DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY

Translating Causal Claims: Principles and Strategies for Policy-Relevant Criminology

Citation
Sampson, Robert J., Christopher Winship, and Carly Knight. 2013. “Translating Causal Claims: Principles and Strategies for Policy-Relevant Criminology.” Criminology & Public Policy 12, no. 4: 587–616.

Published version
https://doi.org/10.1111/1745-9133.12027

Link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:12967679

Terms of use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles (OAP), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility
https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story
The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

Overview of: "Translating Causal Claims
Principles and Strategies for Policy-Relevant Criminology"

Robert J. Sampson

Christopher Winship

Carly Knight

Harvard University

Research Summary

This article reviews the causal turn in the social sciences and accompanying efforts by criminologists to make policy claims more credible. Although there has been much progress in techniques for the estimation of causal effects, we find that the link between evidence and valid policy implications remains elusive. Drawing on criminological theory and research insights from disciplines such as sociology, economics, and statistics, we assess principles and strategies for informing policy in a causally uncertain world. We identify three distinct domains of inquiry that form a part of the translational process from evidence to policy and that complicate the straightforward exportation of causal effects to policy recommendations: (a) mechanisms and causal pathways, (b) effect heterogeneity, and (c) contextualization. We elaborate these three concepts by examining research on broken windows theory, policing, video games and violence, the Moving to Opportunity voucher experiment, incarceration, and especially the rich set of experimental studies on domestic violence that originated in Minneapolis, MN in the early 1980s. We also articulate a set of conceptual tools for advancing the goal of policy translation and offer recommendations for how what we call "policy graphs"—causal graphs used to

Direct correspondence to Robert J. Sampson, Department of Sociology, William James Hall, Harvard University, 33 Kirkland St., Cambridge, MA 02138 (e-mail: rsampson@wjh.harvard.edu).

analyze the policy implications of a system of causal relations—can potentially integrate the theoretical and policy arms of criminology.

Policy Implications

Evidence, even if causal, does not necessarily inform policy. In fact, the question of "what works," the focus of the growing evidence-based movement in criminology, turns out to be a different question than, "what will work?" Evidence-based policy research must therefore be concerned with much more than providing policymakers with research on causal effects, however precisely measured. The implication is that we must separate criminology's increasing focus on causality from its policy turn and formally recognize that the latter requires a different standard of theory and evidence than does the former. In particular, criminologists interested in making policy claims must ask hard questions about the potential mechanisms through which a treatment influences an outcome, heterogeneous effects across people and time, contextual variations, and all of the real-world phenomena to which these challenges give rise—such as unintended consequences, policies that change incentive and opportunity structures, and the scale at which policies change in meaning. Theoretically guided causal graphs enhance this goal and help inform policy in a causally uncertain world. Translational criminology is ultimately a process that entails the constant interplay of theory, research, and practice.

Keywords

causality, mechanisms, effect heterogeneity, policy graphs

RESEARCH ARTICLE

Translating Causal Claims

Principles and Strategies for Policy-Relevant Criminology

Robert J. Sampson

Christopher Winship

Carly Knight

Harvard University

Criminology has deliberately and increasingly turned its attention to influencing public policy. The founding of *Criminology & Public Policy* as an official publication of the American Society of Criminology is a major recognition of this trend. At the same time, criminology and the social sciences at large have undergone what some have termed a "causal revolution," a movement characterized by increased attention to, and higher standards for, causal claims. These two trends are controversial and seemingly in conflict with each other. As Blomberg, Mestre, and Mann (2013) noted in their introduction to this special issue, many criminologists eschew making policy recommendations or "public" claims (Tittle, 2004). In a field characterized by uncertain knowledge—especially on the root causes of crime—and contested claims for which very few stylized facts are agreed upon, strong policy advice may be premature (Manski, 2013; Rein and Winship, 1999).

We agree with those criminologists who see a social world characterized by contingency, and yet we support criminology's policy turn. This article seeks to resolve this tension by

assessing principles and strategies for informing policy in a causally uncertain world. Our claim for a new mode of "translational criminology" (Laub, 2012) does not double down on the standards for establishing causality, the current trend. To accommodate policy-relevant research, for example, many criminologists have shifted their object of study to proximate and malleable factors that are amenable to intervention by the state, and they have made great strides in developing experimental approaches and technical tools for the identification of causal effects. In a causally complex world, however, policy research requires more than the estimation of causal effects, even if precisely and well identified. Rather, it requires system-level knowledge of how policy is expected to work within a larger social context. Methodological fine-tuning and even technical certainty, we argue, cannot substitute for theory, substantive knowledge, and attention to context.

To translate criminological findings into policy recommendations instead requires a set of strategies that move us beyond the narrow confines of causal identification. In this article, we identify three distinct topics or domains of inquiry that must be part of the translational process and that complicate the straightforward exportation of causal effects to policy: (a) *mechanisms and causal pathways,* (b*) effect heterogeneity*, and (c) *contextualization.* We elaborate each of these in turn, accompanied by conceptual tools for advancing the goal of policy translation. We specifically offer recommendations for how what we call "policy graphs"—causal graphs used to analyze the policy implications of a system of causal relations—can potentially integrate the theoretical and policy arms of criminology.

The Shift in Causal Standards

Over the past several decades, the social sciences have been swept by an interest in causality and in developing a technical apparatus with which to identify it. The very language of causality—from selection bias to endogeneity concerns—has spilled over the borders of experimental research to become part of a common social scientific lexicon. The sources of the causal revolution are complex, but the disciplines of economics and statistics are major drivers that have integrated an epistemological agenda with powerful methodological tools. In economics, Heckman (2005) has drawn on a venerable tradition of structural equation modeling and on the analysis of alternative courses of economic action to put forth what he terms a "scientific model of causality" (to which we return in the subsequent discussion). In statistics, Donald Rubin and others (Holland, 1986; Rubin, 1974; Rubin, 1978), building on the foundational work of Neyman (1935; 1990 [1923]), have articulated the counterfactual model of causal inference (also referred to as the "potential outcome" model) that supplies an exact definition of a causal effect with implications for how it should be identified. Here, the fundamental problem of causal inference is that for any given unit of treatment we cannot actually observe that unit's counterfactual outcome—that is, what would have happened if it did or did not receive the treatment. Experiments are able to overcome this problem through the power of randomization. By randomizing a treatment and averaging over observations, experiments provide an unbiased estimate of the *average causal effect*.

It is not difficult to understand, then, how experiments have won the title of the "gold standard" of empirical research. Compared with experiments, methods based on observational data alone are no longer considered by many social scientists with the same level of confidence in estimating causal parameters. The result is that evidence-based policy has largely become

equated with evidence from randomized controlled trials (RCTs).[1] Because experiments often are not possible, however, a litany of quasi-experimental methods (e.g., the use of instrumental variables) and sophisticated modeling (e.g., matching and propensity scores) have been developed that attempt to mimic the classic experimental design. The result of these moves is that empirical standards of evidence today are as high as they have ever been, and concerns with identification of causal estimates now animate much of social science methodology.

Criminology and Policy Today

It is within this general social scientific context that criminology has undergone its own causal turn. Criminology has historically been focused on "backward-looking causality" or what often is termed "the causes of an effect." In the case of the causes of crime, classic criminological subjects such as poverty or subcultural values are typically considered root causes. Yet the turn toward causality and policy has pushed much of criminology away from this kind of focus. As famously maligned by James Q. Wilson in *Thinking About Crime* (1975), root causes are not only steeped in causal uncertainty but also, from the perspective of policy, may be irrelevant. Rather than investigate causes of crime that governments (at least in criminal justice) are generally powerless to change, Wilson argued that criminology should seek solutions elsewhere and in essence turn its back on theory.

The rise of the counterfactual paradigm, coupled with policy demands, has pushed research toward trying to identify possible interventions the government would be better equipped to undertake. Policy-based criminology has thus largely adopted a "forward-looking"

[1]See the discussion in Cartwright and Hardie (2012) and Ludwig, Kling, and Mullainathan(2011); see also coalition4evidence.org/.

approach—"the effects of a cause".[2] Root causes have been replaced with a focus on treatments, say, policing. The question has shifted from "what causes crime" to "did a program work?" In addition, causal standards have been raised throughout the field, such that it is now common in the criminological literature to see the use of propensity scoring and instrumental variable approaches in warranting causal claims, even for classic questions on root causes.

Many aspects of this turn are clearly salutary. Identifying treatments that "work" is no easy matter, and causal clarity can hardly be considered a bad thing. But increasingly, it seems, the causal turn and the policy turn have led to the posing of increasingly narrow questions and to the conflation of "what policy works" to the issue of "did the treatment have a causal effect?" This emphasis is perhaps best indicated by the Department of Justice's new "clearinghouse" website for the assessment of existing research–"crime solutions" (see crimesolutions.gov/about.aspx). The idea is to offer policy makers guidance on what to do based on prior research that has been deemed by review panels to meet rigorous standards of causal evidence. As stated on the website, "crimesolutions.gov uses rigorous research to inform practitioners and policy makers about *what works* in criminal justice, juvenile justice, and crime victim services" (accessed May 23, 2013, emphasis in original). Individual studies are ranked, with randomized experiments getting the highest rating in terms of ensuring the internal validity of results.

Yet internal validity and what works in terms of policy are two separate, and only loosely connected, questions. No causal estimate, however precise, is the same as a policy prescription for "what *will* work." In the next section, we will expand on what we believe are additional and

[2]Statistics has largely abandoned the causes of effects, as Holland (1986) proposed.

necessary questions that must be taken into account when translating empirical results into policy recommendations. Establishing internal validity is only a first step.

Lost in Translation? From Causal Claims to Policy Intervention

Despite the obvious importance of experimental and quasi-experimental evidence in causal analysis, some scholars, most notably the economist James Heckman, have pointed to their underappreciated limitations for policy analysis (Heckman, 2005, 2008; Heckman and Smith, 1995). Heckman (2008: 4–5) argued that three types of questions are involved in policy analysis:

> (P1) Evaluating the effects of historical interventions on outcomes, including their impact on the treated and society at large.
>
> (P2) Forecasting the effects (constructing counterfactual states) of interventions implemented in one environment in other environments.
>
> (P3) Forecasting the effects of interventions (constructing counterfactual states associated with interventions) never historically experienced to various environments.

Ideally, experiments can inform the answers to the first question (P1). According to Heckman (2008), theory and ultimately structural equation modeling are needed to provide answers to the second (P2) and third (P3) questions. Heckman's argument looms large when we consider that in most policy contexts, questions P2 and P3 are what really matter—that is, what will happen in

contexts different from where an experiment has occurred and where the intervention differs in important ways from that carried out in the experiment.

Similar to Heckman (2008), we argue that there is a large gulf between the kinds of information causal analysis typically provides and the kind of information that well-informed policy demands. To traverse this gap, we advance three broad topics that must form a part of the translational process from experimental results to policy recommendations: (a) the identification of mechanisms and pathways, (b) effect heterogeneity, and (c) contextual validity. In each case, theory is essential to assessing the importance and possible policy implications of experimental or other causally based evidence.

Consider first the gold standard of an RCT. The experiment establishes the existence of a link between a treatment and an outcome for a particular population in a particular context. Within this circumscribed context, the goal of the experiment is to generate an internally valid estimate. We depict this bare-bones causal model in Figure 1. Establishing this link is no small feat. It is of great importance to policymakers as well as to academics to be able to identify with precision that some programs can reduce recidivism or juvenile offending, or any number of seemingly intractable criminological outcomes.

(Figure 1 about here)

Yet once this link is established, the translational process has only just begun. First, policy requires information about what will happen in *different* contexts, the very thing for which experiments make no claim to be able to estimate. Although the problem of external validity is well known in the social sciences, it has not been fully confronted by criminologists. It is

remarkable, for example, that whereas the website of crimesolutions.gov contains clear descriptions of causal evidence and internal validity, there is currently no entry whatsoever for external validity.[3] Causal evidence is instead defined in terms of "evidence that documents a relationship between an activity, treatment, or intervention (including technology) and its intended outcomes, including measuring the direction and size of a change, and the extent to which a change may be attributed to the activity or intervention." This definition may nicely fit the requirements for causal evidence for internal validation. However, causal evidence for *policy recommendations* outside this context should be held to a more demanding evidentiary—and theoretical—standard.

Second, and more fundamentally, in most cases a policy is not a treatment. Thus, to recommend policy requires more than considering how a treatment would be expected to work across diverse locales. When one considers policy not as a randomized trial but as a change in institutional structure, it becomes clear that theory must be brought to bear for prediction. A policy is, by definition, a change in the rules of the game. As a result, "policy translation" involves both the problem of what happens when "C" in Figure 1 changes *and* the problem of accounting for changes in organizational, political, or wider social structure when the treatment in Figure 1 scales up into official policy.

The point merits repeating: Even the most internally valid RCT, one that provides near incontrovertible evidence as to the existence of a link between treatment and effect, can only be uncertainly applied to a formal policy context. This limitation arises with equal force to

[3]Accessed May 23, 2013. The classic treatment of internal and external validity is from Cook and Campbell (1979).

nonexperimental designs.[4] The fundamental disjuncture of evidence and policy application raises the problem of causal *interpretation*. No matter how experiments may be understood by researchers, experimental results often are interpreted by policymakers as a direct test of policy, as evidential support that clinches a conclusion of "what works" (Cartwright, 2007). But the simplicity of the causal graph in Figure 1 is deceptive, such that causal analyses tend to carry with them an implicit exportability claim (Barinboim and Pearl, 2013). Conditionality and constraints fall away, and experiments all too often are misread as a general statement of how the world works. To move beyond these limitations we must consider what is inside the experimental black box—to look at why and how *T* is linked to *Y*.

Strategies for Moving Forward

The call for a holistic and contextual approach to understanding causal relationships may sound intractable. How are we to reconcile complex causality—with multiple pathways, heterogeneous effects, and interdependent systems—with policy recommendations that are useful and manageable? What are the strategies for improving criminological research? Recognition of complexity need not imply nihilism in practice. The purpose of this article is thus not to discourage policy-relevant criminological research or critique experiments but to suggest topics that a translational criminology must address, topics that should feature in any discussion of "what works" or, perhaps more accurately, "what will work." In doing so, we note that experiments are part of the solution; especially those conducted within criminal justice agencies and that test mechanisms (Ludwig et al., 2011). Another part of the solution involves modes of

[4]We focus here on experiments because they are held up as the gold standard of research design and because other methods typically face the limitations of experiments and many additional problems as well.

inquiry that, under the name of causal rigor, often have been branded as inferior: descriptive data and "speculative" theory. We can move forward by remaining cognizant of and theorizing how the relationships identified by experiments fit within a larger social structure.

Although a growing number of tools are used to model complexity, we use a strategy that that can aid in giving proper attention to the sorts of complexity that we identify: *causal graphs*. Causal graphs have, for decades, been used by social scientists to understand systems of causal relationships. They are at the core of path analysis (Duncan, 1966, 1975) and are an important component of structural equation modeling (Bollen, 1989). In the past decade they have gained a renewed importance with Judea Pearl's (2009 [2000]) influential work on directed acyclic graphs (DAGs).[5]For Pearl, the fundamental purpose of DAGs is that they provide a simple but powerful tool that allows for the analysis of the conditions under which an observed association can be identified as a causal effect. Although translation, not identification, is our key concern, we draw from Pearl and others the broader point that causal graphs are a useful way of specifying the theoretical structure of a problem and illustrating the interdependencies in causal systems. From this perspective, making policy requires knowledge of an interlocking system of organizations and actors, demanding a birds-eye view of the wider causal picture.

Because we are not concerned with identification but with how causal relationships fit within wider systems, we do not need the extensive mathematical machinery associated with DAGs.[6] Moreover, we will label causal effects as positive or negative in our various examples,

[5] See Bollen and Pearl (2013) for a thorough discussion of the complementarity and interrelationship between structural equation models and DAGs as modes of representing causal structures.

[6] See Morgan and Winship (2007) for a basic introduction to DAGs and Elwert (2013) for a more through presentation of their workings.

which is something not generally done with DAGs.[7] As our focus is on policy, we will term our causal graphs "policy graphs" and broaden the initial example in Figure 1.

In using causal graphs to understand policy implications, we also revisit the contrast made between forward-looking causation and backward-looking causation. Although useful, this distinction may obscure the fact that *both* kinds of causation must refer to the casual system in which effects are located. That is, to apply forward-looking causation to policy, the constellation of mechanisms surrounding it must be elaborated. To understand how a specific treatment affects an outcome, and thus what the effects of a policy intervention are likely to be, one needs to understand the causal system of which a specific estimated effect is a part. Responsible policy analysis cannot be done theory free (Laub, 2004: 14–19).

In short, the use of causal graphs in policy analysis provides an explicit and concrete way to bring theory into one's analysis or even better, one's research design. Specifically, causal graphs are a way of representing the theoretically derived causal system related to an outcome of interest. They are important both for research design done prior to the collection of data and for the interpretation of empirical findings derived from statistical analyses. As such, we should keep in mind the interaction of various causal processes in the production of an outcome, rather than focusing on the effect of a singular treatment abstracted from its setting. We examine in depth how this strategy might work with respect to the three pragmatic challenges a translational criminology must address: (a) mechanisms and pathways, (b) effect heterogeneity, and (c) context.

Mechanisms and Pathways

[7] Because DAGs are nonparametric, labeling of effects is not appropriate.

Experiments provide a causal graph of the world in which a treatment leads to an outcome: $T \rightarrow Y$. Mechanisms disaggregate this relationship and in doing so provide what Heckman (2005, 2008) would term a scientific understanding of causality. VanderWeele (2009), drawing on work by Aalen and Frigess (2007), has helped us understand this argument by making the distinction between *counterfactual-based causality* and *mechanistic causality*. As VanderWeele (2009: 222) described the distinction:

> Counterfactual-based causality is essentially concerned with the effects of a particular intervention or exposure without regard to the mechanisms by which these effects arise. Conclusions about causal effects are drawn either through randomized trials or through the careful design and analysis of observational data in which the researcher attempts to control for all the variables that confound the exposure-outcome relationship.

This approach is described as ‘‘black box’’ causality because the methods used to estimate causal effects can be valid *irrespective of how the exposure produces its effect*.

Mechanistic causality is different. We quote Tyler again at length (2009: 222):

> Mechanistic causality, on the other hand, attempts to understand the mechanisms governing the various processes which give rise to particular outcomes. Assessing mechanistic causality requires closer observations and a good deal of scientific knowledge. The model for mechanistic causality is the natural sciences in which attempts are made to identify the natural laws and precise workings behind the phenomena we observe. In the mechanistic approach, one attempts to ‘‘look inside the black box.’’ A similar distinction to that made by Aalen and Frigess is also made by Heckman (2005, 2008). Heckman calls the counterfactual-based causality described above ‘‘statistical

> causality.'' He contrasts this with what he calls the ''scientific model of causality'' or ''econometric causality.'' Heckman criticizes the statistical literature on causality for not making use of theory and for not taking into account agent choice, equilibrium processes and feedback which he argues are the mechanisms by which outcomes are generated.

There are three reasons why the theoretical specification and analysis of mechanisms area fundamental part of policy analysis (see also Ludwig et al., 2011; Rosenzweig and Wolpin, 2000). First, mechanisms are necessary for interpretation of what is a cause and what is merely a risk factor in crime (Wikström, 2011). Second, policy generally is concerned with achieving a particular causal *process*, not simply a causal *effect*. Finally, policy efficacy requires considering alternative, cost-effective processes for bringing about the desired outcome; mechanisms can identify these processes.[8]

To elaborate on these points, we begin by considering the "broken windows" theory of crime. Introduced by James Q. Wilson and George L. Kelling in 1982 in the *Atlantic Monthly*, the theory contends that targeting minor forms of disorder—broken windows, loitering, and graffiti—reduces serious crime. Disorder is hypothesized to provide visual cues of the state of neighborhood social control from which potential offenders derive expectations of whether further antisocial behavior will be tolerated or reported. Thus, the theory goes, the appearance of disorder begets further disorder: Broken windows breed an environment conducive to crime.

[8]Increasingly, funders of interventions have come to a similar conclusion and are pushing back against the emphasis on the black box of average causal effects. The President of the William T. Grant Foundation recently wrote: "In today's vernacular, we need more research attention paid to *why* and *under what conditions* things works as the missing ingredients in the 'what works' agenda" (Granger, 2011: 29).

Represented in Figure 2, broken windows (*BW*) provides visual cues (*VC*) of social disorganization that in turn incentivizes crime (*C*):

(Figure 2 about here)

Broken windows theory has proven to be one of the most influential theories in criminology, prompting a wave of "zero-tolerance" and "order-maintenance" policing in New York City, Chicago, and Los Angeles (Duneier, 1999; Harcourt, 2001). For example, New York City Mayor Rudy Giuliani's crackdown on misdemeanor crimes was widely touted as a principal cause of the drop in crime in the 1990s. Even today, the NYC Police Department continues to cite aggressive policing as a decisive factor in the crime drop. However, although there is some empirical support that broken windows policing lowers crime (Kelling and Sousa, 2001), findings have been mixed overall (Harcourt, 1998; Harcourt, 2001; Harcourt and Ludwig, 2006; Sampson and Cohen, 1988). Even the evidentiary basis in Zimbardo's 1969 vandalism experiment, often cited as the foundation of the broken windows thesis, is questionable. In that study, cars were purposively left abandoned in the Bronx, NY, and Palo Alto, CA. In the Bronx, the car was immediately vandalized, whereas in Palo Alto, the car was left untouched for a week. Only after Zimbardo himself smashed in the windows did the Palo Alto car succumb to further destruction. The causal link from visual cues to crime was thus contextually dependent even at the outset (Sampson, 2013: 17).

Mechanisms feature prominently in critiques of broken windows. This brings us to our first point: Mechanisms are a necessary part of the *interpretation* of causal claims (Rosenzweig and Wolpin, 2000). To the extent that we can identify an empirical relationship, mechanisms tell

us why that relationship exists. Consequently, they also warn us when a relationship may be spurious, or caused by some other process than the one generally touted to be at play (Knight and Winship, 2013; Wikström, Oberwittler, Treiber, and Hardie, 2012). For example, Sampson and Raudenbush (1999; 2004) called into question the mechanism linking visual cues to crime according to the original broken windows theory, considering both spurious pathways (common causes) and interpretive processes. In analyzing the social-psychological processes behind the formation of perceptions of disorder, they find only a modest association between *perceived* disorder and *actual* disorder. Rather, stereotypes about the neighborhood, based in large part on its racial composition, have a much larger effect on perceptions. In other words, there is no guarantee that repaired windows will be noticed. Neighborhoods with large minority populations also generate enhanced perceptions of disorder among all race/ethnic groups, suggesting a general *cultural* process, not an unmediated effect of visual cues.

By testing the broken window's hypothesized mechanism, Sampson and Raudenbush (2004) complicated the story. If broken windows policing works and if their findings are correct, then this style of policing may work through a different process than changes in perception of actual disorder. Or, to the extent that broken windows theory is correct, then policies targeting disorder may be least likely to work in some of the most disadvantaged areas with the most entrenched neighborhood stereotypes. To go back to Zimbardo's (1969) experiment, broken windows might be an important signal in some communities while they may be meaningless in others. The wider point that emerges from this observation is that the identification of mechanisms is inextricably tied to interpretation of the causal claim. Answers to when, where, and for whom broken windows policing works are incoherent without recourse to mechanisms.

Our second point builds on the first: Mechanisms are necessary for the *justification* of public policy. In many cases, it is of great normative importance whether the treatment exerts its causal force through one or another mechanism. Policy, in other words, is not always concerned simply with achieving an outcome by manipulating a cause: Rather, it aims to achieve an outcome via *a certain route.* Consider, for example, that there are a number of pathways through which broken windows policing, with its targeting of misdemeanor crimes, could affect the crime rate. As depicted in the enhanced causal graph presented in Figure 3, the standard mechanism proposed by broken windows theory is given by the pathway in which broken windows policing (*BWP*) affects visual cues (*VC*) that in turn affects crime (*C*) ($BWP \rightarrow VC \rightarrow C$). However, as Sampson and Cohen (1988) suggested, this is hardly the only pathway by which policy could act: Broken windows policing may directly increase police presence, raising the likelihood of arrest for serious crimes. This possibility is represented by a pathway from broken windows policing (*BWP*) to arrests for violent crime (*AV*) to crime (*C*) ($BWP \rightarrow AV \rightarrow C$). In this case, what seems to be deterrence would actually be the result of a higher arrest rate for serious crime and subsequent incapacitation, rather than the result of a higher arrest rate of misdemeanors or a change in offender perceptions. Harcourt (1998) discussed a related possibility, in which aggressive policing may lead to the arrest of individuals the police would not have grounds to arrest otherwise, preempting potential offenders through increased surveillance.

(Figure 3 about here)

Which of these pathways is responsible for the observed effect matters for policy because the political attractiveness of a certain intervention may depend on the mechanism being used to

change behavior. Zero-tolerance policing may be politically efficacious if it works by sending a signal to offenders that the neighborhood will not tolerate crime. It is less so, however, if it works through the police indiscriminately abusing the power to arrest. This latter mechanism is doubly pernicious given that police legitimacy would likely be undermined by aggressive and discriminatory behavior (Fagan and Davies, 2000), perhaps leading to increased crime in the long run.

Mechanisms are also an important part of identifying *efficient* policy. Even granting that a study has identified evidence of a cause, without an understanding of the often competing underlying mechanisms, we cannot determine whether that treatment is a more justified policy response than another, potentially effective and cheaper intervention. For example, if zero-tolerance policies promote higher rates of arrest for violent crime and it is these arrests that are doing the causal work, then arresting individuals for relatively minor misdemeanors would be a waste of resources. In other words, causal certainty about the effect of a treatment does not in itself imply certainty that a policy change will yield effects in ways that follow the logic of the intervention.

Despite the importance of mechanisms, they rarely feature in policy recommendations of "what works." For instance, an RCT conducted by Braga and Bond (2008) to test broken windows theory with respect to hot-spot policing was given the highest evidentiary rating by crimesolutions.gov. Yet what qualified as broken windows policing consisted of multiple interventions, including increased misdemeanor arrests, better lighting, providing youth with recreational opportunities, working with local shelters to provide housing for homeless individuals, and connecting problem tenants to mental health services. The latter components were "intended to create opportunities for high-risk individuals to assist police efforts to promote

social order."[9] A benefit of the experimental approach was that particular broken windows policing "package" could be randomized. The drawback, of course, is that it is impossible to know which component of this strategy worked. To our mind, also it is problematic to assume that the purpose of giving social services is to help recipients assist police officers—indeed this assumption is loaded with unexamined theoretical baggage. Even if an increase in social services does promote cooperation with the police, this is a very different mechanism than that typically associated with increased misdemeanor arrests for social disorder.

In sum, the debate over broken windows theory serves to illustrate the basic point that contention over policy rarely concerns only causal identification. Rather, it must grapple with how policy works. In saying this, we are making a stronger point than that criminological research also should be concerned with identifying mechanisms. Going further, we have argued that policy claims often make strong theoretical assumptions, or are themselves unarticulated theories about social processes. In advocating for broken windows policing, policymakers are not simply interested in a causal effect but in a causal *process*. Criminological research, then, should be understood as adjudicating among competing models of social processes.

Heterogeneous Effects

Mechanisms complicate our simple causal model by identifying the various pathways through which a treatment works. Effect heterogeneity, on the other hand, implies that multiple causal models exist that correspond to different subpopulations. Effect heterogeneity occurs when a

[9]See also crimesolutions.gov/ProgramDetails.aspx?ID=208.In another evaluation of "broken windows" policy, the treatment included better technology for detecting crime patterns (commonly known as COMPSTAT)—see crimesolutions.gov/ProgramDetails.aspx?ID=87.Improved technology is far removed from the mechanisms posited in the original broken windows theory.

given treatment has a different causal effect for different individuals or subgroups (e.g. males or females). In extreme cases, a treatment may be beneficial for some individuals, detrimental for others, and have no effect for others still. Medical researchers are increasingly aware of the ubiquitous nature of effect heterogeneity, to the point that a lengthy *New York Times* essay recently posed the provocative question: "Do Clinical Trials Work?" (Leaf, 2013).

This question matters for criminal justice policy too but in a way that goes beyond the specific method of an RCT. At the most basic level, policy is concerned not only with *how* effects are brought about but also with *for whom* they are brought about. Given the centrality of questions concerning quality and fairness to policy, the issue of differential effects must be a primary part of policy analysis. For example, a policy that benefits Whites while harming Blacks is not an example of a policy that most would consider "works" even if its average causal effect was salutary. Moreover, behavioral assumptions implicit in unaccounted heterogeneous effects may cause us to *understate* the effect of a policy. When a random experiment provides a treatment both to those for whom the treatment will give a high return as well as to those for whom the treatment will give a low return, forced treatment leads to a lower average outcome than would be obtained if the treatment were limited to those likely to benefit.

Consider the case of the Moving to Opportunity experiment (Kling, Liebman, and Katz, 2007; Kling, Ludwig, andKatz, 2005) in which families were randomly assigned vouchers to relocate to low-poverty neighborhoods. Subsequent analysis revealed that although the move was beneficial for young females (lowered arrests for property crime better mental health), the move had deleterious effects on young male delinquency (heightened aggressive behavior and probability of arrest for property crime). This sort of effect heterogeneity implies separate graphs for each gender (Figure 4). Here, a move (*M*) facilitated behavioral adjustment (*BA*) for females

while hindering it for males. This distinction is represented by the positive arrow for females ($M \overset{+}{\rightarrow} BA$) and the negative arrow for males $\left(M \overset{-}{\rightarrow} BA\right)$ and is responsible for different probabilities of arrest in each group (*A*). In assessing policy, this differential effect must be a part of the conversation. A single causal effect would mask real differences and all the normative and political implications they entail.

(Figure 4 about here)

Heterogeneity is not just a quality of subgroups. It also is a temporal phenomenon. That is, the efficacy of a treatment depends on its location within the cluster of activities, networks, and institutions that define the actor at any given point. A treatment given to a young man will differ from a treatment given to an old man; a first treatment will differ from a recurring one. Moreover, a treatment at a given age may show an instantaneous effect (or null result) that changes or reverses in a later follow-up. As a consequence, the population-level effects of a policy change, one that individuals may confront daily, will likely diverge from the effects of a single randomized intervention. Our point is not to say that experiments are not useful in estimating treatment effects. Rather, and again, the point is to separate the policy question of what "works" or "will work" from the effects identified by researchers. Policies unfold over a time horizon more expansive than those accommodated by a single empirical study.

Perhaps the best example here is noncriminological but which nonetheless has implications for crime over the life course. The famous Perry Pre-school Project was a randomly assigned treatment that initially boosted IQ in children, but the effect quickly faded. It was only years later that researchers discovered significant treatment effects on lifetime outcomes up to 40

years of age. Accounting for this temporal heterogeneity but also probing the causal mechanisms that produced it, Heckman, Pinto, and Savelyev (in press) showed that the Perry Project significantly enhanced adult outcomes including education, employment, earnings, marriage, participation in healthy behaviors, *and* reduced participation in crime and welfare. They also argued that experimentally induced changes in noncognitive personality traits, rather than IQ, explain a sizable portion of the adult treatment effects. This example shows that the treatment was temporally heterogeneous, had unanticipated spillover effects, and operated through a particular causal pathway—what criminologists would call "self-control." In the next section we provide a more detailed example of how the effect of an intended criminological treatment—arrest for domestic violence—similarly varies over the life course and through mechanisms not anticipated by the original design. [10]

Context

Academics often dutifully note the importance of context. Context is said to provide boundary conditions or limit causal claims. Yet context is more than an unarticulated background or boundary against which to generalize causes and effects. Context is an entrenched causal web that intervenes and shapes every point of an unfolding causal process, dictating the nature of incentives, opportunities, and institutional relationships that define the policy world. As such, policy researchers must rethink their understanding of the role of context, moving it from the periphery to the center of analysis.

[10]In another example, temporal heterogeneity combines with effect heterogeneity. The gender interaction in delinquency uncovered in the interim follow-up of the Moving to Opportunity experiment seems to have eroded in the long-term follow-up (Sanbonmatsu et al., 2011).

In the subsequent discussion, we elaborate three dimensions of context, from most basic to most demanding, that must be kept in mind when translating empirical results into policy. First, and most straightforwardly, context can be understood as the macro-environment that directly affects the actor, whether neighborhood, city, or country. Consider a study on recidivism by Kubrin and Stewart (2006). The authors pointed out, rightly, that most studies tend to examine the individual-level characteristics associated with recidivism, ignoring how the effects of these characteristics may be dependent on local factors, in this instance, place. Using nonexperimental data, they found that neighborhood context accounted for nearly 13% of the variance in recidivism, with offenders returning to disadvantaged communities reoffending more, net of individual-level factors. In another study showing the power of place-based context to influence recidivism, Kirk (2009) employed a quasi-experimental approach that used the residential destruction resulting from Hurricane Katrina as an exogenous source of variation that influences where a parolee will end up once released from prison. He found that a forced move away from a parolee's former geographic home substantially lessens the likelihood of re-incarceration. Such studies thus have suggested that neighborhood context influences the efficacy of criminal justice policies. More generally, context can be understood as the macro-level environment influencing behavior.

Second, and expanding further, context can be understood as the opportunity structure within which an actor exists. Context shapes incentives and limits choice. As such, the context of the actor in the real world often differs markedly from that of the actor in the experimental one. For example, psychological laboratory experiments have established a positive relationship between video game use and violent behavior. Video games are often said to increase aggressive tendencies through a process of social learning that is posited to support a violent personality and

encourage criminal behavior. Yet extrapolating from the lab to the street is a difficult endeavor and ignores the trade-offs between video game playing and other options youths have to occupy their time. Considered against the whole gamut of possible activities, video game playing may not be so socially deleterious. In particular, a recent study by Cunningham, Engelstatter, and Ward (2011) found that playing video games may *incapacitate* violent activity by taking youth off the streets, on the whole, decreasing criminal behavior on aggregate.

The causal graph given in Figure 5a represents the causal pathway indicated by psychology experiments in which video gaming (*VG*) leads to violent behavior (*VB*) through the creation of aggressive tendencies (*AT*), hence, $(VG + AT + VB)$. But when the causal graph is amended to take into account opportunity structure, it is no longer clear whether the positive association identified in the laboratory should hold. As shown in Figure 5b, video game playing (*VG*) may still increase the likelihood of aggressive tendencies (*AT*) and thus increase violent behavior (*VB*) ($VG \overset{+}{\rightarrow} AT \overset{+}{\rightarrow} VB$). However, this criminogenic pathway may be mitigated by a prosocial pathway in which video-game playing lowers opportunities for delinquency and, thus, violent behavior$\left(VG \overset{-}{\rightarrow} OP \overset{+}{\rightarrow} VB\right)$. Given a positive relationship between opportunities for violence and violent behavior, any drop in opportunity for violence should lessen observed violent behavior.

(Figure 5 about here)

What we observe in the causal graphs in Figure 5 is that if the effect of video game along the prosocial pathway ($VG \overset{-}{\rightarrow} OP \overset{+}{\rightarrow} VB$) is stronger than that along the antisocial pathway ($VG \overset{+}{\rightarrow} AT \overset{+}{\rightarrow} VB$), then video-game playing will actually lead to a *decrease* in violent behavior.

This finding recalls the criminological literature on routine activities theory that explains crime not by reference to singular treatments but as the intersection of activities, opportunities, and environment (Cohen and Felson, 1979).

Not least, context concerns the interdependence of institutions and societal responses—a policy intervention in one part of the criminal justice system will have reverberations, quite possibly changing the intended outcome of the intervention. This result is likely to develop in any social context, but it is a particularly salient feature in criminology, where interlocking institutions and interdependent social networks are the norm. As far back as the 1960s, the President's Commission on Law Enforcement and the Administration of Justice introduced the influential concept of the criminal justice "system."

The implication is that feedback loops and unintended causal consequences are theoretically expected. For example, consider America's grand "natural experiment"—mass incarceration. Although incapacitation may reduce violent crime by removing offenders from the community, Sampson (2011) suggested that removal also decreases the ratio of males to females, which in turn increases family disruption and rates of violence. Based on research that has shown that imprisonment has negative effects on employment, especially the marginalization of Black men from the labor market (Western, 2006), Sampson argued that imprisonment may indirectly lead to future crime through its disruptive effect on Black family structure.

We can illustrate this theoretical model in Figure 6. Figure 6a represents the simple relationship among arrest (*A*), incapacitation (*I*), and crime (*C*) ($A \overset{+}{\rightarrow} I \overset{-}{\rightarrow} C$). If Sampson's (2011) contention is true, then estimating only this simple pathway is incomplete without knowledge of the path from incarceration to increased family disruption and, hence, increased violence. The alternative pathway of incarceration's effect through removal (*R*) is represented in

the causal graph of Figure 6b ($A \overset{+}{\rightarrow} R \overset{+}{\rightarrow} C$). This and other potential pathways are especially salient when interpreting a randomized experiment on enhanced policing. An experiment that randomizes by neighborhood or block will capture the causal pathway from arrest to lower crime via incapacitation. But it is less likely to capture the effect of increased incarceration on employment opportunities and family structure, which operates in a temporally and geographically broader context (Western, 2006). Yet the overall effect of increased arrest for a community will depend on the balance between the two pathways, a calculation of which informed policy must assess. Unlike heterogeneous treatment effects and mechanisms, these differences cannot be modeled within an experiment alone, and they require recourse to theory to estimate their effects.

(Figure 6 about here)

In short, contextualism challenges a framework that assumes that we can manipulate a treatment, *ceteris paribus*—that we can isolate an intervention that is exogenous to the system and assume that incentive structures or practices among individuals and organizations will not change. A concern for modeling context also points to the importance of *replications*, *meta-analyses*, and *large-scale observational studies* specifically geared toward investigating macro-level factors.

Putting it All Together: The Minneapolis Domestic Violence Experiment and Follow-ups

Up to this point, we have discussed three complexities to the bare-bones RCT causal model of the world: mechanisms, causal heterogeneity, and context. These subjects, we argue, are as necessary to policy considerations or recommendations as are precise causal estimates. Yet they typically do not feature in criminological conversations about policy that works and, when they do, not in a formal or systematic way. We thus turn to a final extended example to illustrate further our argument and provide guidelines. We chose the Minneapolis (MN) Domestic Violence Experiment (hereafter MDVE) and a group of further studies inspired by the original MDVE as a case of experimental criminology that was ultimately "done right." Through experiments, replications, observational studies, and importantly, criminological theory, the MDVE and its aftermath accumulated information about how mechanisms, heterogeneity, and the importance of context would shape policy on the ground.

Prior to the MDVE, law enforcement had been reluctant to arrest or intervene in cases of domestic violence. This changed during the 1980s after *Thurman vs. City of Torrington* (1984) established police liability in the case of domestic assault, thereby incentivizing states to enact policies that eliminated officers' discretion in cases of domestic abuse. By randomizing police response, the MDVE found that the arrest of an alleged offender caused a marked drop in rates of reoffending (as compared with counseling or separation from the partner). Seizing on the seemingly strong results, legislators, policy experts, and academics began supporting mandatory arrest policies (Mignon and Holmes, 1995).

Although the original researchers repeatedly cautioned against premature extrapolation of the experimental results, the findings drew national attention and quickly ushered in rapid and perhaps unprecedented change in how states and cities administered police responses to domestic violence. Protestations to the limitations of external validity notwithstanding, 24 states adopted

mandatory arrest policies (Miller, 2005), an illustration of the power of the experiments' implicit "exportability" claim. Indeed, the MDVE often has been cited as evidence that mandatory arrest, as a general policy instrument for crimes other than domestic violence, is an effective tool (Davis, 2008).

It is useful, however, to go back to the seminal *American Sociological Review* paper in 1984 that reported the results of the MDVE (Sherman and Berk, 1984).[11] The experiment was presented as a test between two competing theories of the effect of punishment. On the one hand, deterrence theory suggests that punishment will lower recidivism, especially when punishment is certain, swift, and severe. On the other hand, labeling theory predicts that arrest may be criminogenic (Becker, 1963; Lemert, 1951). By randomly assigning arrest, the MDVE sought to adjudicate the effect of arrest. In the end, labeling theory was not supported. At least for the particular population under study, the causal relationship was represented by the simple causal graph in Figure 7: Randomized arrest (*A*) increases deterrence (*D*) [i.e., ($A \xrightarrow{+} D$), which in turn lessens the probability of repeated violence (*RV*), or $D \xrightarrow{-} RV$.

(Figure 7 about here)

Yet over the course of further research, this simple causal graph began to fray. In the 6 years after MDVE, the National Institute of Justice funded five replication studies, whose results varied widely. On the one hand, studies in Omaha, NB; Milwaukee, WI; and Charlotte, NC, not only found no evidence for the deterrent effect of arrest, but also they reported *increases* in

[11]This article is perhaps the most cited criminology paper (with more than 1,100 citations) to appear in the *American Sociological Review* in modern times. A policy brief also was published around the same time as the *ASR* paper.

subsequent crimes. Colorado Springs, CO, and Dade County, FL, on the other hand, did find evidence of deterrence.

Our framework suggests that this kind of difference can be expressed by widening the causal graph. As we discussed with regard to mechanisms and effect heterogeneity, an average causal effect may mask important differences in the routes through which, and the groups for which, an outcome is realized. Analyzing the data from the replication studies 8 years after the original MDVE article was published, Sherman, Smith, Schmidt, and Rogan(1992) offered a possible explanation for these diverse findings: differences in how subsamples of individuals responded to arrest as a result of the operation of two possible mechanisms. They argued that although arrest serves as a formal sanction, arrest also is linked to informal sanctions that decrease the likelihood of reoffending for a subset of the population. In particular, individuals who were more socially and institutionally embedded (employed or married) were hypothesized to experience more informal sanctions and greater social controls from their partners and social networks (Sampson and Laub, 1993). Arrest was thus arguably more effective in this group, whereas for men with fewer social bonds, arrest lost its crime-reducing sting.

Put in the language of this article, differences in the domestic violence experiments could be explained by a combination of the heterogeneity of effects and differences in mechanisms. From this perspective, the causal arrow that represents "deterrence" in the simple causal graph from Figure 7 must be further broken down. In Figure 8, we demonstrate how the effect of arrest works through two separate pathways, one involving formal sanctions ($A \overset{+}{\rightarrow} FS \overset{-}{\rightarrow} RV$) and the other an effect of informal sanctions on socially embedded men ($A \overset{+}{\rightarrow} IS \overset{-}{\rightarrow} RV$), both of which lead to decreases in violence (*RV*). We can then hypothesize that the efficacy of each pathway may well vary with individual characteristics of the offender.

(Figure 8 about here)

Reexamining the data and consistent with Figure 8, Sherman et al. (1992) found that offenders with a greater number of social ties and thus greater "stakes in conformity" were less likely to reoffend than those missing such ties. To the extent that "the effectiveness of legal sanctions rests on a foundation of informal control" (p. 688), differences in a city's economic situation and strength of individual ties are predicted to result in different directions of the effect of arrest. That these results are dependent on the degree of "social bonds" suggested a different causal graph from the one posited in the initial experiment. Similar heterogeneous effects of arrest were found in a reanalysis by Berk, Campbell, Klap, and Western (1992). In a subsequent work, Sherman (1993) offered a theory to account explicitly for how arrest "either reduces, increases, or has no effect on future crimes, depending on the type of offenders, offenses, social settings, and levels of analysis" (p. 445). His theory of "defiance" helps explain the conditions under which punishment, or in this case arrest for domestic violence, increases crime.[12]

Within our framework, policy analysis also must be cognizant of temporal heterogeneity. Indeed, a recent 23-year follow-up study by Sherman and Harris (2013) investigated the effect of arrest from the Milwaukee domestic violence experiment. In what is probably the longest follow-up ever of a randomized trial testing the effect of criminal sanctions, they found that arrest had no effect for employed individuals and actually *increased* the prevalence of reoffending for

[12]Sherman (in press) also noted the widespread heterogeneity in estimates derived from "hot-spot" policing research—"While the average effect is beneficial, the range of effects is very great. Whether another agency implementing some form of hot spot policing will achieve a large or small effect remains highly uncertain from the available research."

unemployed individuals. Crucially, the harmful effect of arrest was only apparent after 6 years and continued growing for the next 20 years. Much like the Heckman et al. (in press) article cited earlier on the Perry Preschool Project, the authors conclude that short-term evaluations are inappropriate for understanding longer life-course outcomes. We would generalize further and emphasize the logical conclusion, often overlooked in criminological discussion of policy implications, that any single empirical result must not be conflated with having predicted the long-term outcomes of a policy regime change.

Scaling Up and the Importance of Context

> "If I called the police to get him out of my house, I'd get evicted." Victim of domestic violence in Milwaukee on reluctance to call 911 (quoted in Eckholm, 2013: A1)

Even with no effect heterogeneity and full knowledge of the mechanisms operating within a particular study, the context challenge implies that a single experiment cannot provide evidence of the consequences of scaling up. Although this challenge might not matter in medical trials, the canonical example of an experimental science, crime, and criminal justice are quintessentially social phenomena. In regard to the MDVE, the treatment was randomly assigned arrest in a study of 314 individuals in a city of nearly 400,000, not an alternative policy universe in which arrest was mandatory. Yet there are many reasons to assume that scaling up will change the nature of the intervention, altering both offenders' understanding of the likely consequences and, importantly, also the victim's likelihood of reporting.

Using nonexperimental methods, Iyengar (2007) has found evidence that suggests that the rise in mandatory arrest laws may be associated with a greater probability of spousal homicide,

an effect mediated by the lesser probabilities that victims will call the police when it is known that their partner will be arrested. As Iyengar stated, the MDVE's use of randomized assignment provided no evidence as to whether mandatory arrest would lower recidivism under a policy regime change of larger scale. Once it was known that a victim calling in would unequivocally result in the offender's arrest, there was a marked drop in reports of domestic violence. This drop was associated with an increase in spousal homicides. Taking into account the effects of scaling up makes clear *policy itself* also must be a part of our causal graph of the world (Figure 9). By doing so, one creates what one might call a "policy graph"—a causal graph in which the policy itself is a variable. This move responds to Heckman's (2005) call for policy research to take into account agent choice and feedback processes based on expectations.

By making the arrest policy (*AP*) part of the graph in Figure 9, we draw attention to two important causal processes that link the policy of mandatory arrest to actual instances of arrest: (a) the perceived likelihood of sanctions (PLS), here on the part of the offender that his or her partner will report domestic abuse, and (b) the likelihood that the partner will make the call reporting domestic abuse (*C*). First, arrest policy (*AP*) may lessen the likelihood that the offender believes the partner will call the police (*PLS*). Second, arrest policy may lessen the actual probability that the victim will report abuse (negative arrow between *AP* and *C*). To the extent that these two mechanisms are in play, we could then witness an overall *decrease* in calls (*C*), while witnessing an *increase* in domestic violence (*DV*). Even if we assume that an increase in domestic violence will be associated generally with an increase in reports of abuse (positive between *DV* and *C*), the overall outcome will depend on the strength of the negative relationship between arrest policy and calls.

For example, if it is certain that the potential offender would be arrested if a call were made, that offender may not believe his or her partner would take such an action, especially if an arrest would lead to a loss of household income or even eviction for victims who are also tenants. In many cities including Milwaukee, the site of one of the domestic violence experiments, landlords can evict tenants who frequently call the police or house criminals. Based on so-called "crime-free housing" ordinances, the idea is to put responsibility on landlords to weed out disruptive tenants (Eckholm, 2013). Although 911 calls may in themselves be defined as disruptive, arrest is more so and leads to a criminal record of someone in or associated with the household. As the victim quoted at the beginning of this section reveals, these ordinances can thus dampen the willingness of citizens to call on the police for help, in turn leading to unintended consequences. In fact, Desmond and Valdez (2013: 117) found that a third of all evictions in Milwaukee involve domestic violence. Desmond (2012: 88) reaches a strong conclusion overall: "In poor black neighborhoods, eviction is to women what incarceration is to men: a typical but severely consequential occurrence contributing to the reproduction of urban poverty."

It stands to reason that in addition to potential monetary losses (e.g., child support) there are real incentives for victims to *not* call the police under mandatory arrest regimes. It is further reasonable to assume that potential offenders are aware of these incentives and potential effects on victims' behavior. With respect to our model, if potential offenders do not believe there will be a cost associated with their actions, then a decrease in perceived likelihood of calling may serve to increase domestic violence (negative arrow between *PLS* and *DV*). Therefore the behavior of both victims and offenders is potentially altered in previously unanticipated and possibly countervailing ways once domestic violence policies are scaled up.

(Figure 9 about here)

Perhaps no other result provides a more powerful reminder of the importance of considering the potential unintended consequences of moving from experiment to policy. In the case of the MDVE, it is clear that the policy intervention, and its unintended effects on the incentives and opportunities available to individuals within the new policy regime, must be theorized as part of the translation from experiment to policy. Noting the difference between the causal graph implicit in the initial experiment (Figure 7) and the processes graphed in Figures 8 and 9, we draw a combined causal graph in Figure 10. A result of the accumulation of replication, theory, and observational data, Figure 10 underscores that the relationship identified by the experiment—that between arrest and reoffending—constitutes just one part of an interlocking causal web. The effect of arrest policy (*AP*) is necessarily different from that of randomized arrest (*A*). It is conditioned by how arrest policy alters the legal structure in which offenders and victims act, by the mechanisms that impact how offenders react to arrest, and by the differences among individuals and their local contexts.

(Figure 10 about here)

To summarize, the results of the initial MDVE must be understood as part of a larger research program. The first, single experiment provided evidence of a time-bound relationship between arrest and rates of reoffending. Yet, as research about the boundary conditions of this effect mounted, the contingencies of moving from treatment to policy became clearer. The MDVE body of research should thus be considered, from the point of view of research, a

success. Multiple methods and theory were brought to bear on many independent streams of data in a sequential process. Through the accumulation of both experimental and observational studies, the simplistic causal graph of a treatment influencing an outcome was gradually transformed into a policy graph, one that gives information regarding mechanisms, heterogeneous effects, and the potential unintended consequences of scaling up.

In terms of policy, by contrast, the MDVE proved a mixed blessing. The initial experiment was taken as causal evidence of "what works." Mandatory arrest was expanded before its potential effects were clear and follow-up studies revealed the contingency of results. Because policy makers did not pay sufficient attention to the accumulation of multiple data or the formulation of good theory, they implemented measures that, in many cases, proved to be counterproductive (Sherman, 1993; Sherman and Harris, 2013). It is probably unknowable how much net harm (or good) was done as a result of the early adoption of mandatory arrest policies.

Toward Complex Parsimony

Throughout this article, we have argued for a theoretically informed approach to policy, one that gives as much weight to understanding the social structure of criminal justice policy as it currently does to identifying causal effects within it. Although our focus has been on the interpretation of experimental results, our argument applies equally to problems in observational research design. Experiments and related designs will be most informative if researchers seriously consider theory even before data are collected. We have argued specifically that three key topics—*mechanisms*, *effect heterogeneity,* and *context*—are central to the project of policy creation, although they often are addressed, if at all, as an afterthought, typically as either

boundary conditions or details of the causal narrative. By contrast, we believe that these topics merit the core of our attention as criminologists at both the design and the analysis stage.

In exploring the role of mechanisms, heterogeneity, and context, we have adopted the use of causal graphs because we believe they make more explicit the causal claims and theory that are implicated in policy research. Just the drawing of a graph is an informative exercise for what at first seems to be a simple relationship. We also use these graphs because our principal goal in this article is to complicate underlying causal models—to make the policy intervention a part of the graph itself—that is, to create a "policy graph." This representation can, we believe, eliminate or at least temper the implicit exportability claim of experimental results, making clear the conditionality of causal results and their place within a wider causal system. Furthermore, we agree with Ludwig et al.'s (2011) claim that experiments testing causal mechanisms can yield generalizable, policy-relevant information even if they test interventions that do not correspond to realistic policy options. Testing the visual cue hypothesis of broken windows theory by randomly assigning disorder (e.g., abandoned cars) in a field experiment is an example of this strategy. More generally, although it may seem counterintuitive, the best way to inform policy is not always to test policy (Ludwig et al., 2011: 30).

The ultimate principle is that to provide effective policy, causal effects must be understood within a larger organizational, political, and social structure. Causal graphs transformed into policy graphs provide one way of representing that structure and can complement ongoing causal inquiry: Once we have a theoretical model of a given causal structure, we can go about the complicated task of estimating parts of the causal web and

understanding how its different components are related and interact with each other.[13] Causal graphs and theory also can be used to design experiments themselves and test hypothesized policy mechanisms (Ludwig et al., 2011). A key strength of our recommended approach is that to understand whether a particular causal effect is identified, *we do not need to estimate the whole system*. What we do need to do is to draw on theory to gain an understanding of how parts of the system are stitched together. Otherwise, we are left with static, detached segments of purported causal relationships; such a balkanized view of reality tells very little about the dynamic relationships of complex causal systems. Moreover, it is only by asking questions about the larger causal structure that we can examine when and how a set of experimental results generalizes to other situations. In the context of policy, we must thus reject the separation of forward and backward causality—to understand policy going forward cannot be divorced from a "backward-looking" understanding of the causes of effects.

It should be emphasized again that descriptive data and noncausal analysis constitute a crucial part of the construction and evaluation of a policy graph. Consider, for example, that much of our knowledge on "broken windows" noted earlier was based on careful observation rather than on experimentation or counterfactual causal analysis. Or consider the important gains derived from meta-analyses, demographic-like analysis of stocks and flows through the criminal justice system, and research on the previously neglected links between high rates of incarceration and inequality (Western, 2006) that motivated the hypothesized links in Figure 6. More generally, noncausal analysis of offender patterns derived from longitudinal research on the life

[13]A mature criminological science that combines causal evidence with theory and the accumulation of knowledge across multiple studies and different contexts may lead to areas of consensus that can yield reasonably strong policy inferences. This seems to be occurring in policing research, as reviewed by Nagin and Weisburd (2013, this issue) and Sherman (2013).

course is important for understanding temporal heterogeneity and informing sentencing policy. Blumstein and Nakamura (2009), for example, used criminal-history information to develop longitudinal estimates of expected career lengths and what they called "redemption times." These estimates make no causal claims but are relevant to building a rational policy on setting release times from prison and for employment policy concerning ex-offenders. Descriptive data and theoretically driven analysis thus form essential building blocks of causal policy graphs.

Conclusion

Policy research must be concerned with much more than providing policymakers with information about average causal effects. As we have argued, "what works" does not reduce to the estimation of a causal effect, however precisely measured. *We must instead separate criminology's increasing focus on causality from its policy turn, recognizing that the latter requires a different standard of theory and evidence than does the former.* We must in turn be willing to ask questions related to mechanisms, heterogeneous effects, and context, and all of the real-world phenomena to which these difficulties give rise—such as the possibility of unintended consequences, of policies that change incentive and opportunity structures, and more. There are tools that can enhance this goal, hopefully leading to a set of topics, all of which must be addressed, by researchers who wish to inform policy in a causally uncertain world.[14]

We thus agree with Blomberg et al. (2013) that causal uncertainty does not negate criminological contributions to policy. On the contrary, causal uncertainty can be the subject of

[14]We set aside the disturbing reality that in today's political climate, many policymakers are openly hostile to science and seek to avoid any policy relevant criminology that conflicts with prior world views. The assertion of values over evidence is a deep issue that needs to be confronted by criminologists, but it is well beyond the scope of our article.

investigation and criminologists can do a better job of making explicit their assumptions. It is here that concepts, theory, and descriptive analyses—including insights from practitioners themselves—are essential (Laub, 2012; Sherman, in press). Indeed, it may be that for some of the issues raised in this article, practitioners (e.g., cops on the beat) may be better "theorists" of what policy changes will trigger on the ground than academic criminologists who theorize at a considerable remove. Along with a commitment to conducting ongoing systematic research or experiments, criminal justice agencies can s be co-producers of the sort of feedback information that is necessary to address questions of heterogeneity, mechanisms, and context.

Finally, although James Q. Wilson might have been right in 1975 to note the inability of criminal justice institutions to change root causes, he was wrong, we think, to suggest the casual irrelevance of criminological theory. Even the most "root"-like causes, such as concentrated disadvantage, unemployment, and legacies of racial inequality are central to our understanding of how policies will likely be received by those subjected to intervention by the state—and, therefore, how the policies seemingly recommended by experiments or other causal analyses are likely to work in practice and in the future. The domestic violence results are a clear case in point. Observed in this light, translational criminology is a *process* that entails the constant interplay of theory, research, and practice. Evidence, even if causal, does not "speak for itself" to policy.

References

Aalen, Odd O. and Arnoldo Frigessi. 2007. What can statistics contribute to a causal understanding?*Scandinavian Journal of Statistics,* 34:155–168.

Barinboim, Elias and Judea Pearl. 2013. Causal transportability with limited experiments.In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press.

Becker, Howard. 1963. *Outsiders: Studies in the Sociology of Deviance*. New York: Free Press.

Berk, Richard A. , Alec Campbell, Ruth Klap, and Bruce Western. 1992. The deterrent effect of arrest in incidents of domestic violence: A Bayesian analysis of four field experiments. *American Sociological Review*, 57: 698–708.

Blomberg, Thomas G., Julie Mestre, and Karen Mann. 2013. Seeking causality in a world of contingency: Criminology, research, and public policy. *Criminology &Public Policy*. This issue.

Blumstein, Alfred and Kiminori Nakamura. 2009. "Redemption" in an era of widespread criminal background checks. *Criminology,* 47:327–359.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, Kenneth A. and Judea Pearl. 2013. Eight myths about causality and structural equation models. In *Handbook of Causal Analysis for Social Research*. New York: Spinger.

Braga, Anthony A. and Brenda J. Bond. 2008. Policing crime and disorder hot spots: A randomized controlled trial.*Criminology,* 46:577–607.

Cartwright, Nancy. 2007. Are RCTs the gold standard?*Biosocieties,* 2:11–20.

Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford, U.K.: Oxford University Press.

Cohen, Lawrence E. and Marcus Felson. 1979. Social change and crime rate trends: A routine activity approach.*American Sociological Review,* 44:588–608.

Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.

Cunningham, Scott, Benjamin Engelstatter, and Michael Ward. 2011. Understanding the effects of violent video games on violent crime.Retrieved from ssrn.com/abstract=1804959.

Davis, Richard L. 2008. *Domestic Violence: Intervention, Prevention, Policies and Solutions*.Boca Raton, FL: CRC Press.

Desmond, Matthew. 2012. Eviction and the reproduction of urban poverty. *American Journal of Sociology,* 118:88–133.

Desmond, Matthew and Nicol Valdez. 2013. Unpolicing the urban poor: Consequences of third-party policing for inner-city women. *American Sociological Review*, 78:117–141.

Duncan, Otis D. 1966. Path analysis: Sociological examples.*American Journal of Sociology,* 72:1–16.

Duncan, Otis D. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.

Duneier, Mitchell. 1999. *Sidewalk*. New York: Farrar, Straus & Giroux.

Eckholm, Erik. 2013. Victims' dilemma: 911 calls can bring eviction. *New York Times*. A1, August 17.

Elwert, Felix. 2013. Graphical causal models. In (Steven L. Morgan,ed.),*Handbook of Causal Analysis for Social Research*. New York: Springer.

Fagan, Jeffrey and Garth Davies. 2000. Street stops and broken windows: Terry, race, and disorder in New York City.*Fordham Urban Law Journal,* 28:457–504.

Granger, Robert C. 2011. The Big Why: A learning agenda for the scale-up movement.*Pathways Magazine,*28–32.

Harcourt, Bernard E. 1998. Reflecting on the subject: A critique of the social influence conception of deterrence, the broken windows theory, and order-maintenance policing New York style. *Michigan Law Review,* 97: 291–389.

Harcourt, Bernard E. 2001. *Illusion of Order: The False Promise of Broken Windows Policing*. Cambridge, MA: Harvard University Press.

Harcourt, Bernard E. and Jens Ludwig. 2006. Broken windows: New evidence from New York City and a five-city social experiment.*University of Chicago Law Review,* 73:271–320.

Heckman, James J. 2005. The scientific model of causality. *Sociological Methodology,* 35: 1–97.

Heckman, James J. 2008. Econometric causality. *International Statistical Review,* 76: 1–27.

Heckman, James J., Rodrigo Pinto, and Peter Savelyev. In press. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*.

Heckman James J. and Smith. 1995. Assessing the case for social experiments. *The Journal of Economic Perspectives,* 9: 85-110.

Holland, Paul. 1986. Statistics and causal inference. *Journal of the American Statistical Association,* 81:945–970.

Iyengar, Radha. 2007. Does the certainty of arrest reduce domestic violence? Evidence from mandatory and recommended arrest laws. NBER Working Paper No. 13186.

Kelling, George L. and William H. Sousa Jr. 2001. Do police matter? An Analysis of the impact of New York City's police reforms. Center for Civic Innovation. Civic Report.

Kirk, David S. 2009. A natural experiment on residential change and recidivism: Lessons from Hurricane Katrina.*American Sociological Review,* 74:484–505.

Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz. 2007. Experimental analysis of neighborhood effects.*Econometrica,* 75:83–119.

Kling, Jeffrey R., Jens Ludwig, and Lawrence F. Katz. 2005. Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment.*Quarterly Journal of Economics,* 120:87–130.

Knight, Carly and Christopher Winship. 2013. The causal implications of mechanistic thinking: Identification using directed acyclic graphs (DAGS). In (Steven L. Morgan, ed.),*Handbook of Causal Analysis for Social Research*. New York: Springer.

Kubrin, Charis E. and Eric A. Stewart. 2006. Predicting who reoffends: The neglected role of neighborhood context in recidivism studies.*Criminology,* 44:165–197.

Laub, John H. 2004. The life course of criminology in the United States: The American Society of Criminology 2003 Presidential Address. *Criminology* 42: 1–26.

Laub, John H. 2012. Translational criminology.*Translational Criminology,* 4-5.

Leaf, Clifton. 2013. Do clinical trials work?*New York Times*. SR1, July 14.

Lemert, Edwin M. 1951. *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior*. New York: McGraw-Hill.

Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. Mechanism experiments and policy evaluations.*Journal of Economic Perspectives,* 25:17–38.

Manski, Charles F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press.

Mignon, Sylvia I. and William M. Holmes. 1995. Police response to mandatory arrest laws.*Crime & Delinquency,* 41:430–442.

Miller, Susan. 2005. *Victims as Offenders: The Paradox of Women's Violence in Relationships*.New Brunswick, NJ: Rutgers University Press.

Morgan, Stephen and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Nagin, Daniel S. and David L. Weisburd. 2013. Evidence and public policy: The example of evaluation research in policing.*Criminology &Public Policy*. This issue.

Neyman, Jerzy. 1935. Statistical problems in agricultural experimentation.*Journal of the Royal Statistical Society Series B,* 2:107–180.

Neyman, Jerzy. 1990 [1923]. On the application of probability theory to agricultural experiments. essay on principles. Section 9.*Statistical Science*, 5:465–480.

Pearl, Judea. 2009 [2000]. *Causality: Models, Reasoning, and Inference.* Cambridge, U.K.: Cambridge University Press.

Rein, Martin and Christopher Winship. 1999. The dangers of "Strong"causal reasoning in social policy.*Society,* 36:38–46.

Rosenzweig, Mark R. and Kenneth I. Wolpin. 2000. Natural "natural experiments" in economics.*Journal of Economic Literature,* 38:827–874.

Rubin, Donald B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology,* 66: 688–701.

Rubin, Donald B. 1978. Bayesian-inference for causal effects: The role of randomization.*Annals of Statistics,* 6:34–58.

Sampson, Robert J. 2011. The incarceration ledger: Toward a new era in assessing societal consequences.*Criminology &Public Policy,* 10:819–828.

Sampson, Robert J. 2013. The place of context: A theory and strategy for criminology's hard problems:2012 Presidential Address to the American Society of Criminology.*Criminology,* 51:1–31.

Sampson, Robert J. and Jacqueline Cohen. 1988. Deterrent effects of the police on crime: A replication and theoretical extension.*Law & Society Review,* 22:163–190.

Sampson, Robert J. and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points through Life*. Cambridge, MA: Harvard University Press.

Sampson, Robert J. and Stephen W. Raudenbush. 1999. Systematic social observation of public spaces: A new look at disorder in urban neighborhoods.*American Journal of Sociology,* 105:603–651.

Sampson, Robert J. and Stephen W. Raudenbush. 2004. Seeing disorder: Neighborhood stigma and the social construction of broken windows.*Social Psychology Quarterly,* 67:319–342.

Sanbonmatsu, Lisa, JensLudwig, Lawrence F.Katz, Lisa A.Gennetian, Greg J. Duncan, Ronald C.Kessler, EmmaAdam, Thomas W.McDade, and Stacy Tessler Lindau. 2011. Moving to opportunity for fair housing demonstration program: Final impacts evaluation. Washington, DC: U.S. Department of Housing and Urban Development.

Sherman, Lawrence W. 1993. Defiance, deterrence, and irrelevance: A theory of the criminal sanction. *Journal of Research in Crime and Delinquency,* 30: 445–473.

Sherman, Lawrence W. In press. The rise of evidence-based policing: Targeting, testing, and tracking. *Crime and Justice in America: 1975-2025*.

Sherman, LawrenceW. and Richard A. Berk. 1984. The specific deterrent effects of arrest for domestic assault. *American Sociological Review,* 49:261–272.

Sherman, Lawrence W. and Heather Harris. 2013. *Effects of Arrest over the Life-Course: A 24-Year Follow-up of the Milwaukee Domestic Violence Experiment*. Paper presented at The Stockholm Criminology Symposium, Stockholm, Sweden, June 11.

Sherman, Lawrence W, Douglas A Smith, Janell D. Schmidt, and Dennis P. Rogan. 1992. Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review,* 57:680–690.

Tittle, Charles R. 2004. The arrogance of public sociology. *Social Forces,* 82:1639–1643.

VanderWeele, Tyler J. 2009. Mediation and mechanism. *European Journal of Epidemiology,* 24:217–224.

Western, Bruce. 2006. *Punishment and Inequality in America*. New York: Russell Sage Foundation.

Wikström, Per-Olof H. 2011. Does everything matter? Addressing the problem of causation and explanation in the study of crime. In (J. M. McGloin, C. J. Sullivan, and L. W. Kennedy,eds.), *When Crime Appears: The Role of Emergence*. London, U.K.: Routledge.

Wikström, Per-Olof H., Dietrich Oberwittler, Kyle Treiber, and Beth Hardie. 2012. *Breaking Rules: The Social and Situational Dynamics of Young People's Urban Crime*. Oxford, U.K.: Oxford University Press.

Wilson, James Q. 1975. *Thinking About Crime*. New York: Random House.

Wilson, James Q. and George Kelling. 1982. Broken windows: The police and neighborhood safety. *Atlantic,* 127:29–38.

Robert J. Sampson is Henry Ford II Professor of the Social Sciences at Harvard and Past President of the American Society of Criminology. His research focuses on crime, urban inequality, the life course, neighborhood effects, and the social structure of cities. His most recent book, *Great American City: Chicago and the Enduring Neighborhood Effect*, was published in paperback in 2013 by the University of Chicago Press.

Christopher Winship is Diker-Tishman Professor of Sociology at Harvard and a faculty member in the Kennedy School of Government. In criminology, he has studied The Ten Point Coalition, a group of black ministers working with the Boston police to reduce youth violence, and changes in the racial differential in imprisonment rates. With Stephen Morgan, he is the author of *Counterfactuals and Causal Inference* (Cambridge, 2007).

Carly Knight is a PhD student in the Department of Sociology at Harvard and a Doctoral Fellow with the Multidisciplinary Program in Inequality and Social Policy. Her research interests include comparative/historical sociology, economic sociology, criminology, social mechanisms and causality, and the philosophy of science.

Figure 1. Treatment Affects Outcome Conditional on Experimental Context

$$C : Treatment \longrightarrow Outcome$$

Figure 2. Classic Broken Windows

$$BW \longrightarrow VC \longrightarrow C$$

Figure 3. Broken Windows, Alternative Mechanism

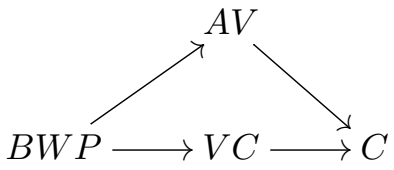
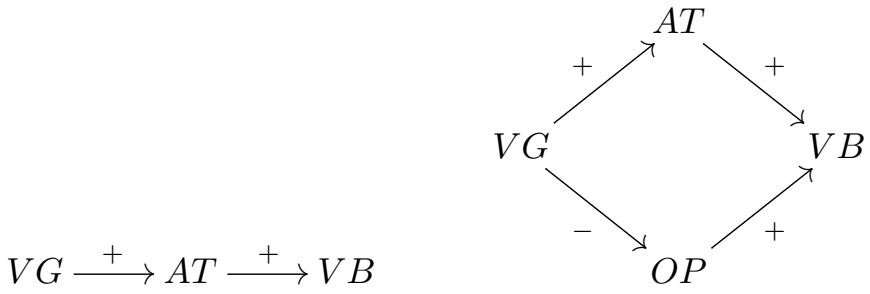


Figure 4. Heterogeneity in the Moving to Opportunity Experiment

$$Male : M \xrightarrow{-} BA \xrightarrow{-} A$$

$$Female : M \xrightarrow{+} BA \xrightarrow{-} A$$

Figure 5. Video Games and Violent Behavior



(A) Simple Graph (B) Extended Graph

Figure 6. Countervailing Effects of Incarceration

(A) Simple Graph (B) Extended Graph

Figure 7. Domestic Violence and Deterrence

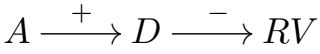


Figure 8. Domestic Violence: Mechanisms and Effect Heterogeneity

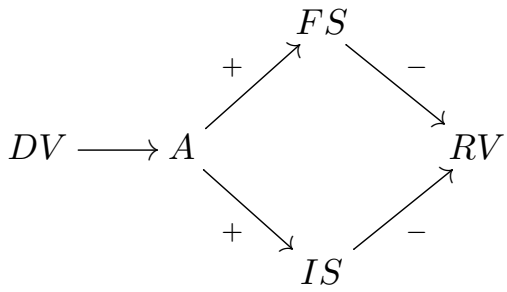


Figure 9. Domestic Violence: Graphing the Policy Intervention

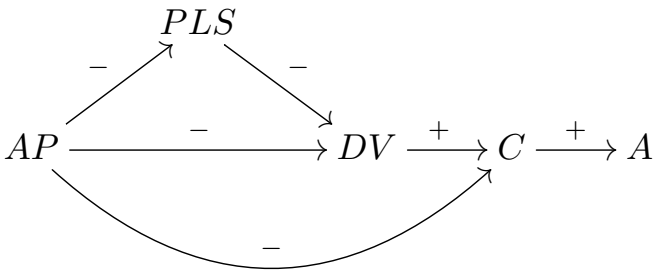
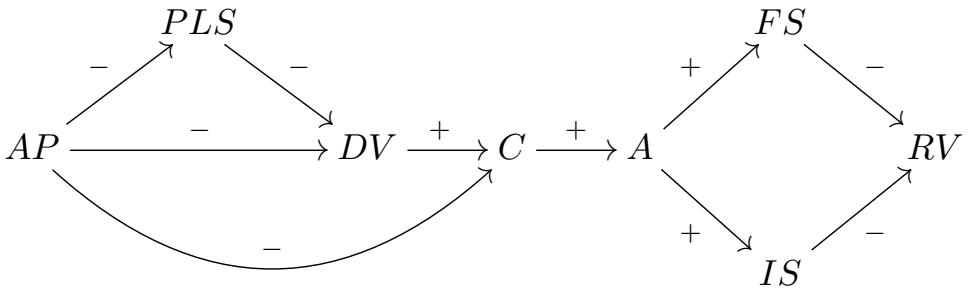



Figure 10. Combined Causal Graph of Domestic Violence