptations. In both cases, according to the letter of the law, persons may be guilty technically. Money after all was taken, even if it was thrust upon the "guilty" party. But it is not clear that any broad social purpose is served by offering very attractive temptations to persons who may be weak and vulnerable, absent indications of prior wrongdoing on their part.

Advances in banking technology may unintentionally make it morally and technically easier for such fraud to occur. For example, Louise Van Vooren died in 1976. The government continued to send her social security checks directly to her bank for

automatic deposit until 1981. During that time her daughter drew on the money that was regularly deposited in her deceased mother's account. This seems to involve a lesser degree of moral turpitude than cases where the deceased's signature is forged directly on the social security check. As with the above welfare cases, should such unwitting government encouragement in a violation be treated the same way as more autonomous violations?

20. See, for example, Westin and Baker (1972), Rule et al. (1980), and Perrolle (1983), for treatments of privacy and computers.

21. For example, in 1983 a federal judge in the District of Columbia ruled that a form mailed to 4 million social security recipients "makes a mockery of the consent requirement." The crippled, blind, and disabled recipients of supplemental security income were led to believe that their assistance might be denied if they refused to authorize access to their otherwise confidential tax returns.

22. For example, U.S. Department of Health and Human Services auditors found that the Social Security Administration's system for transferring large volumes of data between centralized computers and local offices could be improperly accessed rather easily. This was also the case for access to Social Security Administration terminals (U.S. Department of Health and Human Services, 1981: 11).

23. For a discussion of the politics of conferring "routine use" status, see Kircher (1981).

24. For example, Office of Management and Budget guidelines for federal matching programs require that information concerning "routine use" matches be published in the *Federal Register* in reasonable proximity to their implementation. Technically, those subject to data searching are given notice in this way. However such publication requirements may have little meaning, since those subject to data searching are unlikely to read the *Federal Register.* Furthermore, "the reasonable proximity" requirement does not assure publication before the search is implemented. For example, a match conducted on federal student loans in August 1982 was not published in the *Federal Register* until December 1982 (U.S. Senate, 1982: 182).

25. What is often the *fait accompli* and incomprehensible and hidden nature of the process for determining guilt may show some parallel to the use of witches and trial by ordeal and other magical means for determining guilt.

26. For example, see U.S. Senate (1982: 4-40).

27. One difficulty in assessing impact is whether or not the rates of infraction stay the same. For example, 1981-1983 saw a significant increase in the use of systematic data searching and a concomitant rise in the discovery of fraud. But it is difficult to know how much of this is due to better discovery and how much to a worsened economy that may have resulted in increased rates of fraud.

28. Undercover means, for example, are expensive, restricted in scope, intrusive, and may "discover" crimes that would not have occurred were it not for the instigative activity of the investigation. Yet they can make discoveries not possible with other means. The investigator can exercise considerable initiative over the process. In contrast, efforts to increase citizen reporting are still relatively passive and dependent on whether or not, and with what, citizens choose to come forward. Undercover means are inexpensive and can cover a broad range of persons and areas. Anonymous means such as hotlines can encourage responsible as well as irresponsible accusations. Systematic data searching can be broad in scope and

relatively inexpensive, and can avoid problems such as generating crime or maliciously inspired accusations, but, as noted, it has other costs.

29. Of course, keeping it a secret may work against the goal of deterrence. An implicit choice may be made between minimizing neutralization and maximizing deterrence. One solution is to hint at the powerful means of discovery being used without being specific. But leaks and the experience of apprehended persons work against this.

30. Lipsky (1980: 122), for example, finds that the routinization of bureaucratic functions reduces the chance to discover unique circumstances requiring flexible responses. The problem is compounded when a computer rather than a human agent is involved. Reliance on the computer (or any other machine) as a surrogate for human decision making may permit violations that deviate from the average to go undetected.

31. For example, in using a social security number other than one's own the unsophisticated person may simply make up a number and run the risk of being detected because he or she has chosen a number that was never issued. But sophisticated offenders will simply take a genuine number belonging to someone else and use that. Their chance of being discovered via a match of claimed to real social security numbers is slight. On the frequency and ease with which false identification is used, see the Report of the Federal Advisory Committee (1976).

32. This depends on the relative distribution of offender types. There is likely to be significant variation across offenses.

33. Beyond pushing toward discovery of a particular type of offender within an offense category, the computer may subtly influence the type of offenses to which police devote their energies. For example, a former chief of the Kansas City Police Department believes that computerization has led to an undue focus on minor offenses (unregistered cars, parking scofflaws) that can be dealt with very efficiently at the expense of other more important and difficult to solve crime problems (cited in Goldman 1983). The effectiveness of the means becomes an important, and often barely recognized, factor in deciding what ends will be pursued.

REFERENCES

ALTHEIDE, D. (1975) "The irony of security." Urban Life 4: 179-195.

——— and J. JOHNSON (1980) Bureaucratic Propaganda. Boston: Allyn & Bacon.

Arson News (1981) "Arson control has most successful year." January.

BLACK, D. (1980) The Manners and Customs of Police. New York: Academic.

BURNHAM, D. (1983) The Rise of the Computer State. New York: Random House.

COHEN, S. (1979) "The punitive city: notes on the dispersal of social control." Contemporary Crises 3, 4: 339-363.

FOUCAULT, M. (1977) Discipline and Punish: The Birth of the Prison. New York: Pantheon.

GOLDMAN, D. (1983) "The electronic Rorschach." Psychology Today (February): 36-43.

GREENWOOD, P. (1982) Selective Incapacitation. Santa Monica, CA: Rand Corporation.

KATZ, J. (1978) "Concerted ignorance: the social construction of cover-up." Urban Life 8: 295-316.

KIRCHER, J. (1981) "A history of computer matching in federal government programs." Computerworld (December 14).

KUTTNER, B. (1983) "The declining middle." Atlantic (July): 60-72.

LAUDON, K.C. (1983) "Data quality and due process in large record systems: criminal record systems." Communications of the Assn. for Computing Machinery.

LIPSKY, M. (1980) Street Level Bureaucrats. New York: Russell Sage.

MARX, G. T. (1983) "Notes on the discovery, collection and assessment of hidden and dirty data," in J. Kitsuse and J. Schneider (eds.) Studies in the Sociology of Social Problems. Norwood, NJ: Ablex.

——— (1982) "Who really gets stung: some questions regarding the new police undercover work." Crime and Delinquency 28: 165-193.

——— (1981) "Ironies of social control: authorities as contributors to deviance through escalation, nonenforcement and covert facilitation." Social Problems 28: 221-246.

MAWBY, R.I. (1981) "Overcoming the barriers of privacy." Criminology 18: 501-521.

National Legislative Conference on Arson (1982) Anti-Arson Manual. Columbus, OH: Author.

New York Civil Liberties Union (1982) An Evaluation of New York State's Wage Reporting System: The Real Cost of Computer Matching. New York: Author.

PARKER, D. (1976) Crime by Computer. New York: Scribner.

PERROLLE, J. (1983) "Computer generated social problems." Presented at the meeting of the Society for the Study of Social Problems, Detroit.

President's Council on Integrity and Efficiency (1983) A Summary Report of Inspectors General's Activities. Washington, DC: Government Printing Office.

REICHMAN, N. (1983) "Ferreting out fraud: the manufacture and control of fraudulent insurance claims." Ph.D. dissertation, Massachusetts Institute of Technology.

REISS, A. (1971) The Police and the Public. New Haven, CT: Yale Univ. Press.

Report of the Federal Advisory Committee on False Identification (1976) The Criminal Use of False Identification. Washington, DC: Government Printing Office.

REUTER, J. and J. RUBENSTEIN (1978) "Fact, fancy and organized crime." Public Interest: 45-67.

RULE, J., D. McADAM, L. STEARNS, and D. UGLOW (1980) The Politics of Privacy. New York: New American Library.

SELZNICK, P. (1948) "Foundations of the theory of organization." Amer. Soc. Rev. 13.

SILBEY, S. and E. BITTNER (1982) "The availability of law." Law and Policy Q. 4: 399-434.

SKOLNICK, J. and J. WOODWORTH (1967) "Bureaucracy, information and social control: a study of a morals detail," pp. 99-136 in D. Bordua (ed.) The Police. New York: John Wiley.

SPITZER, S. (1979) "The rationalization of crime control in capitalist society." Contemporary Crises 2, 3: 187-206.

STINCHCOMBE, A. (1963) "Institutions of privacy in the determination of police administrative practice." Amer. J. of Sociology 69: 150-160.

Technology Review (1983) "When computers track criminals." April: 75-76.

U.S. Department of Health and Human Services, Office of the Inspector General (1981) Annual Report. Washington, DC: Government Printing Office.

U.S. General Accounting Office (1982) IRS Can Do More to Identify Tax Return Processing Problems and Reduce Processing Costs. Washington, DC: Government Printing Office.

U.S. Senate, Committee on Governmental Affairs, Subcommittee on Oversight of Government Management (1982) Oversight of Computer Matching to Detect Fraud and Mismanagement in Government Programs. Washington, DC: Government Printing Office.

U.S. v. Harrison (1982) 667 F2d 1158, Fourth Circuit Court of Appeals.

VAUGHAN, D. (1980) "Crime between organizations: implications for victimology," pp.77-97 in G. Geis and E. Stotland (eds.) White Collar Crime: Theory and Research. Beverly Hills, CA: Sage.

WESTIN, A. and M. BAKER (1972) Databanks in a Free Society: Computers, Record-Keeping and Privacy. New York: Quadrangle.

WHITESIDE, T. (1978) Computer Capers. New York: New American Library.