CHAPTER 12

Systematic Social Observation in Criminology

STEPHEN D. MASTROFSKI, ROGER B. PARKS, AND JOHN D. MCCLUSKEY

INTRODUCTION

Systematic social observation (SSO) came to criminology at the hand of Albert J. Reiss, Jr., who, in the 1960s, encouraged social scientists to shed some "nonsensical" views about the limits and benefits of different forms of observing social phenomena (Reiss 1968, 1971b). Reiss objected to the notion that direct observation of social phenomena in their natural setting was work for solo researchers using qualitative methods, while survey research was suitable as a group enterprise with many researchers using a systematized protocol to gather quantified data. Reiss argued that both direct social observation and survey research were in fact forms of observation that must confront the same set of challenges to produce interpretable information, that both were amenable to either solo or group practice, and that both could be used effectively for discovery or validation of propositions about social phenomena. Beyond these insights, Reiss's important contribution to criminology was the development and practice of the techniques of SSO. Acknowledging that others before him had associated social field observation with the sorts of systematic protocols that had become popular in survey research, Reiss demonstrated how SSO could be used to answer important questions about what influences police–citizen interactions, with implications for theories about police–citizen relationships and for public policies concerning justice, race relations, and crime control. Since Reiss, criminologists have expanded the application of SSO more broadly, but it is still used relatively infrequently.

In this chapter, we describe SSO, noting what distinguishes it from other forms of observation, as well as shared features. We review methodological issues and then turn to practical considerations in fielding SSO studies. We provide a brief description of SSO's contributions to the field of criminology and suggest future directions and uses for SSO.

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_12,

DEFINITION OF SYSTEMATIC SOCIAL OBSERVATION

> An observational method is *the selection, provocation, recording, and encoding of that set of behaviors and settings concerning organisms "in situ" which is consistent with empirical aims* (Weick 1968:360).

When the set of behaviors and settings of empirical interest are those of humans, we call this *social observation*. Social observation becomes *systematic* when

> . . . observation and recording are done according to explicit procedures which permit replication and . . . rules are followed which permit the use of the logic of scientific inference. The means of observation and recording, whether a person or some form of technology must be independent of that which is observed and the effects of observing and measuring must be measurable (Reiss 1971b:4)

SSO, so defined and employed, is a powerful tool for the study of human behaviors, especially human interactions, as they occur.

Of course, SSO is not the only means for observing human behaviors; certainly, it is not the most commonly employed in criminology. In institutionally structured settings, it is common to rely on the traces of behaviors captured in transactions recorded by institutional data systems. In more open settings, it is common to rely on research subjects for accounts of their past behaviors, often in structured interviews. In these methods, researchers rely on retrospective observation at second hand, accepting the observation and recording procedures employed by institutions, and hypothecating the procedures employed by their untrained subjects as observers. The power of SSO derives from its direct connection to the behaviors of empirical interest.

ORIGINS AND APPLICATIONS

In the late 1920s and early 1930s, scholars studying early childhood social development practiced systematic social observation, using methods ". . . designed . . . to ensure consistent recordings of the same events by different observers . . ." (Arrington 1943:83). Arrington's discussion of "the feasibility of systematic observation of social behavior under life conditions" (1943:89), the technology employed, and the reliability and validity of measures obtained by observation remains relevant today. Later reviews of SSO research and methods by Weick (1968), Hartmann and Wood (1982), and McCall (1984) document its spread to additional fields of inquiry and review the issues of feasibility, technology, reliability, and validity found in those fields.

In addition to early childhood development, Arrington cites systematic observation of adult factory workers in the 1930s by Thomas et al. (1933), Whitehead (1935), and Roethlisberger and Dickson (1939). Social psychologists interested in parent and child interactions employed SSO in the 1950s (e.g., Barker and Wright 1955). Boyd and Devault (1966) review SSO research in education, especially classroom behaviors, and Rosenshine and Furst (1973) discuss its application to the study of teaching. The methods of SSO were developed and applied to the study of police by Reiss (1968, 1971b) in the late 1960s, and by Sykes (1977, 1978) and Ostrom and her colleagues (Whitaker et al. 1982) in the 1970s. Mastrofski et al. (1998) further refined and applied these methods to policing in the 1990s. And a body of SSO research on controlling violence and aggression in bars and nightclubs emerged in the 1990s (Graham et al. 2000, 2004, 2005, 2006a, b; Homel and Clark 1994; Homel et al. 1997; Macintyre and Homel 1997).

Others have used SSO to study the pace of life in large U.S. cities (Levine 1997), the presence of social incivilities on residential blocks (Raudenbush and Sampson 1999; Taylor 1997; Taylor et al. 1985), the causes and effects of display of emotion by retail clerks (Rafaeli and Sutton 1989, 1990), the style of management in organizations (Mintzberg 1972; Luthans and Lockwood 1984), and many other phenomena. Recent SSO applications include measuring school environments (Wilcox et al. 2006), evaluating therapeutic communities for drug offenders (Bouffard et al. 2003), expression in service encounters (Pugh 2001), physician counseling of patients (Ellerbeck et al. 2001), the delivery of outpatient medical services (Stange et al. 1998), and evaluating iris scanning technology in three New Jersey schools (Uchida et al. 2004).

METHODOLOGICAL ISSUES AND DECISIONS

Here, we consider the suitability of SSO for the research question, sampling, instrumentation, recording observations, and dealing with various sources of error.

Suitability of SSO

What make SSO especially valuable to researchers gathering data directly in the natural setting are precision of the observations and the independence of the observer from that being observed (Reiss 1971b:4). For example, some classic qualitative field research pioneered researcher access to the police occupation (Skolnick 1966; Van Maanen 1974), but the necessarily selective samples of these solo researchers appear to have overstated the uniformity of police practice (Skogan and Frydl 2004:27). SSO researchers have observed considerable variation in the way police use their authority (Black 1980:ch. 5; Mastrofski et al. 2002; Terrill et al. 2003), and some have shown the high degree of variability that may be found with the same officer over time (Snipes and Mastrofski 1990). Precision is also accomplished through the sequencing of events and the detailing of context (Barnsely 1972:222), matters that may not be well documented by routine official records or accurately recalled by participants when interviewed – for example, how police encounters with the public escalate into rebellion or violence (Sykes and Brent 1983; Terrill 2001). Sometimes, SSO precision derives from the application of complex standards or expectations to the practices of persons with obligations to perform in particular ways: the extent to which legal actors conform to constitutional standards (Gould and Mastrofski 2004) or professional standards (Bayley and Garofalo 1989; Terrill 2003). But, SSO can also be used to determine the extent to which justice officials comply with the preferences of citizens they encounter or whether citizens comply with the preferences of justice officials (Mastrofski et al. 1996, 2000).

The observer's independence of the phenomena observed is a benefit and also a limitation of SSO, and it can be challenging to maintain. Of course, the survey interviewer is also independent of the respondent, but almost entirely dependent upon the respondent to serve as an informant. Interviewers used in victimization surveys are at the mercy of the candor and recall of their research subjects in attempting to determine the occurrence and nature of a crime. The SSO observers are not so dependent; they may rely upon what their senses tell them, recording elements of an event consistent with a protocol that examines the degree of evidence that a crime occurred.

SSO may be especially desirable when the question demands detailed knowledge of situations, conditions, or processes that are not otherwise well-illuminated or where there is reason to question the validity of knowledge based on other forms of observation. SSO may also be useful in studying people who might find it difficult to provide an objective or accurate account of what the researcher wishes to know (such as their behavior and the context of that behavior in highly emotional situations) (McCall 1984:266). Where there are strong temptations to omit, distort, or fabricate certain socially undesirable features, such as illegal, deviant, or otherwise embarrassing situations, SSO offers an independent account. This is, for example, a limitation of survey-based citizen self reports of encounters with police to deal with a problem caused by the survey respondent (Langan et al. 2001), and especially problematic if there is systematic variation in the degree of error across important subgroups within the sample, for example according to race (Engel and Calnon 2004:68).

While much of the SSO research has focused at the level of individual persons as decision makers, the 1980s saw the beginning of studies that use an ecological unit, such as the neighborhood block face, as the unit of SSO analysis. Noting that neighborhood residents find it difficult to offer accurate descriptions of their neighborhood's social and physical environment, Raudenbush and Sampson (1999), building on research by Taylor et al. (1984, 1985, Perkins et al. 1993), highlighted the value of an "ecometric" approach that uses SSO in conjunction with neighborhood survey research to more fruitfully characterize the state of neighborhood physical and social structure.

SSO may be especially well-suited to situations and events where all of the relevant actors and events pertinent to the phenomenon of interest can be observed from start to finish in a limited, well-defined time period. For example, the police decision on how to deal with a traffic violator is clearly bounded in time and place. To the extent that (as currently popular theory suggests) the decision is heavily influenced by the context of the immediate situation (e.g., the offense, the evidence, the driver's demeanor), the decision on how to treat the traffic offender lends itself to SSO (Schafer and Mastrofski 2005). However, observing behaviors relevant to a police chief's decision on establishing departmental traffic enforcement policy would require advance knowledge about when and where such activities take place, something that is seldom available.

Although some forms of large-scale human interactions would not seem so amenable to SSO, a creative approach may overcome the challenges. For example, studying the dynamics of mass protest and its potential for facilitating illegality (riots, for example), could be accomplished when researchers have sufficient advance notice of the organization of protests. Placing observers in the protest strategically, according to the dictates of one's theoretical framework, would allow SSO researchers to generate on-scene, objective accounts of whether certain presumed precursors to riot occur, when and how they occur, and the consequences.

In general, SSO lends itself to observing phenomena that occur either with high frequency, such as drivers' noncompliance with speed limits on public highways (Buckman and Lamberth 2001; Meehan and Ponder 2002; Smith et al. 2004), or at predictable times and places, such as criminal trials (Mileski 1971), field tests of prison security systems (Wilson and Mastrofski 2004), or scheduled meetings between probation officers and offenders. Events that occur less frequently, such as acts of social disorder in public places may require considerably more observation time to obtain reliable estimates (Raudenbush and Sampson 1999), or they may be so infrequent and unpredictable as to make SSO simply infeasible, such as the police use of lethal force or the life-course of criminality in a sample of individuals.

Processes, such as plea negotiations in the criminal process, may be ubiquitous, but they typically occur in ways that make it difficult for the researcher to follow negotiations in specific cases, since those negotiations do not occur at predictable times and places.

One of the most frequent uses of SSO has been investigating how criminal justice workers operate in the context of role expectations generated by their organization or profession (McCall 1984:267). Primarily focused on police, SSO research in criminology has been very concerned with how officers negotiate the tension between the formal (legal, bureaucratic, and professional) standards set for them and those that issue from the occupational culture. SSO could also be applied to role conformance in the context of informal or illegitimate organizations, such as gangs.

SSO is often used in conjunction with other forms of observation. Some studies have used SSO to measure the extent to which treatment conditions in randomized trials have been maintained (Sherman and Weisburd 1995:685). Sometimes, SSO is linked on a case-by-case basis to other forms of data collection on research subjects, such as census data (on neighborhoods), survey interviews of police officers, and follow-up interviews with citizens who were observed in encounters with police (Parks 1984; Percy 1980; Smith 1984, 1986). And sometimes, SSO is used to supply data not otherwise available, such as objective measures of the physical and social disorder in urban neighborhoods (Raudenbush and Sampson 1999).

Perhaps, the most frequent reason that criminologists have turned to SSO is their dissatisfaction with the data they could obtain by other means, such as official records and surveys of research (Buckle and Farrington 1984:63). For example, official records not only measure criminality, but may be affected by other considerations taken into account by those in the agency who record and process that information. Self-report and victim surveys, even though fairly commonplace, also have a number of biases and limitations (O'Brien 2000). As with any method, SSO has its methodological strengths and drawbacks, which will be discussed below.

Unit of Analysis and Sampling

Planning the selection of what is to be observed is often not an explicit feature of qualitative observation, but it is an essential element for SSO. Like survey interviewing, SSO requires a careful focusing of what is to be observed, and makes it possible to estimate parameters and evaluate error (Reiss 1968:357–358).

The first step is establishing the unit of analysis. Given that much SSO focuses on social interactions, there are three distinct approaches (McCall 1984:268–269). One uses a time period as the unit of analysis, observing what happens within each discrete time segment, such as what behaviors police officers show in a 15-min segment (Frank et al. 1997) or the level of social disorder on a street segment during a 1-h period (Weisburd et al. 2006). Another uses a behavior or act as the unit, tracking the sequencing of different behaviors over time, such as the behavioral transactions between officers and citizens who engage each other (Sykes and Whitney 1969). And a third approach is to socially construct an event, such as a face-to-face encounter between a police officer and citizen (Reiss 1971b) or a public meeting between police and members of a neighborhood organization (Skogan 2006).

Sometimes, defining the boundaries of an event-based unit of analysis, such as the encounter, is relatively straightforward. Noting when the parties to the encounter begin and end their interaction is not too difficult, but even here lie complexities. For example, a police

officer engages two citizens in a domestic dispute and then disengages them to go to another part of the house out of sight and sound of these two to engage another citizen for several minutes – before returning to re-engage the original citizens. To determine whether this constitutes one or multiple encounters requires careful thought about what constitutes an event of interest.

Sometimes, the challenge is not definitional, but rather observational, because the event of interest does not readily manifest itself in what can be observed. Take, for example, the police decision to stop a suspicious citizen, a decision that concerns criminologists wishing to determine whether police stops are influenced by the race of the citizen. The challenge for the field observer is determining when a citizen is actually observed by the officer, and when that observation is worthy of suspicion or in fact elicits suspicion in the officer. Researchers have attempted to do this by observing the officer for outward manifestations of police suspicion (the "doubletake") and to ask observed officers as part of the research protocol to announce to the researcher when they have formed suspicion so that the observer may inquire further what the officer observed (features of the citizen versus behavior of the citizen) and what judgments were made (Alpert et al. 2005). Of concern here is the regularity and reliability with which officers provide outward manifestations of suspicion forming and the regularity and reliability with which observers will accurately note them.

Once the unit of analysis is decided, the researcher must consider the sampling frame. Of course, the same principles of sampling apply to SSO as any other data collection method, such as survey research (Reiss 1971b:9–11). The researcher must consider where and when the units of interest may be found, and determine an efficient method of capturing a representative sample. An example of a straightforward sampling strategy is an SSO study of shoplifting that randomly selected shoppers entering a store, systematically varying the location of observers among entrances to the store (Buckle and Farrington 1984).

SSO researchers often use more complex sampling strategies focusing on geographic space. We have sampled police beats and specific days and times within them, oversampling places and times where higher levels of police–citizen encounters were expected (Mastrofski et al. 1998). Some researchers rely upon the observed subjects making their own choices as to where observers conduct their observations. This makes sense when the object of study is a specific research subject, but when the object of study is the geographic entity itself, an independent sampling plan is required. For example, a study of public order in a park required researchers to conduct hourly park patrols to observe and record activities of persons by location within the park (Knutsson 1997). Some researchers have used a smaller geographic unit than a police beat or park. Several studies use the face block to apply SSO to the measurement of disorder on public streets, defined in terms of traces of physical and social disorder (trash, graffiti, loitering, public intoxication) (Taylor et al. 1985; Sampson and Raudenbush 1999). In one study, observers drove through street segments videotaping what was viewable from the vehicle (*ibid*). Others have performed live observation at the "epicenter" of each block face (best location to observe the most activity), randomly selecting short periods of time for observation from that location and recording them on check sheets (Weisburd et al. 2006). But observers could focus on single addresses, as might be done if one were interested in observing the extent of different kinds of desired and undesired social activity at crime hotspots. Even smaller spatial units have served as the sampling frame. A study of the relationship between crowding and aggression in nightclubs selected high traffic areas within the establishment (10 sq m) to observe levels of patron aggression for 30-min time periods (Macintyre and Homel 1997).

While much of the extant SSO research must develop time- or area-based sampling frames that capture stochastically-distributed events, some SSO studies have focused on scheduled events, such as the delivery of therapeutic community programs in corrections institutions (Bouffard et al. 2003) or the previously-mentioned police–community neighborhood meetings. Sampling of regularly scheduled events is common in research on educational practices and physician behavior, a practice easily replicated for certain aspects of the legal process of interest to criminologists.

Sometimes, the practicalities of conducting successful field observation make the research vulnerable to sample biases. In cases wherein consent of those to be observed must be secured, a clear bias is introduced when those who refuse to be observed differ in their behaviors from those who are willing to be observed (Fyfe 2002; Mastrofski 2004). The sheer physics of observation can introduce bias as well. For example, the observation of disorder on Chicago block faces required light that was sufficient for observation only between 7 am and 7 pm (Sampson and Raudenbush 1999), meaning that researchers were unable to measure many forms of disorder that predominantly occur in the darkness (Kelling 2001). It would also be challenging to observe many aspects of law enforcement inquiry and exchanges in the investigative and prosecutorial processes, because much of the effort is not limited to face-to-face encounters, but rather occurs through telephone and computer, modes of communication that may necessitate very different sampling frames and observational methods. Particularly challenging are studies that require a sampling of *cases* rather than individual decision-makers, inasmuch as it is difficult to track and observe the behavior of many different persons who may be involved in making decisions about a case.

Instrumentation

Principles that apply to other forms of research also apply to the creation of instruments for structuring and recording SSO (Reiss 1971b:11). Sometimes, the instrument takes the form of a tally sheet or log for recording the frequency at which phenomena were observed, such as counting disorderly elements at block faces (Sampson and Raudenbush 1999) – or the timing and duration of events, such as police presence in a hotspot (Sherman and Weisburd 1995). Often, however, the instrument takes the form of a questionnaire, in this case, directed to the observer. For example, a study of police use of force might ask a series of close-ended questions about the citizens involved in an incident (their personal characteristics, their appearance, their behavior), the behavior of police (how much force was used at which junctures), and other features of the situation (location of the event, its visibility, the presence of bystanders) (Terrill and Mastrofski 2002).

SSO instruments, which can be brief or extensive, have the desired effect of focusing observers' attention on items selected for observation. Field researchers have demonstrated a substantial capacity to recall the relevant features of long sequences of these events, given the repetitive use of the protocols. Nonetheless, greater complexity in the coding system heightens the risk of error (McCall 1984:272). The accuracy of such recall is undoubtedly variable, but we are unaware of the research that has assessed most of the correlates of recall accuracy (e.g., observer characteristics, instrument characteristics, and the observational setting).

As with any questionnaire, the selection of fixed responses requires careful consideration, and detailed instructions may be necessary to standardize the classification of what

is observed. In research we have conducted, more than a page of text was required to instruct how citizens encountered by police should be sorted into four categories of wealth, based on their appearance, language, and context of the situation.

In addition to fixed-response items, SSO can be easily supplemented with open-ended items and general descriptions of events that are entered as qualitative data – narrative accounts. Sometimes, they constitute elaborations of what has been coded on fixed-response items, such as observer commentary on unusual events (Sampson and Raudenbush 1999:616). Sometimes, they introduce other forms of observation, such as when participants are "debriefed" to an event to learn how they perceived and interpreted an event that had just transpired (Mastrofski and Parks 1990). And sometimes, observers are asked to provide lengthy and less structured descriptions of the event (Reiss 1971b:13; Mastrofski et al. 1998). These narrative accounts are used to validate the reliability of observers' classification of phenomena, to document the time sequencing of certain events that is too cumbersome to incorporate into fixed-response items, to learn more about the perceptual framework of the observer, or to capture or clarify features of the event that were not anticipated in the construction of the fixed-response part of the instrument. The risk in using these less-structured modes alone is that narratives could be sufficiently *unsystematic* that they undermine much that SSO is designed to accomplish. Our approach to this problem has focused on detailing a priori the sorts of things that should appear in observers' narrative accounts, thereby adding more structure to the qualitative data than typically occurs in ethnographies (Mastrofski et al. 1998).

Recording Observations

Two issues arise in recording of phenomena observed through SSO: (a) whether it occurs contemporaneous with the observation or later, and (b) whether technological recording devices are employed. Resolving these issues requires choosing the highest priority and what must suffer as a consequence. The more contemporaneous the recording of an observation, the less the vulnerability to recall error and various forms of bias (e.g., when coding decisions about earlier observations are influenced by what subsequently was observed to happen), but in many cases, the act of recording may increase the reactivity of the observed parties to the process of being observed, as for example, when observers posing as shoppers follow actual shoppers to observe whether they are shoplifting (Buckle and Farrington 1984, 1994). Reiss's pioneering SSO of police required observers to complete some paper-and-pencil logs in the field to record the time, location, and participants of events (for example, what might fit on a clipboard) (1971b:14), but many later researchers opted for recording the information in less structured form (in small notebooks that would fit in one's pocket). In both cases, the data encoding occurred subsequently at a separate research office (Mastrofski et al. 1998). Observers attempted to reduce any concerns felt by the observed police by making their field recorded materials available to the officers to review, and by stressing that making these notes was simply part of their job, just as the police were required to record events as part of theirs. The lower level of obtrusiveness afforded by the limited-field-note approach must be balanced with concern for the accuracy of recall, a particularly salient issue when the observations per unit are extensive or complex. Consequently, these projects imposed a number of safeguards. Observers were trained to record keywords to stimulate memory, and completion of data entry took place within hours of departure from the field setting.

Some SSO projects ban any note taking by observers on scene, instead relying entirely on memory. For example, studies of pub violence placed pairs of unobtrusive observers (posing as patrons) in drinking establishments, making visits of 2 h or more (Homel and Clark 1994; Homel et al. 1997). Observers were expected to note hundreds of items, including covering the physical environment, staff levels and characteristics, behavior of patrons and staff regarding deviance, alcohol consumption and efforts to control it and violence, forms of aggression, and the dynamics of incidents involving violence. Observers recorded their observations according to written protocols as soon as possible after leaving the pub.

Employing technological aids is usually intended to increase the accuracy or detail of observation from that which would be otherwise available. Handheld electronic recording devices have been used in observing police–public interactions (Sykes 1977) and in observing the social and physical environment of neighborhoods (Gravlee et al. 2006). Audio-taping of calls for service to police telephone operators has been used to gather data on police workload (Scott 1981). Videotaping neighborhood block faces from a slow-moving motor vehicle has been used to observe neighborhood disorder (Sampson and Raudenbush 1999). Handheld personal digital devices allow contemporaneous observation and recording of brief, frequent events and are most practical when the number of aspects to be observed per event are small in number, which minimizes the interference of recording events occurring in close succession with observing them (McCall 1984:272).

The major advantage of initially recording events electronically and then encoding those records later for analysis is not only the elimination of recall problems, but also that more detailed and accurate observations may be made, and the testing of inter-observer reliability is facilitated. Further, field researchers may in some instances feel safer when they make the initial recording of their observations from the security of a moving vehicle or when events are recorded with a remote and unobtrusive device (e.g., time-lapse photography of a street corner or address). Nonetheless, there are a number of drawbacks that have concerned other researchers. Video recordings cannot exactly replicate what all of the available senses would communicate to an observer who was there "in the moment." The use of drive-by photography can be expensive; it raises ethical concerns because denizens of the neighborhood find them intrusive and perhaps anxiety producing; it may raise legal and human subjects protection issues if alleged criminal acts are recorded; and they may cause significant reactivity among those being observed (Caughy et al. 2001; Gravlee et al. 2006; Kelling 2001; but cf Smith et al. 1975). In other instances, however, the pervasiveness of already-present surveillance technology that records observations in readily shared (digital) formats (closed circuit television in public and mass-private settings or in-car police video cameras to record traffic stops) may afford researchers a relatively unobtrusive source of data that does not encounter these problems. Dabney et al. (2004), for example, used augmented video surveillance to study shoplifters. The researchers followed a sample of drugstore customers at a single location equipped with high resolution video surveillance cameras. They were able to determine which customers engaged in shoplifting and coded data on customers' personal characteristics, as well as behavior. Increasingly, much technology-based surveillance derives its unobtrusiveness, not from its being unknown to subjects, but that it is taken-for-granted (Shrum et al. 2005:11). And with the advent of nonlinear editing packages for digital video, the data itself (traditionally analyzed in quantitative or text format) can be readily manipulated and analyzed as images (Shrum et al. 2005:5).

Dealing with Error, Reliability, and Validity

SSO data are subject to the same range of threats that befall other methods. Here we consider errors introduced by the observer, reliability, and validity issues.

Observer Error. Observers can introduce error intentionally (cheating) or unintentionally (bias or reactivity). Cheating is rarely reported in SSO, although its frequency is unknown. In our own experience (with approximately 100 observers covering about 15,000 h of observation across five projects), we have uncovered but one instance, a situation where the observer fabricated an excuse to switch officers so that she could accompany one in whom she had developed a romantic interest. It seems likely that most instances of SSO cheating go undetected. Detection in this case occurred because the observer carelessly recorded an account of her actions on the same medium used to submit her observation data, which one of her supervisors read when reviewing her materials.

A more subtle form of cheating, shirking, may occur in SSO if observers attempt to reduce their postobservation workload by failing to record events that would require extensive narratives and structured coding. Although there has been no direct systematic assessment of the extent and impact of this form of shirking in SSO, one researcher did examine the effects of time-on-the-job on researcher productivity. Presumably, project burnout might encourage some to engage in workload shedding as the project progresses, but the analysis showed that observers' productivity was not significantly lower late in the project, when compared with the earlier time period (Spano 2005:606–608).

One can assume that, as with any employment situation, instances of cheating are reduced to the extent that the research enterprise (a) selects as observers persons who are committed to integrity, (b) who are socialized during training to maintain it, (c) whose morale is an important concern of those running the operation, and (d) whose field work and work products are carefully monitored and given feedback at the time that researchers are in the field.

Potential sources of unintended biases in SSO are the mindset and prejudices that observers bring to the field or develop on the job, as these may affect what they observe and how they interpret it. The limited extant research exploring these issues for SSO does not offer clear and consistent findings but is dependent upon the particular information in question. Reiss (1968, 1971b:17–18) found that an observer's professional background (law student, sociology student, or police officer) did have consequences for some types of information, but not others. A later study attempted to determine whether a statistical relationship between observed police orientation to community policing and officer success in securing citizen compliance could be attributed to observers' own views on community policing (Mastrofski et al. 1996:295). A clear association was not found between the observers' attitudes and the effects that their analysis produced.

While there is no reason to expect observer bias in SSO to be any more or less a problem than in survey research, the nature of SSO does present some interesting challenges. For example, it is more difficult to check for biases in SSO, at least when the field observation cannot be directly observed, as is usually the case when observers accompany criminal justice decision makers or offenders far from a controlled or predictable setting. A significant contributor to reduction in survey interviewer bias (and shirking) in recent years has been the change from in-person interviews conducted in the field to telephone interviews by interviewers located in call centers, where close monitoring is both possible and unobtrusive.

There are, however, a number of methods for detecting biases. If other, independent sources of data are available, comparisons can be made between what SSO generated and what the observer generated. Although official crime and justice documents often fail to capture the full range of events that interest the researcher, where there is overlap, crosschecking the observations of the SSO field worker is possible, such as crime and arrest reports filed by police officers. The extensive records of officer activities captured in modern police communication systems – both voice and data – are another potential source for cross-validation. Another option is to test observers in a simulated environment, for example, having them evaluate a video of one or more situations and comparing their observations to each other or some presumed objective standard (Mastrofski et al. 1998:9). Another is to look for unexpected or unusual patterns in observer-generated data from field observations, especially when there are a sufficient number of observations to make it reasonable to expect certain sorts of observations to fall within a given range (Mastrofski et al. 1998:14). And of course, one can monitor observers by periodically surveying or debriefing them about their beliefs and perceptions or by requiring that they constantly self-observe in a diary-like format (Reiss 1971b; Spano 2005). One must remain sensitive, however, to the weak and often subtle links between observer attitudes and their observational judgments and behaviors. Well-trained and supervised observers may have or develop stronger judgments and beliefs about what they observe, but still remain true to the research protocols designed to limit the effects of personal biases.

Some types of observation judgment are undoubtedly more vulnerable to personal bias than others. For example, some research required field observers to judge whether police officers applied excessive force against citizens (Worden 1996:36). But it may prove more effective to bifurcate the process into (a) recording narrative accounts of what happened (without asking the field observer to make a judgment about excessive force), and (b) having separate, specially trained experts review these accounts and make an independent judgment about whether they constitute a violation of some standard, legal or otherwise (Gould and Mastrofski 2004).

One particular source of observer error is extended exposure to field conditions. Researchers may "go native" (adopting the perspective of their research subjects), or they may experience "burnout," cynicism about what they are observing. The closer the observer is to the phenomena observed, and the more beneficial the establishment of rapport with the research subject, the greater the risk of going native. An examination of these effects in an SSO study of police found little evidence of a significant change in attitudes toward police between the beginning and end of the study, and surprisingly found that observers who developed more positive attitudes toward the police over time were *less* likely to render assistance to police officers while observing them, albeit the relationship was not statistically significant (Spano 2005:605–606).

Of all the sources of observer error in SSO, reactivity receives the most attention. Reactivity occurs when the method of observation alters what is being observed (Kazdin 1979). In general, the less obtrusive the observation of human behavior, the lower is the risk of reactivity. Much has been written about "interviewer effects" – reactivity in survey research and ways to reduce it (Webb et al. 1981). Much less is known about reactivity in SSO. Measuring SSO's reactivity is challenging, because often SSO is the only means available to measure the object of study, and even when other methods of observation are available, they may have their own reactivity issues. It may be possible to make inferences by comparing the results of different forms of observation, some presumably more conspicuous than others (Brackett et al. 2007). But for SSO by observers who are noticeably present, the challenges of comparing behavior

when observers are present to when they are not is nigh impossible unless there are other unobtrusive ways of monitoring the behavior in question both when observers are and are not present. This is rarely possible.

No responsible researcher dismisses the risks of reactivity in SSO, but a number suggest that it is dependent on the context of the observational setting and the nature of the relationship between observer and observed (McCall 1984:273–274). In SSO of police, it has been argued that reactivity to the observer can be reduced by the observer downplaying any evaluative role and emphasizing one's naivety as a "learner" (Mastrofski et al. 1998; Reiss 1968). Yet, even this approach may generate more "teaching" or show-off activity in police subjects. Observed officers who engage in such teaching have been noted to self report making contact with citizens to illustrate elements of police work to the observer (Mastrofski and Parks 1990:487). And some types of observers (for example females in the presence of male police officers), may produce more of this effect than others (Spano 2007:461).

Certainly, some types of research subject behavior are more susceptible to the presence of an observer than others; the stronger the social valence of a particular pattern, the greater the risk that behavior will be altered. Nonetheless, many well-ingrained patterns of behavior may be especially difficult to alter (Reiss 1971a), including such things as racial discrimination and violations of constitutional rights, especially if the making of this particular determination is not an obvious purpose of the research. And it has been noted that research subjects often habituate to the presence of observers over time (McCall 1984:274; Spano 2007:461). Compared to other forms of data (interviews, self-reported behavior, agency records) SSO may well be less reactive. Interview respondents can easily offer inaccurate or distorted answers that are difficult to detect. An official can easily misrepresent what he or she reports on agency records where there is a weak organizational capacity to detect it. When SSO researchers are present to observe the event, especially when the subject is engaging other people, changing his or her habitual behavior may be far more challenging (Reiss 1971b:24). Such manipulations may be especially difficult where skill levels and experience influence the research subject's course of action.

Controlling reactivity is accomplished by (a) making the observation less obtrusive, (b) making the observation less consequential, and (c) selection, training, and monitoring observers to remain within the appropriate limits of their research role (McCall 1984:275). Exposing research subjects to repeated observation facilitates the habituation effect, as does establishing rapport, although it may be difficult to know in advance how much observation is needed to reduce reactivity to an acceptable level (Barnsely 1972:227). Assurances of confidentiality (when possible) that are also reinforced by action can reduce the consequences of observing something that research subjects would regard as socially undesirable. Selecting persons willing and able to be taught to avoid comments and behaviors that provoke reactivity (e.g., the expression of judgments), and following up that training with supervision while in the field can reinforce the appropriate researcher role for a given study.

However, some of these reactivity countermeasures can suppress one kind of reactivity and stimulate others. For example, using an expert to observe the same kind of expert may make it easier for the observer to establish rapport, but it might also make it more difficult for the observer to refrain from offering an opinion or assistance, based on a claim of ignorance or inexperience. Repeatedly reminding research subjects of the project's confidentiality guarantees reminds them that they are the object of an outsider's attentions and will result in a record of their activities. And the best persons and methods for establishing rapport with a research subject may in fact also stimulate researcher involvement in activities that are

decidedly beyond the research role (Reiss 1968). Clearly, some compromises must be made in drawing the line between what is and is not acceptable observer behavior in attempting to control reactivity.

Reliability. One of the distinct advantages of SSO over solo field research is that it facilitates the testing and improvement of the reliability of observations (Reiss 1968, 1971b). Early on, much attention was given to the use of multiple observers and estimating their inter-rater reliability. Where many researchers can independently observe the same phenomenon (by having multiple observers on scene or by using video recordings), the testing of inter-rater reliability is accomplished by measuring the extent of agreement among the pool of observers for the same set of events (Raudenbush and Sampson 1999:14; Homel et al. 1997:63; Wilcox et al. 2006). Sometimes, disparate independent observations of the same event are resolved by a process of discussion and negotiation (Graham et al. 2006b:284). Where multiple independent observations of the same event are not possible, and that is often the case in situations wherein having more than one observer would be too disruptive, observers might be tested by using their detailed narrative descriptions to determine (a) if they are properly classifying phenomena according to the protocol and (b) the extent of agreement among persons who use those narratives to make classifications (Reiss 1971b:22). For example, this has been done for characterizing a wide range of police and citizen behaviors in predicting citizen compliance with police requests (McCluskey 2003:60–74).

Recently, SSO researchers have broadened their reliability concerns to incorporate measurement accuracy and stability (McCall 1984:275). Raudenbush and Sampson (1999) apply psychometrics to the development of "ecometrics" to better understand the error properties of SSO data gathered from observing physical and social disorder in urban neighborhoods. They adapt three psychometric analytic strategies: item response modeling, generalizability theory, and factor analysis to illuminate the error structure of their observational data and to make judgments about the best ways to limit different sources of error in future observational studies. For example, they find that physical disorder can be more reliably measured at lower levels of aggregation than social disorder, due to the much lower frequency of the latter in their observations (Raudenbush and Sampson 1999:30).

Validity. While texts often note that field studies are less vulnerable to validity problems than surveys because the method places the observer "there" while events are unfolding (Maxfield and Babbie 1998:286), they are of course only valid insofar as they produce data that measure what the researcher intends for them to measure (McCall 1984:276). For the most part, the challenges of assessing the validity of SSO data are those common to other forms of data collection, and we will not elaborate them here. However, we will raise three points that may be useful to those contemplating the use of SSO.

First, SSO in its purest form is not well-suited for observing internal thought processes and motivations. An example is inferring the level of intent to harm someone solely from naturalistic observation of human interaction, such as might take place at school or bars (Boulton 1991; Graham et al. 2006a). Intent is a key concept in theories of aggression and violence. The search for reliable, observable cues that reveal the degree of intent is challenging, not only because the links between cues and intentions are not empirically well validated, but also because of the ambiguity of certain cues: when is a person so intoxicated that his behavior cannot have been guided by a high degree of intent to harm (Graham et al. 2006b:285)?

Observing judgments can also be challenging for SSO. Theories about the impact of procedural justice on citizen compliance and law abidingness have been expressed in terms that readily admit their testing from survey data (e.g., whether a respondent perceived a given judicial action as procedurally just) (Tyler and Huo 2002). While many social scientists have been willing to accept the respondent's answer as a valid reflection of his or her assessment, this has not been readily available to practitioners of SSO. Consequently, they have three options. One is to take whatever behavioral indicators are available and infer thoughts and feelings from them. This may be appropriate for some things (general citizen satisfaction with the police) (see McIver and Parks 1983:31), but rather risky for others (citizen judgments about specific aspects of police action). Another option is to measure procedural justice objectively from what can be observed (for example, whether the officer showed respect, explained what he was doing, referred to evidence) (Mastrofski et al. 1996; McCluskey et al. 1999). The problem is, of course, that these things may not in fact register as procedurally just for the citizen being observed. However, this objectivity does have the clear policy-relevant advantage of allowing the researcher to determine how influential a given police behavior is on a citizen's response, such as compliance. A third option is to observe events and then later debrief those involved to elicit what they were thinking and feeling at the time (Mastrofski and Parks 1990). This mixed approach to gather both objective and subjective data about an event has been used in police research (Alpert et al. 2005; Mastrofski et al. 1998), but its error and validity properties have not been explored.

A second concern with the validity of measures derived from SSO is their generalizability, especially across different cultures. For example, an SSO study that attempted to predict when police would show disrespect toward a suspect used as one predictor whether the police were disrespectful (Reisig et al. 2004). The sample of citizens was diverse in terms of race, age, and wealth, and it is conceivable that conceptions of what is disrespectful could be conditioned by those variables. That police disrespect failed to predict suspect disrespect was consistent with the researchers' theoretical expectations, but the failure to find the effect could have been due to variability in the sample of how disrespect is subjectively defined. Under these circumstances, tests of criterion validity are appropriate, but not conclusive, according to the relevant cultural classifications where differences are expected.

A third validity issue has to do, not with what the observers recorded, but what judgments are made about it. For example, SSO researchers may use experts to generate judgments about the likelihood that an observed event would be treated in a particular way – for example, the probability that an officer's behavior would be regarded in court as a violation of Fourth Amendment standards for search and seizure (Gould and Mastrofski 2004). It is appropriate to think of the constitutionality of a police action in terms of probabilities because there are many courts and many judges, and they can vary in their decision, especially in hard or borderline cases. The question is, did the researchers produce a valid assessment of what the relevant pool of judges would have produced? Unable to submit their observations to such a sample of judges, the researchers performed a sensitivity analysis to see if shifts in the scale of probabilities (more and less lenient judgments about constitutional violations) would alter results.

PRACTICAL CONSIDERATIONS

There are a number of practical issues that arise when fielding an SSO research project. Among these are matters pertaining to the research staff and matters relating to the research subjects.

Organization and Staffing of the Research

SSO typically involves the efforts of a group of researchers, and as with any group effort, organizational issues require attention. SSO researchers are often recruited from the ranks of graduate and undergraduate university students, but not always; they may come from the ranks of the persons being observed or from some other group (Reiss 1968). For example, an SSO study of shoplifters used trained psychologists experienced in observing children in classrooms. They had received additional training from store detectives on how to observe unobtrusively (Buckle and Farrington 1984:67). The complexity of the tasks required (interacting with observed research subjects and gathering/recording data) – even when the observer has experience in the role of the people being observed – will likely call for persons with a certain temperament as well as certain aptitudes and skills, especially for projects that place the observer in a participant role.

Our experience with SSO of patrol officers has shown a need for the following skills and traits to deal with the complexities of police field work and a demanding observation protocol: strong writing skills, ease with social conversation, curiosity about policing, self discipline and strong work ethic, team player, candor and integrity, and facility with word processing software. Absent from this list is expertise in policing and the law, in part because we provide researchers with the basics of what they need to know during training, and when expertise is necessary (such as judging the constitutionality of an action), we draw on a separate pool of experts *ex post facto* to code the data collected by our field staff. It is much harder to find persons who make good field researchers *and* have the requisite subject matter expertise. Where SSO projects require little or no field involvement with research subjects, we anticipate that a different set of skills and priorities would be relevant. On the other hand, we can imagine certain settings (courts) where observers with extensive legal training would be highly desired to note and interpret events. Finding the best mix of skills and traits depends upon the particulars of the job task, and innovative SSO projects may require an iteration or two in the field for the investigator to learn precisely what is most valuable in a field researcher.

Regardless of the recruitment pool, researchers who are inexperienced with SSO may underestimate the amount of time and effort it takes to train their field staff. Naturally, the more complex and elaborate the research instrument, the more training time required, and the more direct involvement or exposure the field researchers will have with the research subjects, the more training required. In our SSO of police, we have expended considerable training time on familiarizing prospective field researchers with the manners and customs of the officers and the communities where they will be observed – long before they logged any training time in the patrol car. This has included using representatives from the police organization to deliver presentations on the department.

The importance of establishing the specific role expectations for researchers cannot be overstated (McCall 1984:273; Reiss 1968, 1971b). Where to draw the lines between permissible and impermissible participation with research subjects is not only important for methodological purposes, but also for legal and ethical ones. Given that many SSO projects place observers in the field for extended time, it is important to train observers in how to maintain a professional, yet congenial relationship with research subjects, even when encouraged or pressured to engage in activities that are inappropriate. Establishing some bright lines around forbidden practices is essential.

As with survey research, quality control of SSO calls for attentive monitoring of the field researchers and their work products. Using supervisors experienced in conducting that form of SSO is a major advantage, not only because they will know more about the challenges, but

also they will have greater credibility with the observers they supervise. They should meet frequently with the field researchers to discuss not only their research products, but also their feelings and state of mind. Computer software can be developed to check for logical inconsistencies in coding and to flag potential problems that deserve double checking (Mastrofski et al. 1998:14). If project resources do not allow detailed checking of each item or case, then researchers should consider sampling items or cases to catch potential problems as quickly as possible. Errors can pile up quickly, so it is valuable to give rapid feedback to field researchers.

Management of field observers goes well beyond the need to check for problems with various sources of measurement error that have already been described. There are a number of issues that will likely arise, depending upon the length and social context of the observation period. Supervisors need to be readily available to counsel people who have observed traumatic events. The daily exposure to bleak social conditions and human frailty can have significant effects on the observers' emotions and outlooks. And of course, supervisors need to monitor and manage, to some extent, relationships among the researchers themselves, some of whom will experience conflicts and crises found in any organization.

A host of other practical issues arise in fielding SSO, many of which are reviewed by Reiss (1971b:24–27) and Mastrofski et al. (1998). These include the legal liability of the observers (should their presence or action stimulate a lawsuit), the legal liability of the project's organization, and the availability of civil action to the investigator and the employees for harms that may arise in the course of the research. Of particular concern may be the risk of injury to field researchers and their vulnerability to legal action, including the subpoena of observers and their work products. In the case of protecting the confidentiality of observations, federally-sponsored research may offer some protections, but they are limited, which raises concern about the power of the research organization to compel its employees to maintain certain confidentiality guarantees in the face of legal action (Mastrofski et al. 1998:16). In general, investigators should seek the advice of qualified legal counsel to manage these risks.

Beyond that, investigators are well advised to incorporate the project's key governing policies into the terms of the field observers' contracts. Due diligence requires that this carry over to training and supervision for the SSO project. Much of the pressure for revealing protected information is delivered not through formal legal channels, but informally while researchers are in the field. Learning how to deflect these and when to summon supervisors for assistance should be part of the training. We recommend that investigators incorporate some role-playing sessions involving scenarios where observers are presented with inducements or pressure to violate policies, providing instruction and practice on how to respond to these situations.

Although SSO can be quite labor intensive, its efficiency (amount of desired data acquired per unit of cost), can be quite variable. For example, it is undoubtedly more efficient to gather data on shoplifting by surveying a sample of shoppers or gathering data from store records about shoppers and shoplifters. Indeed, one such study committed approximately 755 h of field observer time (two observers per sampled shopper) to yield only nine instances of shoplifting (Buckle and Farrington 1984). SSO studies of patrol work tend to yield on average slightly less than one police–citizen encounter per hour of field work (Parks et al. 1998:2–23). A survey researcher doing follow-up calls with citizens who have had recent contact with the police could be more efficient, but even that can be time consuming, given the difficulty of catching such persons at home and persuading them to participate. On the other hand, SSO studies that deal with how certain kinds of subjects (criminal justice officials or offenders) spend their time can be fairly efficient, since each moment of observation

time yields relevant data. Of course, if electronic recordings of the phenomenon of interest are already available (for example, CCTV recordings of specific locations), costs may be limited primarily to interpreting and coding those recordings, making this method more efficient than alternatives to gathering data. Ultimately, efficiency is usually only one consideration for researchers, and so the real issue is whether there are sufficient gains in access to relevant data, and improvements in validity and reliability of those data to justify the SSO investment.

Preparing Organizations and Research Subjects for SSO

Many research subjects have some degree of familiarity with what is involved in survey research, interviews, and other forms of data gathering that involve subject interaction with researchers. But most know nothing about SSO, what it requires of the research subject, or what it requires of the researcher. Preparation of the research subjects is important, not only so that the requirements for informed human subjects research are satisfied, but so that the project is conducted with minimum error and hassle. If the research subjects work within a formal organization, advance time spent explaining the project and its methods at all levels of the organization, but especially the immediate supervisors of the persons to be observed, is time well spent. Conducting a few "demonstration" observation sessions will also help to reduce anxieties or uncover unanticipated problems that can be resolved before actual data collection is launched. Over the course of the study, frequent contact with persons at various levels within the organization is advisable, to detect and deal with problems quickly.

Among the essential elements of a good relationship with the research subjects is making clear to them how human subjects research issues will be handled. Not infrequently, research subjects are skeptical about the willingness of field researchers to maintain confidentiality guarantees, and some may go so far as to test them in subtle, as well as not-so-subtle ways. This may arise in the course of casual conversation about something that researchers may have observed previously with other research subjects who are known to both the researcher and the current research subject. Sensitizing both the researchers and the research subjects to the need to avoid discussion of these events is particularly important. On the not-so-subtle side, supervisors of research subjects (e.g., police sergeants) may exert tremendous pressure on field observers to reveal information that may be inculpatory or exculpatory for an observed subordinate. Invoking project policy and, if necessary, summoning a research supervisor may be the most effective way of avoiding unnecessary revelations of protected information.

SSO CONTRIBUTIONS TO CRIMINOLOGY AND FUTURE OPPORTUNITIES

SSO has been heavily concentrated on gathering data on two aspects of criminology: the behavior of police patrol officers and disorder in the public parts of urban neighborhoods. We offer a comprehensive review of neither body of literature, but summarize key contributions that SSO has made in advancing knowledge in each area. We then suggest a few areas that seem especially promising for future SSO applications.

SSO data have dominated the empirical research that describes and accounts for variation in the discretionary choices of officers: making stops and arrests, issuing citations, using force,

assisting citizens, and displaying procedural justice (Skogan and Frydl 2004:ch. 4). One of SSO's special contributions has been the scope of explanatory elements made available for the researchers' models. These include many details of not only the officer's behavior, but also the context in which it occurs (nature of the participants, their behavior, the location, and the neighborhood). SSO has also been instrumental in detailing the nature of the *process* of police–citizen interaction, opening our eyes to the interactive quality of temporally ordered micro-transactions or stages that may occur in even a relatively short police–citizen face-to-face encounter (Bayley 1986; Sykes and Brent 1983; Terrill 2003). And SSO has also allowed researchers to observe elements of organizational control and community influence on the work of police officers. For example, we learn more about the influence of police supervisors on subordinates' practices (Engel 2000), and we learn the dynamics of police–community interaction and their consequences when police and neighborhood residents deal with each other at community problem-solving meetings (Skogan 2006).

A second area where SSO research has concentrated is the examination of neighborhood physical and social disorder. Ralph Taylor and colleagues pioneered this focus in the 1980s in Baltimore (Taylor 2001; Taylor et al. 1985), and it has been used to test the impact of police interventions in hotspots, showing that police interventions in these "micro-places" not only reduce crime and disorder, they also diffuse those benefits to nearby areas (Sherman and Weisburd 1995; Weisburd et al. 2006). The largest project in this area has focused on Chicago neighborhoods and has produced a number of insights relevant to the testing and development of theories of the role of neighborhood disorder in causing crime in urban neighborhoods (Raudenbush and Sampson 1999). Using SSO-based measures of "objective" disorder described earlier in this chapter, researchers have examined the sources and consequences of public disorder. The research has demonstrated the importance of "collective efficacy" in predicting lower crime rates and observed disorder, controlling for structural characteristics of the neighborhood (Sampson and Raudenbush 1999). Collective efficacy also predicted lower levels of crime, controlling for observed disorder and the reciprocal effects of violence. The researchers found that the relationship between public disorder and crime is spurious, with the exception of robbery, which is contrary to the expectations of the well-known "broken windows" theory of neighborhood decline.

In general, SSO has afforded precision that has in many cases shown the phenomena of interest to be more complex than other forms of data collection had indicated. For example, we have already noted that SSO researchers have found rich variation among police officers in their patterns of discretionary choice, and even noted the instability of those patterns for individual officers over time. And the independence of SSO observers from the phenomenon of interest has provided a means to understand the contributing factors to the social construction of phenomena, such as the contributions of a neighborhood's racial profile in assessing its level of disorder (Sampson and Raudenbush 2004).

Beyond the above areas of recent high visibility in criminology, there are many opportunities to expand the use of SSO to improve knowledge. Largely untapped is the observation of crime and disorder, especially at the micro-level, where observers have the opportunity to make detailed observations of offenders in the act. SSO studies of shoplifting and aggressive or disorderly behavior in bars and clubs show that this is feasible where observers can easily blend into the environment. Where that is not possible, access to unobtrusive surveillance technologies appears to offer opportunities for detailed observation that reduce reactivity concerns. Indeed, it seems highly likely that in the future, criminologists will take advantage of the ubiquity of electronic surveillance in society generally to capture events that would otherwise be costly to observe. For example, the growing sophistication of surveillance and

identification technology, may make it possible to use facial identification software to gather data for a network analysis of persons who frequent hotspots. This includes not only the growing use of video recording devices by government and private sector organizations, but the now ready availability of miniaturized recording devices to the general public (through cell phone recording devices).

In searching for efficient ways to use SSO, criminologists should bear in mind the growing body of evidence about the predictability of crime and disorder occurring in small geographic spaces ("hotspots") (Weisburd 2008). Because much "street" crime is so highly concentrated in a relatively small portion of addresses or face blocks, the location of observers or observational devices can very efficiently generate lots of information on what occurs, especially in public areas. In addition, given heightened levels of obtrusive surveillance (security procedures in public places), SSO should prove an excellent way of better understanding how security and surveillance operate, why certain methods are effective, and the collateral impacts of various methods of monitoring and control designed to increase public safety.

Another venue for SSO to be used fruitfully is in experimental studies. We have already identified some field experiments that used SSO to validate the application of treatment protocols and to measure outcomes. But beyond that, SSO can be used to measure key aspects of the process that presumably operate to link treatments to outcomes. For example, if the physical redesign of bars and serving practices of bartenders is intended to reduce violence in those establishments, do patrons in fact alter their patterns of behavior in the ways that are expected to produce less violence (Graham et al. 2004)?

CONCLUSION

Four decades after Albert Reiss showed criminologists the utility of systematic social observation, it remains a method used infrequently in the discipline. This is undoubtedly due in no small part to two things. First, criminologists are rarely exposed to training and opportunities to do SSO during their course of study. Hence, they remain ignorant of its applications or unfamiliar with how to do it. This could be remedied by increasing researchers' exposure to the method in coursework and the early, formative years of their work. But doing this requires a cadre of researchers sufficiently skilled and experienced to teach it. One way to prime the pump would be holding methods workshops, sponsored by organizations such as ICPSR. Second, those who know a little about it may often expect that it requires more time and resources than they have available. This may indeed be the case, but we suspect that many projects could be taken on a smaller scale with a narrower scope of questions than the better-known, large SSO projects. Some researchers may decline to use SSO because of reactivity concerns, but the available evidence suggests that these problems are often manageable and may be no more severe in any event than found with other data gathering methods.

Increased use of SSO will undoubtedly attract and stimulate greater scrutiny of its limitations, as well as its advantages. Certainly, the error properties of most SSO data sets have been underexplored, and more attention is needed here. We expect that expanding the use of SSO and more comprehensively assessing its strengths and limits can be fruitfully combined into a more comprehensive assessment of other methods of gathering data on crime and justice phenomena.

SSO deserves the consideration of researchers because of its many advantages. For the study of crime and justice phenomena, it offers enhanced prospects of validity, and in many

situations it provides for increased confidence in reliability, because of the researcher's direct access to the phenomenon of interest and greater control and transparency of data encoding. Further, it affords greater precision in capturing details of the phenomenon and its context, such as the sequencing of what happens before, during, and after those events. In many cases, it may be the least problematic method for acquiring information. Criminology, which has strong roots in the traditions and methodologies of sociological research, remains heavily reliant on the use of sample surveys and official records (McCall 1984:277; Reiss 1971b). But, as the field matures and diversifies intellectually, more of its researchers may, with justification, be inclined to make systematic social observation the method of first, not last resort.

REFERENCES

Alpert GP, Macdonald JM, Dunham RG (2005) Police suspicion and discretionary decision making during citizen stops. Criminology 43(2):407–434

Arrington RE (1943) Time sampling in studies of social behavior: a critical review of techniques and results with research suggestions. Psychol Bull 40(2):81–124

Barker RG, Wright HF (1955) The Midwest and its children. Row, Peterson, Evanston, IL

Barnsely JH (1972) The social reality of ethics: the comparative analysis of moral codes. Routledge and Kegan Paul, Boston

Bayley DH (1986) The tactical choices of police patrol officers. J Crim Justice 14:329–348

Bayley DH, Garofalo J (1989) Management of violence by police patrol officers. Criminology 27:1–25

Black D (1980) The manner and customs of the police. Academic, San Diego, CA

Bouffard JA, Taxman FS, Silverman R (2003) Improving process evaluations of correctional programs by using a comprehensive evaluation methodology. Eval Program Plann 26:149–161

Boulton MJ (1991) A comparison of structural and contextual features of middle school children's playfull and aggressive fighting. Ethol Sociobiol 12:119–145

Boyd RD, Devault MV (1966) The observation and recording of behavior. Rev Educ Res 36(5):529–551

Brackett L, Reid DH, Green CW (2007) Effects of reactivity to observations on staff performance. J Appl Behav Anal 40:191–195

Buckle A, Farrington DP (1984) An observational study of shoplifting. Br J Criminol 24(1):63–73

Buckle A, Farrington DP (1994) Measuring shoplifting by systematic observation: a replication study. Psychol Crime Law 1(2):133–141

Buckman WH, Lamberth J (2001) Challenging racial profiles: attacking Jim Crow on the interstate. Rutgers Race Law Rev 3:83–115

Caughy MO, O'Campo PJ, Patterson J (2001) A brief observational measure for urban neighborhoods. Health Place 7:225–236

Dabney DA, Hollinger RC, Dugan L (2004) Who actually steals? A study of covertly observed shoplifters. Justice Q 21(4):693–728

Ellerbeck EF, Ahluwalia JS, Jolicoeur DG, Gladden J, Mosier MC (2001) Direct observation of smoking cessation activities in primary care practice. J Fam Pract 50(8). http://www.jfponline.com/toc.asp?FID=278&issue=August%202001&folder_description=August%202001%20(Vol.%2050,%20No.%208) Accessed January 31, 2009

Engel RS (2000) The effects of supervisory styles on patrol officer behavior. Police Q 3(3):262–293

Engel RS, Calnon J (2004) Examining the influence of drivers' characteristics during traffic stops with police: results from a national survey. Justice Q 21(1):49–90

Frank J, Brandl SG, Watkins RC (1997) The content of community policing: a comparison of the daily activities of community and "beat" officers. Policing Int J Police Strateg Manage 20:716–728

Fyfe JJ (2002) Too many missing cases: holes in our knowledge about police use of force. Report prepared for the National Research Council, Committee to Review Research on Police Policy and Practices Data Collection Workshop, Washington, DC

Gould JB, Mastrofski SD (2004) Suspect searches: assessing police behavior under the constitution. Criminol Public Policy 3:316–362

Graham K, West P, Wells S (2000) Evaluating theories of alcohol-related aggression using observations of young adults in bars. Addiction 95:847–863

Graham K, Osgood D, Zibrowski E, Purcell J, Gliksman L, Leonard K, Pernanen K, Saltz R, Toomey T (2004) The effect of the Safer Bars programme on physical aggression in bars: results of a randomized controlled trial. Drug Alcohol Rev 23:31–41

Graham K, Bernards S, Osgood D, Homel R, Purcell J (2005) Guardians and handlers: the role of bar staff in preventing and managing aggression. Addiction 100:755–766

Graham K, Bernards S, Osgood D, Wells S (2006a) Bad nights or bad bars? Multi-level analysis of environmental predictors of aggression in late-night large-capacity bars and clubs. Addiction 101:1569–1580

Graham K, Tremblay PF, Wells S, Pernanen K, Purcell J, Jelley J (2006b) Harm, intent, and the nature of aggressive behavior: measuring naturally occurring aggression in barroom settings. Assessment 13(3):280–296

Gravlee CC, Zenk SN, Woods S, Rowe Z, Schulz AJ (2006) Handheld computers for direct observation of the social and physical environment. Field Methods 18:382–397

Hartmann DP, Wood DD (1982) Observational methods. In: Bellack AS, Hersen M, Kazdin AE (eds) International handbook of behavior modification and therapy. Plenum, New York, pp 109–138

Homel R, Clark J (1994) The prediction and prevention of violence in pubs and clubs. In: Clarke RV (ed) Crime prevention studies, vol 3. Criminal Justice Press, Monsey, NY, pp 1–46

Homel R, Hauritz M, Wortley R, McIlwain G (1997) Preventing alcohol-related crime through community action: The Surfers Paradise Safety Action Project. In: Homel R (ed) Policing for prevention: reducing crime, public intoxication and injury. Criminal Justice Press, Monsey, NY, pp 35–90

Kazdin AE (1979) Unobtrusive measures in behavioral assessment. J Appl Behav Anal 12:713–724

Kelling GL (2001) "Broken windows" vs. "A new look." Law Enforcement News, March 15, 9

Knutsson J (1997) Restoring public order in a city park. In: Homel R (ed) Policing for prevention: reducing crime, public intoxication and injury. Criminal Justice Press, Monsey, NY, pp 133–151

Langan PA, Greenfeld LA, Smith SK, Dunrose MR, Levin DJ (2001) Contacts between police and the public, findings from the 1999 national survey. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC

Levine R (1997) A geography of time. Basic Books, New York

Luthans F, Lockwood D (1984) Toward an observation system for measuring leader behavior in natural settings. In: Hunt J, Hosking D, Schriesheim C, Steward R (eds) Leaders and managers: international perspective of managerial behavior and leadership. Pergamon, New York, pp 117–141

Macintyre S, Homel R (1997) Danger on the dance floor: a study of interior design, crowding and aggression in nightclubs. In: Homel R (ed) Policing for prevention: reducing crime, public intoxication and injury. Criminal Justice Press, Monsey, NY, pp 91–113

Mastrofski SD (2004) Controlling street-level police discretion. Ann Am Acad Pol Soc Sci 593:100–118

Mastrofski SD, Parks RB (1990) Improving observational studies of police. Criminology 28(3):475–496

Mastrofski SD, Snipes JB, Supina A (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33:269–305

Mastrofski SD, Parks RB, Reiss AJ Jr, Worden RE, DeJong C, Snipes JB, Terrill W (1998) Systematic observation of public police: applying field research methods to policy issues. NCJ 172859. National Institute of Justice, Washington, DC

Mastrofski SD, Snipes JB, Parks RB, Maxwell CD (2000) The helping hand of the law: police control of citizens on request. Criminology 38:307–342

Mastrofski SD, Reisig M, McCluskey JD (2002) Police disrespect toward the public: an encounter-based analysis. Criminology 40:519–552

Maxfield MG, Babbie E (1998) Research methods for criminal justice and criminology, 2nd edn. West, Belmont, CA

McCall GJ (1984) Systematic field observation. Annu Rev Sociol 10:263–282

McCluskey JD (2003) Police requests for compliance: coercive and procedurally just tactics. LFB Scholarly Publishing LLC, New York

McCluskey JD, Mastrofski SD, Parks RB (1999) To acquiesce or rebel: predicting citizen compliance with police requests. Police Q 2(4):389–416

McIver JP, Parks RB (1983) Evaluating police performance: identification of effective and ineffective police actions. In: Bennett RR (ed) Police at work: policy issues and analysis. Sage, Beverly Hills, pp 21–44

Meehan AJ, Ponder MC (2002) Race and place: the ecology of racial profiling African American motorists. Justice Q 19(3):399–430

Mileski M (1971) Courtroom encounters: an observational study of a lower criminal court. Law Soc Rev 5:473–538

Mintzberg H (1972) The nature of managerial work. Harper Collins, New York

O'Brien RO (2000) Crime facts: victim and offender data. In: Sheley JF (ed) Criminology. Wadsworth, Belmont, CA, pp 59–83

Parks RB (1984) Comparing citizen and observer perceptions of police-citizen encounters In: Whitaker GP (ed) Understanding police agency performance. U.S. Government Printing office, Washington, DC, pp 121–135

Parks RB, Mastrofski SD, Reiss AJ, Worden RE, Terrill WC, DeJong C, Stroshine M, Shepherd R (1998) St. Petersburg project on policing neighborhoods: a study of the police and the community. Indiana University, Bloomington, IN

Percy S (1980) Response time and citizen evaluation of police. J Police Sci Adm 8(1):75–86

Perkins DD, Wandersman A, Rich RC, Taylor RB (1993) The physical environment of street crime: defensible space, territoriality and incivilities. *J Environ Psychol* 13:29–49

Pugh SD (2001) Service with a smile: emotional contagion in the service encounter. Acad Manage J 44(5):1018–1027

Rafaeli A, Sutton RI (1989) The expression of emotion in organizational life. In: Cummings LL, Staw BM (eds) Research in organizational behavior. JAI Press, Greenwich, CT, pp 1–42

Rafaeli A, Sutton RI (1990) Busy stores and demanding customers: how do they affect the display of positive emotion? Acad Manage J 33:623–637

Raudenbush SW, Sampson RJ (1999) Toward a science of assessing ecological settings, with applications to the systematic social observation of neighborhoods. Sociol Methodol 29:1–41

Reisig M, McCluskey JD, Mastrofski SD, Terrill W (2004) Suspect disrespect toward the police. Justice Q 21: 241–268

Reiss AJ Jr (1968) Stuff and nonsense about social surveys and observation. In: Becker HS, Geer B, Riesman D, Weiss RS (eds) Institutions and the person: papers presented to C. Everett Hughes. Aldine Publishing Company, Chicago, pp 351–367

Reiss AJ Jr (1971a) The police and the public. Yale University Press, New Haven, CT

Reiss AJ Jr (1971b) Systematic observation of natural social phenomena. In: Costner HL (ed) Sociological methodology. Jossey-Bass, Inc, San Francisco, pp 3–33

Roethlisberger FJ, Dickson WJ (1939) Management and the worker. Harvard University Press, Cambridge, MA

Rosenshine B, Furst NF (1973) The use of direct observation to study teaching. In: Gage NL (ed) Handbook of research on teaching, 2nd edn. Rand-McNally, Chicago, pp 122–183

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 3:603–651

Sampson RJ, Raudenbush SW (2004) Seeing disorder: neighborhood stigma and the social construction of "broken windows." Soc Psychol Q 67(4):319–342

Schafer J, Mastrofski S (2005) Police leniency in traffic enforcement encounters: exploratory findings from observations and interviews. J Crim Justice 33:225–238

Scott EJ (1981) Calls for service – Citizen demand and initial police response. NCJ 078362. Workshop in Political Theory and Policy Analysis, Bloomington, IN

Sherman LW, Weisburd D (1995) General deterrence effects of police patrol in crime "hot spots": a randomized, controlled evaluation. Justice Q 12(4):625–648

Shrum W, Duque R, Brown T (2005) Digital video as research practice: methodology for the millennium. J Res Pract 1(1):1–19. http://jrp.icaap.org/index.php/jrp/article/view/6/11

Skogan WG (2006) Police and community in Chicago: a tale of three cities. Oxford University Press, New York

Skogan W, Frydl K (eds) (2004) Fairness and effectiveness in policing: the evidence. The National Academies Press, Washington, DC

Skolnick JH (1966) Justice without trial: law enforcement in democratic society. Wiley, New York

Smith DA (1984) The organizational context of legal control. Criminology 22(1):19–38

Smith DA (1986) The neighborhood context of police behavior. In: Reiss A, Tonry M (eds) Crime and justice: a review of research. University of Chicago Press, Chicago, pp 314–332

Smith RL, Mc Phail C, Pickens RG (1975) Reactivity to systematic observation with film: a field experiment. Sociometry 38:536–550

Smith WR, Tomaskovic-Devey D, Zingraff MT, Mason HM, Warren PY, Wright CP (2004) The North Carolina highway traffic study: appendix A. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, Washington, DC. http://www.ncjrs.gov/app/Publications/Abstract.aspx?ID=204021 Accessed January 31, 2009

Snipes J, Mastrofski SD (1990) An empirical test of Muir's typology of police officers. Am J Crim Justice 14(4): 268–296

Spano R (2005) Potential sources of observer bias in police observational data. Soc Sci Res 34:591–617

Spano R (2007) How does reactivity affect police behavior? Describing and quantifying the impact of reactivity as behavioral change in a large-scale observational study of police. J Crim Justice 35:453–465

Stange KC, Zyzanski SJ, Smith TF, Kelly R, Langa D, Flock SA, Jaen CR (1998) How valid are medical records and patient questionnaires for physician profiling and health services research?: a comparison with direct observation of patient visits. Med Care 36(6):851–867

Sykes RE (1977) Techniques of data collection and reduction in systematic field observation. Behav Res Methods Instrum 9:404–17

Sykes RE (1978) Toward a theory of observer effect in systematic field observation. Hum Organ 37:148–56

Sykes RE, Brent EE (1983) Policing: a social behaviorist perspective. Rutgers University Press, New Brunswick, NJ

Sykes RE, Whitney F (1969) Systematic observation utilizing the Minnesota interaction data coding and reduction system. Behav Sci 14:167–169

Taylor RB (1997) Social order and disorder of street blocks and neighbor-hoods: ecology, microecology, and the systemic model of social disorganization. J Res Crime Delinq 34:113–155

Taylor RB (2001) Crime, grime, fear, and decline. Westview Press, New York

Taylor RB, Gottfredson SD, Brower S (1984) Block crime and fear: defensible space, local social ties, and territorial functioning. J Res Crime Delinq 21:303–331

Taylor RB, Shumaker S, Gottfredson SD (1985) Neighborhood-level links between physical features and local sentiments: deterioration of fear of crime and confidence. J Archit Plann Res 21:261–265

Terrill W (2001) Police coercion: application of the force continuum. LFB Scholarly Publishing, LLC, New York

Terrill W (2003) Police use of force and suspect resistance: the micro process of the police-suspect encounter. Police Q 6(1):51–83

Terrill W, Mastrofski SD (2002) Reassessing situational and officer based determinants of police coercion. Justice Q 19:215–248

Terrill W, Paoline EA III, Manning PK (2003) Police culture and coercion. Criminology 41(4):1003–1034

Thomas DS, Loomis AM, Arrington RE (1933) Observational studies of social behavior. Institute of Human Relations, Yale University, New Haven

Tyler TR, Huo YJ (2002) Trust in the law: encouraging public cooperation with the police and courts. Russell Sage, New York

Uchida CD, Maguire ER, Solomon SE, Gantley M (2004) Safe kids, safe schools: evaluating the use of iris recognition technology in New Egypt, NJ. 21st Century Solutions, Inc, Silver Spring, MD

Van Maanen J (1974) Working the street: a developmental view of police behavior. In: Jacob H (ed) The potential for reform in criminal justice. Sage, Thousand Oaks, CA, pp 83–130

Webb EJ, Campbell DT, Schwartz RD, Sechrest L, Grove JB (1981) Nonreactive measures in the social sciences. Houghton Mifflin, Boston

Weick KE (1968) Systematic observational methods. In: Lindsey G, Aronson E (eds) The handbook of social psychology, 2nd edn, vol 2. Addison-Wesley, Reading, MA, pp 357–451

Weisburd D (2008) Place-based policing. Ideas in Policing. Police Foundation, Washington, DC

Weisburd D, Wyckoff LA, Ready J, Eck JE, Hinkle JC, Gajewski F (2006) Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44(3):549–592

Whitaker GP, Mastrofski SD, Ostrom E, Parks RB, Percy SL (1982) Basic issues in police performance. National Institute of Justice, Washington, DC

Whitehead TN (1935) Social relationships in the factory: a study of an industrial group. Hum Factor 9:381–394

Wilcox P, Augustine MC, Clayton RR (2006) Physical environment and crime and misconduct in Kentucky schools. J Prim Prev 27(3):293–313

Wilson D, Mastrofski S (2004) Final report: evaluation of the Biometric Inmate Tracking System at the Consolidated Navy Brig, Charleston, South Carolina. Report to the Space and Naval Warfare Systems Center. Center for Justice Leadership and Management, Manassas, VA

Worden RE (1996) The causes of police brutality: theory and evidence on police use of force. In: Geller WA, Toch H (eds) Police violence: understanding and controlling police use of force. Yale University Press, New Haven, pp 23–51