

## ARTICLE

**Broken Windows: New Evidence from New York City  
and a Five-City Social Experiment***Bernard E. Harcourt<sup>†</sup> & Jens Ludwig<sup>††</sup>*

*In 1982, James Q. Wilson and George Kelling suggested in an influential article in the Atlantic Monthly that targeting minor disorder could help reduce more serious crime. More than twenty years later, the three most populous cities in the United States—New York, Chicago, and, most recently, Los Angeles—have all adopted at least some aspect of Wilson and Kelling's theory, primarily through more aggressive enforcement of minor misdemeanor laws. Remarkably little, though, is currently known about the effect of broken windows policing on crime.*

*According to a recent National Research Council report, existing research does not provide strong support for the broken windows hypothesis—with the possible exception of a 2001 study of crime trends in New York City by George Kelling and William Sousa.*

*In this Article, we reexamine the 2001 Kelling and Sousa study and independently analyze the crime data from New York City for the 1989–1998 period. In addition, we present results from an important social experiment known as Moving to Opportunity (MTO) underway in five cities, including New York, Chicago, and Los Angeles, as well as Baltimore and Boston, that provides a unique opportunity to overcome some of the problems with previous empirical tests of the broken windows hypothesis. Under this program, approximately 4,600 low-income families living in high-crime public housing communities characterized by high rates of social disorder were randomly assigned housing vouchers to move to less disadvantaged and disorderly communities.*

*Taken together, the evidence from New York City and from the five-city social experiment provides no support for a simple first-order disorder-crime relationship as hypothesized by Wilson and Kelling, nor for the proposition that broken windows policing is the optimal use of scarce law enforcement resources.*

---

<sup>†</sup> Professor of Law, The University of Chicago Law School.

<sup>††</sup> Associate Professor of Public Policy, Georgetown University.

Special thanks to Jeffrey Fagan, Hal Holzman, Tracey Meares, Steven Messner, Robert Morrissey, Robert Sampson, Mark Shroder and Todd Richardson for comments on earlier drafts; to Ella Delaney and Tim Ross at the Vera Institute for their assistance in assembling the data for New York City; to Stephen Schacht at the National Opinion Research Center (NORC) for comments and guidance regarding the data collection and analysis; as well as to Sarah Rose and Zac Callen for excellent research assistance. The MTO survey and administrative data results summarized here and first published in Kling, Ludwig, and Katz (2005) and Ludwig and Kling (2005) were supported by grants from the National Science Foundation to the National Bureau of Economic Research (NBER) (Grant Nos 9876337 and 0091854) and the National Consortium on Violence Research (Grant No 9513040), as well as by the U.S. Department of Housing and Urban Development, the National Institute of Child Health and Human Development and the National Institute of Mental Health (Grant Nos R01-HD40404 and R01-HD40444), the Robert Wood Johnson Foundation, the Russell Sage Foundation, the Smith Richardson Foundation, the MacArthur Foundation, the W.T. Grant Foundation, and the Spencer Foundation.

## INTRODUCTION

In 1982, James Q. Wilson and George L. Kelling suggested in an influential article in the *Atlantic Monthly* that targeting minor disorder—loitering, panhandling, prostitution, graffiti—could help reduce more serious crime.<sup>1</sup> The “broken windows” theory produced what many observers have called a revolution in policing and law enforcement.<sup>2</sup> Today, the three most populous cities in the United States—New York, Chicago, and, most recently, Los Angeles—have all adopted at least some aspect of Wilson and Kelling’s broken windows theory, primarily through more aggressive enforcement of minor misdemeanor laws, also known as “order maintenance” policing.<sup>3</sup>

Despite the widespread policy influence of Wilson and Kelling’s 1982 *Atlantic Monthly* essay, remarkably little is known about the effects of broken windows. A number of leading researchers in sociology, law, and police studies—including Wesley Skogan at Northwestern, Robert Sampson at Harvard, Stephen Raudenbush at The University of Chicago, Anthony Braga at Harvard, and Jeffrey Fagan at Columbia, among others—have compiled datasets from different urban areas to explore the broken windows hypothesis, but the evidence remains, at best, mixed. In 2000, John Eck and Edward Maguire re-

<sup>1</sup> James Q. Wilson and George L. Kelling, *Broken Windows: The Police and Neighborhood Safety*, *Atlantic Monthly* 29, 38 (Mar 1982) (arguing that a correlation exists between law enforcement’s failure to control certain types of “quality of life” crimes, such as loitering, public drunkenness, and vandalism, and the increased likelihood that violent crimes, such as robbery, will occur).

<sup>2</sup> See Bernard E. Harcourt, *Illusion of Order: The False Promise of Broken Windows Policing* 46 (Harvard 2001), citing Michael Massing, *The Blue Revolution*, *NY Rev of Books* 32 (Nov 19, 1998).

<sup>3</sup> Then New York City Mayor Rudolph Giuliani first embraced quality-of-life policing in the mid 1990s, at a time when high crime rates began declining impressively in the city. Mayor Giuliani and his first police commissioner, William Bratton, traced their quality-of-life initiative directly back to the Wilson and Kelling essay. See New York City Police Department, *Police Strategy No 5: Reclaiming the Public Spaces of New York* 6 (1994). The City of Chicago implemented an anti-gang loitering ordinance in the early 1990s that it vigorously enforced during the 1993–1995 period resulting in misdemeanor arrests of more than 42,000 individuals. See *City of Chicago v Morales*, 527 US 41, 49 (1999). In October 2002, Los Angeles Mayor James Hahn appointed William Bratton police commissioner on a platform that promised a broken windows approach. According to news reports, “Mr. Bratton said his first priority after being sworn in on [October 28, 2002,] would be ending the smile-and-wave approach to crime fighting. He said he wanted policing based on the so-called broken-windows theory.” Charlie LeDuff, *Los Angeles Police Chief Faces a Huge Challenge*, *NY Times* A22 (Oct 24, 2002). See also Megan Garvey, *Bratton Is Planning a Clean Start; The Police Chief, Who Will Be Sworn in Today, Sees Fighting Graffiti as Key to Reducing Crime*, *LA Times Metro* 1 (Oct 25, 2002) (“Far from trivial, Bratton said, fighting graffiti is the key to reducing crime overall and solving more serious offenses—from drug dealing to murder.”); Tina Daunt and Megan Garvey, *Bratton Lays out Ambitious Set of Goals for LAPD*, *LA Times Metro* 1 (Oct 4, 2002) (reporting that Bratton identifies tackling the graffiti problem facing Los Angeles as one of his first orders of business “because it reflects community pride”).

viewed the empirical evidence and studies on broken windows policing in their contribution to Alfred Blumstein's *The Crime Drop in America*, and found that there is little evidence to support the claim that broken windows policing contributed to the sharp decrease in crime during the 1990s.<sup>4</sup>

However, a recent report by a blue-ribbon panel commissioned by the National Research Council (NRC)—which is itself part of the National Academies of Science, chartered in 1863 by Congress to advise the federal government on scientific matters—suggests that there may be new evidence that supports the broken windows theory.<sup>5</sup> The NRC notes that:

There is a widespread perception among police policy makers and the public that enforcement strategies (primarily arrest) applied broadly against offenders committing minor offenses lead to reductions in serious crime. Research does not provide strong support for this proposition.

....

A recent study of New York [City] precincts, however, indicates a strong relationship between the rate of arrests for minor crimes and crime rates in precincts in New York (Kelling and Sousa, 2001). Using a multilevel research design, the authors provide one of the first indications of a direct link between a generalized program of intensive enforcement and declines in more serious crime. While the study uses an innovative modeling approach to estimate this effect, limitations in the data available raise questions regarding the validity of the results.<sup>6</sup>

---

<sup>4</sup> John E. Eck and Edward R. Maguire, *Have Changes in Policing Reduced Violent Crime? An Assessment of the Evidence*, in Alfred Blumstein and Joel Wallman, eds, *The Crime Drop in America* 207, 228 (Cambridge 2000) ("Overall, the evidence is mixed on the efficacy of generic zero-tolerance strategies in driving down rates of violent crime, though serious questions have been raised about their effects on police-community relations."). See also Harcourt, *Illusion of Order* at 88 (cited in note 2) (concluding, on the basis of existing social-scientific data, that neighborhood disorder is not significantly related to more serious crimes when poverty, stability, and race are held constant); Bernard E. Harcourt, *Reflecting on the Subject: A Critique of the Social Influence Conception of Deterrence, the Broken Windows Theory, and Order-Maintenance Policing New York Style*, 97 Mich L Rev 291, 389 (1998) (challenging broken windows policing by arguing that the alleged correlation between disorder and serious crime fails to take into account other factors that may contribute to the deterioration of a neighborhood).

<sup>5</sup> Wesley Skogan and Kathleen Frydl, eds, *Fairness and Effectiveness in Policing: The Evidence* 2 (National Academies 2004).

<sup>6</sup> Id at 229–30 ("Moreover, in a review of the strategies employed in the New York City program, Kelling and Souza [sic] suggest that disorder policing was often applied selectively at the precinct level, focusing on areas of specific problems."). The report refers to a study prepared by George L. Kelling and William H. Sousa, Jr., *Do Police Matter?: An Analysis of the Impact of*

The 2001 study by George Kelling and William Sousa, titled *Do Police Matter?: An Analysis of the Impact of New York City's Police Reforms*, shows that aggressive misdemeanor arrest policies in New York City account for the significant drop in crime during the mid-to-late 1990s.<sup>7</sup> The Kelling and Sousa report has received significant media attention. In addition to being viewed as the only promising evidence by the NRC, the *Economist* reported on the study,<sup>8</sup> as did the *New York Times*,<sup>9</sup> the *Wall Street Journal*,<sup>10</sup> and the *Boston Globe*,<sup>11</sup> both of the latter in editorials, and the *Atlanta Journal-Constitution*.<sup>12</sup> For example, the *Wall Street Journal's* editorial page argued:

A brand new report from the indispensable Manhattan Institute chronicles these law-and-order achievements and explains what made them possible.

. . . .

---

*New York City's Police Reforms* 18 (Manhattan Institute Center for Civic Innovation Civic Report No 22, 2001) (“[D]espite the root-cause theories that have dominated criminological, criminal justice, and much popular thinking about crime control, police can have a significant effect on crime levels in neighborhoods and communities. One singularly important way of doing this is by restoring and maintaining order, through ‘broken windows’ policing.”).

<sup>7</sup> The Kelling and Sousa report was issued with a simultaneously published editorial comment by the authors in the *New York Post*. “So what does all this mean?” Kelling and Sousa ask. “First, it means that New Yorkers should stop listening to critics who contend that police tactics matter little, if at all, in determining crime rates.” These critics, the authors note, “have been parroting what is virtual dogma in criminal-justice circles, that crime is caused by ‘root causes’ such as racism, poverty and social injustice.” In contrast, the authors declare, “This study places the ‘root cause’ theory of crime in serious jeopardy.” George L. Kelling and William H. Sousa, Jr., *Tough Cops Matter*, *NY Post* 41 (Dec 19, 2001). The *New York Post* carried its own editorial the same day. Editorial, *It's the Cops, Stupid*, *NY Post* 42 (Dec 19, 2001) (“Kelling and his colleague William Sousa demonstrate in a new Manhattan Institute study that . . . it was actually the vision, management and plain old hard work of the police that produced the city's historic crime drop.”).

<sup>8</sup> *New York's Mayor: Long-Running Show Closes on Broadway; New One Previews*, *Economist* 25 (Jan 5, 2002) (citing the findings of the Manhattan Institute study in discussing the close relationship between Mayor Giuliani's successful implementation of the broken windows policing strategy and the decrease in the rate of violent crime in New York City).

<sup>9</sup> Kevin Flynn, *Study Says a Slumping Economy Doesn't Mean Crime Will Rise*, *NY Times* D8 (Dec 19, 2001) (“In fact, the researchers found that many areas of New York that had higher unemployment in recent years actually had slightly sharper declines in crime, in part because of innovative patrol strategies by police officers.”).

<sup>10</sup> Editorial, *New York's Finest*, *Wall St J* A12 (Dec 27, 2001) (praising Mayor Giuliani for his role in crime reduction).

<sup>11</sup> Editorial, *Behind Giuliani's Jab*, *Boston Globe* A14 (Dec 29, 2001) (noting that Kelling and Sousa argue that zero-tolerance police tactics were instrumental in reducing crime, but also raising doubts about the efficacy of the policing strategy).

<sup>12</sup> Colin Campbell, *New York a Blueprint for Cutting Atlanta Crime*, *Atlanta J-Const* 5F (Dec 23, 2001) (citing the Kelling and Sousa study in discussing the reasons for the decline in New York's crime rate during the 1990s and the lessons that Atlanta should take away from the New York experience as it addresses its rising crime rate).

“Do Police Matter?” also does a great public service in thoroughly refuting those media critics and political opponents of the Republican Mayor who’ve insisted for the past eight years that the NYPD had little if anything to do with the fall in crime. In this alternative universe, the city’s drop in crime should be credited to low unemployment from a booming economy. Or the decline in crack cocaine use that had plagued the 1980s. Or the demographic reality that the proportion of young males—the most common offenders—to the general population had dropped.

In fact, none of these alternative explanations stands up to scrutiny.<sup>13</sup>

An even more recent paper by economists Hope Corman and Naci Mocan applies a slightly different empirical approach to data from New York City and claims to support the Kelling-Sousa conclusion.<sup>14</sup> Corman and Mocan analyze monthly time-series data for New York City as a whole and claim that the dramatic increase in misdemeanor arrest rates in New York during the 1990s is responsible for a large share of the city’s drop in crime over this period. So while Kelling and Sousa use variation across precincts over time in misdemeanor arrests and crime rates to identify the effects of the former on the latter, Corman and Mocan use city-wide variation over time to generate a similar finding. Moreover, Corman and Mocan point to deterrence as the most plausible mechanism for this relationship, given that misdemeanor arrests typically result in either no jail time or short spells of incarceration.<sup>15</sup> The Kelling and Sousa study, together with the Corman and Mocan paper, are thus important contributions, representing the best existing evidence supporting the broken windows hypothesis and the related (and widespread) broken windows or order-maintenance policing strategy.

In this Article, we set out to reanalyze and assess the best available evidence from New York City about the effects of broken windows policing. Although Kelling and Sousa were unwilling to share

---

<sup>13</sup> Editorial, *New York’s Finest*, Wall St J at A12 (cited in note 10). Even Queensland’s *Courier-Mail*, reporting on the 2001 study, stated that “in precinct after precinct Kelling and Sousa found a similar pattern—as ‘broken windows’ policing was increased, violent crime declined.” Ron Bruton, *Broken Windows’ Plan Shatters Crime Theory*, *Courier-Mail* 24 (Queensland, Australia) (Jan 5, 2002).

<sup>14</sup> Hope Corman and Naci Mocan, *Carrots, Sticks and Broken Windows*, 48 J L & Econ 235, 262 (2005) (“A 10 percent increase in misdemeanor arrests decreases motor vehicle thefts by 1.6 to 2.1 percent, [and] robberies by 2.5 to 3.2 percent . . . . We do not find strong evidence to support the contention that a broken-windows policing strategy affects the other crimes.”).

<sup>15</sup> Id at 251 (noting that of the 122,797 misdemeanor arrests studied, only 9.4 percent “resulted in a conviction with a jail sentence,” with “an average length of stay” of 27.5 days, resulting in an “expected jail sentence for [a] misdemeanor arrest” of about 2.6 days).

their data with us, we were able to assemble a dataset with the same precinct-level crime and arrest information as used in their study. We demonstrate that the pattern of crime changes across New York precincts during the 1990s that Kelling and Sousa attribute to broken windows policing is more consistent with what statisticians call mean reversion: those precincts that received the most intensive broken windows policing during the 1990s are the ones that experienced the largest increases in crime during the city's crack epidemic of the mid-to-late 1980s. Consistent with findings elsewhere from city-level data,<sup>16</sup> jurisdictions with the greatest increases in crime during the 1980s tend to experience the largest subsequent declines as well. We call this Newton's Law of Crime: what goes up must come down (and what goes up the most tends to come down the most). For similar reasons we argue that the Corman and Mocan study also is unable to determine convincingly that broken windows policing is a causal contributor to crime rates in New York City.

Because our reanalysis of the New York data leaves us with a Scottish verdict—"not proven"—we then turn to data from a unique randomized experiment conducted by the U.S. Department of Housing and Urban Development (HUD) known as Moving to Opportunity (MTO), which provides a new opportunity to test the original Wilson and Kelling broken windows thesis. MTO has been in operation since 1994 in five cities, including the three largest cities in the country that have adopted aspects of broken windows policing (New York, Chicago, and Los Angeles), as well as Baltimore and Boston. Under MTO, a total of around 4,600 low-income families living in public housing communities characterized by high rates of crime and social disorder were randomly assigned housing vouchers to move to less disadvantaged and disorderly communities.<sup>17</sup> The random assignment of families to neighborhoods in MTO helps overcome the problem of determining the causal effects of neighborhood disorder on individual criminal behavior that plagues most previous studies in this literature.<sup>18</sup>

---

<sup>16</sup> See Steven Raphael and Jens Ludwig, *Prison Sentence Enhancements: The Case of Project Exile*, in Jens Ludwig and Philip J. Cook, eds, *Evaluating Gun Policy* 251, 265 (Brookings 2003) (positing that the reduction in violence in such areas finds its root, not in federalized prosecution of eligible gun offenses, but rather in the fact that the violence accompanying the introduction of crack cocaine in the 1980s had run its course by the late 1990s).

<sup>17</sup> See Larry Orr, et al, *Moving to Opportunity: Interim Impacts Evaluation* ii-iv (HUD Office of Policy Development 2003), online at <http://www.huduser.org/Publications/pdf/MTOFullReport.pdf> (visited Jan 19, 2006) (describing the MTO program).

<sup>18</sup> Because most people have at least some degree of choice over where they live and with whom they associate, previous *nonexperimental* studies may confound the effects of neighborhood disorder and other characteristics on people's behavior with the effects of difficult-to-

The implications of MTO for the ongoing debates about the broken windows theory have not yet been explored.<sup>19</sup> Yet the results from MTO suggest that moving people to communities with less social or physical disorder—the key intervening factor in the original Wilson and Kelling broken windows hypothesis—on balance does *not* lead to a reduction in their criminal behavior. It is important to note that MTO changed multiple aspects of people's neighborhoods: MTO families moved to neighborhoods that were less disorderly, but also had fewer low-income families and more high-status households. MTO thus tested the combined effects of less disorder and increased affluence within a community, which is arguably still a policy-relevant "treatment combination" for neighborhoods under the broken windows model because reductions in disorder, like other improvements in neighborhood amenities, should on average translate into increased neighborhood gentrification.

Taken together our examination of data from New York City and MTO provide no support for the idea that broken windows enforcement activities, including order-maintenance policing or other measures designed to reduce the level of social or physical disorder within a community, represent the optimal use of scarce government resources.

This Article is organized as follows. Part I locates the broken windows theory within sociological and policy traditions, and reviews preceding efforts to test the broken windows theory and the practice of broken windows policing. Part II then presents our discussion of the evidence from New York City. Part III presents our findings from the MTO experiment, demonstrating that randomly assigning people to move to less disorderly communities does not yield the simple "less disorder, less criminal activity" result that broken windows policing predicts.

---

measure individual attributes that influence both their involvement with crime and their choice of residential neighborhood.

<sup>19</sup> While recent results of studies on neighborhood effects on criminal behavior have been published in the field of economics, these findings are currently not widely known outside of that discipline and, as a result, their implications for broken windows have never been explored. See Jeffrey R. Kling, Jens Ludwig, and Lawrence F. Katz, *Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment*, 120 Q J Econ 87, 87 (2005) (concluding that the effects of MTO programs on the propensity of a youth to become involved in some form of violent or property crime depends, to an extent, on the differences in how male and female youths adapt and respond to similar new neighborhood environments).

## I. LOCATING THE BROKEN WINDOWS THEORY

## A. The Sociolegal Context

There is a long tradition within sociolegal research of studying visual cues of neighborhood disorder and exploring the relationship between those neighborhood characteristics and deviance. Prompted by a recurring observation of dramatic variations in crime rates across neighborhoods, the tradition grew over decades of research taking seriously the idea that there may be “neighborhood effects” on the production of crime. That is, arrangements in social space may significantly affect human behavior. This research tradition traces importantly to the early Chicago School of sociology—the monographs on neighborhoods and spatial settings, the Jewish ghetto,<sup>20</sup> the Italian “slum,”<sup>21</sup> the Near North side of Chicago,<sup>22</sup> taxi-dance halls,<sup>23</sup> and brothels<sup>24</sup>—and to the later social interactionist research of Erving Goffman, especially his study *Behavior in Public Places*,<sup>25</sup> and others such as Albert Cohen<sup>26</sup> and Jane Jacobs.<sup>27</sup>

---

<sup>20</sup> See generally Louis Wirth, *The Ghetto* (Chicago 1928) (discussing the extent to which isolation related to the Jewish ghetto has shaped the character of the “Jew” and the nature of his social life).

<sup>21</sup> See generally William F. Whyte, *Street Corner Society: The Social Structure of an Italian Slum* (Chicago 1943) (exploring life and crime in a predominantly Italian slum district known as Cornerville).

<sup>22</sup> See generally Harvey Warren Zorbaugh, *The Gold Coast and the Slum: A Sociological Study of Chicago's Near North Side* (Chicago 1929) (discussing the problems created by the divergence of interests and heritages of the differing groups that compose the Lower North Side of Chicago).

<sup>23</sup> See generally Paul G. Cressey, *The Taxi-Dance Hall* (Chicago 1932) (offering an exploration of the typical taxi-dance hall with its customers and employees, a background regarding the history of such halls as urban institutions, and an overview of the kinds of controls established to enforce standards).

<sup>24</sup> See generally Walter C. Reckless, *Vice in Chicago* (Patterson Smith 1969) (discussing vice in Chicago since the closing of its red light district in 1912).

<sup>25</sup> See generally Erving Goffman, *Behavior in Public Places: Notes on the Social Organization of Gatherings* (Free Press 1963) (exploring human behavior in public places).

<sup>26</sup> See generally Albert K. Cohen, *Delinquent Boys: The Culture of the Gang* (Free 1955) (discussing the delinquent subculture, specifically addressing why it arises and persists in certain neighborhoods but not in others, and offering a new perspective on the issue of psychogenic versus cultural-transmission theories of delinquency).

<sup>27</sup> See generally Jane Jacobs, *The Death and Life of Great American Cities* (Random House 1961) (offering a critique of current city planning while discussing the factors that result in positive and negative changes occurring in differing neighborhoods, the roles neighborhoods play in cities, and what new principles should be embraced in city planning and rebuilding). As Andrew Abbott notes, “[The] Chicago [school of thought] felt that no social fact makes any sense abstracted from its context in social (and often geographic) space and social time. Social facts are located.” Andrew Abbott, *Of Time and Space: The Contemporary Relevance of the Chicago School*, 75 Soc Forces 1149, 1152 (1997) (discussing the theoretical position promoted by the Chicago school of thought in the study of sociology).

One of the most striking findings from the neighborhood effects research comes from the dramatic differences across neighborhoods in rates of crime and delinquency—even across neighborhoods with similar concentrations of social disadvantage as measured by average rates of poverty, unemployment, familial and residential instability, and dependence on government benefit programs.<sup>28</sup> Robert Sampson and Stephen Raudenbush trace the rich intellectual history and the variations over time in neighborhood-effects research in their thorough paper, *Systematic Social Observation of Public Spaces*.<sup>29</sup>

A consideration of the research in this area suggests two lasting puzzles. The first puzzle focuses on locating sources of variation in crime across neighborhoods and identifies two leading candidates. First, differences across areas in crime rates could be due to unobservable individual characteristics related to the residents of the neighborhood, raising the possibility of self-selection on the part of the individuals. Put differently, some neighborhoods may have more crime because they are home to a larger share of crime-prone people, although all of the individual attributes that predispose some people to engage in criminal activity are difficult to measure in social science datasets. A second explanation is that variation across areas in crime rates may be due to differences in social processes and conditions across neighborhoods, including disorderliness or informal mechanisms of social control. The notion of social disorganization pioneered by Clifford Shaw and Henry McKay<sup>30</sup> represented one effort to locate the answer to this first puzzle, at least in part, in mechanisms of informal social control and collective action—in identifying an agency of social control that could be disrupted by residential mobility and economic conditions. Sampson, Raudenbush, and Fenton Earls's Project on Human Development in Chicago Neighborhoods (PHDCN) research

<sup>28</sup> See, for example, Robert J. Sampson, Stephen W. Raudenbush, and Fenton Earls, *Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy*, 277 *Science* 918, 923 (1997) (offering evidence that the social cohesion among neighbors combined with their willingness to intervene on behalf of the common good acts as “a robust predictor of lower rates of violence”); Edward L. Glaeser, Bruce Sacerdote, and José A. Scheinkman, *Crime and Social Interactions*, 111 *Q J Econ* 507, 542 (1996) (arguing that the high cross-city variance in the proportion of potential criminals who do not respond to social influences indicates covariance across agents, such as the age of the criminals or the presence of strong families).

<sup>29</sup> Robert J. Sampson and Stephen W. Raudenbush, *Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods*, 105 *Am J Soc* 603, 637 (1999) (arguing that due to shared theoretical features, both public disorder and predatory crimes are explained by a concentration of disadvantage and lowered collective efficacy).

<sup>30</sup> Clifford R. Shaw and Henry D. McKay, *Juvenile Delinquency and Urban Areas: A Study of Rates of Delinquents in Relation to Differential Characteristics of Local Communities in American Cities* 446 (Chicago 1942) (establishing that the distribution of juvenile delinquents in space and time follows the pattern of the physical structure and of the social organization of the American city).

represents another answer focused on informal social processes, more specifically on the notion of “collective efficacy,” which they define as “the linkage of cohesion and mutual trust with shared expectations for intervening in support of neighborhood social control.”<sup>31</sup>

A second puzzle focuses on the issue of remedies. Even if the neighborhood-effects research suggests a causal relationship between, on the one hand, identifiable social processes or neighborhood characteristics and, on the other hand, crime, does the causal explanation offer insight into what can be done to change things in a public policy sense? In this regard, the sociological theories have been relatively quiet, reflecting a general hesitation to move from the positive to the prescriptive.

It is within this rich research field that the broken windows hypothesis emerged in the early 1980s. Though first articulated and tested by Philip Zimbardo, a Stanford psychologist, in the late 1960s, the broken windows theory was most clearly articulated and popularized in James Q. Wilson and George L. Kelling’s 1982 *Broken Windows* article.<sup>32</sup> “[D]isorder and crime are usually inextricably linked, in a kind of developmental sequence,” Wilson and Kelling argued, so that efforts to reduce disorder might ultimately translate into reductions in criminal activity as well.<sup>33</sup> Minor social disorder—littering, loitering, public drinking, panhandling, and prostitution—as well as physical disorder—graffiti, abandoned buildings, and littered sidewalks—if tolerated in a neighborhood, produce an environment that is likely to attract crime. These forms of disorder signal to potential criminals that delinquent behavior will not be reported or controlled—that no one is in charge. To law-abiding citizens, these disorderly conditions signal the need to avoid the streets or even flee the neighborhood. One broken window, left unrepaired, invites other broken windows. These pro-

<sup>31</sup> Sampson and Raudenbush, 105 Am J Soc at 612–13 (cited in note 29) (“Just as individuals vary in their capacity for efficacious action, so too do neighborhoods. And just as individual self-efficacy is situated relative to a particular task rather than global, our notion of collective efficacy here is conceptualized as relative to the task of maintaining order in public spaces.”); Sampson, Raudenbush, and Earls, 277 Science at 919 (cited in note 28) (“[T]he collective efficacy of residents is a critical means by which urban neighborhoods inhibit the occurrence of personal violence, without regard to the demographic composition of the population.”).

<sup>32</sup> Wilson and Kelling, Atlantic Monthly at 29 (cited in note 1). They discuss the broken windows hypothesis as follows:

The key is to identify neighborhoods at the tipping point—where the public order is deteriorating but not unreclaimable, where the streets are used frequently but by apprehensive people, where a window is likely to be broken at any time, and must quickly be fixed if all are not to be shattered.

Id at 38.

<sup>33</sup> Id at 31 (“Social psychologists and police officers tend to agree that if a window in a building is broken and is left unrepaired, all the rest of the windows will soon be broken.”).

gressively break down community standards and leave the community vulnerable to crime. In this way, disorder breeds crime: "Such an area is vulnerable to criminal invasion. Though it is not inevitable, it is more likely that here," Wilson and Kelling wrote, "drugs will change hands, prostitutes will solicit, and cars will be stripped. That the drunks will be robbed by boys who do it as a lark, and the prostitutes' customers will be robbed by men who do it purposefully and perhaps violently."<sup>34</sup>

The broken windows theory thus addresses the first puzzle of the neighborhood-effects literature in a straightforward and provocative way: it is the variation in disorder in neighborhoods that explains the variation in crime, holding structural disadvantage constant. The real trigger is disorderliness itself. The theory was familiar to sociologists because of its proximity to theories of urban decay and social contagion. Urban sociologists interpreted the broken windows hypothesis through the lens of urban decline: disorderliness, dilapidation, abandonment, and social disorder, such as prostitution, public intoxication, and drug use, reflected and reinforced, in a cyclical manner, declining property values, residential instability, and the gradual decay of the urban neighborhood.<sup>35</sup> A closely related interpretation is suggested by Philip Cook and Kristin Goss's review of the standard models of "social contagion."<sup>36</sup> From this contagion perspective, the broken windows phenomenon reflects an information cascade: people with imperfect information about the risks and rewards of criminal activity may infer the net returns to crime from the social environment.<sup>37</sup> Information limitations are at the heart of the information cascade model. Here, the potential criminals do not know the probability of being detected in a neighborhood, but the lack of enforcement of minor crime and disorder fills this void and signals low enforcement. The characteristics

<sup>34</sup> Id at 32.

<sup>35</sup> See Gerald E. Frug, *City Making: Building Communities without Building Walls* 199–200 (Princeton 1999) (discussing the effects of "get-tough" policing strategies on community building efforts in America's urban centers); Wesley Skogan, *Disorder and Decline: Crime and the Spiral of Decay in American Cities* 3 (Free 1990) (arguing that disorder, in the form of graffiti, abandoned cars, vandalism of public and private property, and decaying homes, is an instrument of destabilization and neighborhood decline).

<sup>36</sup> Philip J. Cook and Kristen A. Goss, *A Selective Review of the Social-Contagion Literature* 3–4 (Sanford Institute of Public Policy Studies, Duke University, Working Paper, 1996) (on file with authors) (providing an overview of the relevant literature on "social contagion"—a term used by scholars to describe cases in which attitudes and behaviors, especially of the antisocial variety, spread through populations).

<sup>37</sup> Id at 24 ("The terms 'information cascades' and 'herd behavior' describe instances in which an individual adopts and acts on the judgments of others, not because he has been pressured to do so, but because the leaders' actions are believed to convey information about the individually optimal choice.").

of the local physical environment, which are themselves the product of the accumulated series of behaviors of local residents, thus communicate the statistical likelihood of being apprehended. They are a signaling mechanism that feeds into the calculus of whether to commit crime. This “contagion” interpretation offers a straightforward explanation of broken windows familiar to most sociologists and economists.<sup>38</sup>

As to the second puzzle—concerning the public policy prescriptions—Wilson and Kelling’s original *Broken Windows* essay itself did not compel a particular policy outcome. From a policy perspective, the broken windows hypothesis is, in principle, consistent with a variety of potential policy levers, ranging from changes in policing to community organizing. Nevertheless, most policymakers seem to have understood the theory as implying what has come to be known as “broken windows policing”—also known as “order-maintenance,” “zero-tolerance,” or “quality-of-life” policing. So for instance, in their 2001 study, George Kelling and William Sousa suggest that the most effective way to address disorder and reduce crime is to increase the number of misdemeanor arrests.<sup>39</sup>

## B. Testing the Broken Windows Hypothesis

To date, empirical testing of the broken windows theory has taken one of two forms. A first approach attempts to measure neighborhood disorder and crime, as well as other correlates of criminality, such as poverty and residential instability, in order to determine whether there are statistically interesting correlations among these variables. A sec-

---

<sup>38</sup> For a discussion of the etiology of less and more serious crimes, see Michael R. Gottfredson and Travis Hirschi, *A General Theory of Crime* 85–120 (Stanford 1990) (discussing the relationship between an individual’s development and maintenance of self-control and an individual’s propensity for criminality). For a discussion of how “routine activities” across neighborhoods may affect criminal opportunities and outcomes, see generally Lawrence E. Cohen and Marcus Felson, *Social Change and Crime Rate Trends: A Routine Activity Approach*, 44 Am Soc Rev 588 (1979) (arguing that the structure of routine activities, such as a dispersion of activities away from households and families, increases the opportunity for crime and thus generates higher crime rates). See also Lawrence E. Cohen, James R. Kluegel, and Kenneth C. Land, *Social Inequality and Predatory Criminal Victimization: An Exposition and Test of a Formal Theory*, 46 Am Soc Rev 505, 507 (1981) (offering an additional explanation of the “routine activities” theory of criminal victimization, focusing primarily on the role played by five factors: exposure, guardianship, proximity to potential offenders, attractiveness of potential targets, and definitional properties of specific crimes themselves). Additional discussion of the “social disorganization” model of disorder and neighborhood effects on crime is provided by Robert J. Bursik, Jr., *Social Disorganization and Theories of Crime and Delinquency: Problems and Prospects*, 26 Criminology 519, 538 (1988) (discussing criticisms of the social disorganization model as well as new applications of the framework to substantive areas, such as victimization).

<sup>39</sup> Kelling and Sousa, *Do Police Matter?* at 7 (cited in note 6) (describing successes of the broken windows policing strategy in terms of the increased arrests for misdemeanors and decreased offenses for violent crimes between 1989 and 1998).

ond approach focuses on measures of broken windows policing—for instance, rates of misdemeanor arrests—and conducts relatively similar statistical analyses on these variables in order, again, to identify significant correlations. We begin by reviewing the first approach.

### 1. Disorder and crime.

Early on, many proponents of the broken windows hypothesis pointed to the research of Wesley Skogan, especially his 1990 monograph *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*, and argued that it empirically verified the broken windows theory.<sup>40</sup> Skogan's book addressed the larger question of the impact of neighborhood disorder on urban decline, but in one section Skogan discussed the broken windows hypothesis, ran a regression of neighborhood disorder on robbery victimization, and concluded that "Broken windows' do need to be repaired quickly."<sup>41</sup> Many observers interpreted this conclusion as an endorsement of the broken windows theory and accepted Skogan's view of the evidence. George Kelling, coauthor of *Broken Windows* and of a book entitled *Fixing Broken Windows*,<sup>42</sup> contended that Wesley Skogan "established the causal links between disorder and serious crime—empirically verifying the 'Broken Windows' hypotheses."<sup>43</sup> Dan Kahan at Yale similarly argued that "[t]he work of criminologist Wesley Skogan supplies empirical support for the 'broken windows' hypothesis."<sup>44</sup> Subsequent work by one of us, however, has cast doubt on what conclusions can properly be drawn from Skogan's analysis.<sup>45</sup>

A few years later, Ralph Taylor of Temple University conducted research in sixty-six neighborhoods in Baltimore using longitudinal data, and attempted to determine the relationship between neighbor-

---

<sup>40</sup> Skogan, *Disorder and Decline* at 120–24 (cited in note 35) (offering evidence suggesting that a "community policing" program directed at controlling disorder in Newark, New Jersey, could help stem the process of urban decline).

<sup>41</sup> *Id.* at 75.

<sup>42</sup> George L. Kelling and Catherine M. Coles, *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities* 23–27 (Simon & Schuster 1996) (discussing Skogan's findings regarding the relationship between fear and disorder as evidence supporting the notion that ignoring disorder poses dangers, such as crime and urban decline).

<sup>43</sup> *Id.* at 24.

<sup>44</sup> Dan M. Kahan, *Social Influence, Social Meaning, and Deterrence*, 83 Va L Rev 349, 369 (1997). See also Dan Kahan, *Between Economics and Sociology*, 95 Mich L Rev 2477, 2488 n 62 (1997) ("According to criminologists, the primary cue that crime is tolerated or expected is visible public disorder.").

<sup>45</sup> Harcourt, *Illusion of Order* at 78 (cited in note 2) (concluding that "there are no statistically significant relationships between disorder and purse-snatching, physical assault, burglary, or rape when other explanatory variables are held constant . . . [Thus] the data do not support the broken windows hypothesis").

hood crime and what he termed social and physical “incivilities”—panhandlers, public drunks, trash, graffiti, and vacant lots, among other things. What he found was that while certain types of incivilities were associated with crime or urban decay, others were not. In his 2001 book, *Breaking Away from Broken Windows*, Taylor concludes from his data that different types of incivilities may require different policy responses. “Researchers and policy-makers alike,” Taylor writes, “need to break away from broken windows per se and widen the models upon which they rely, both to predict and to preserve safe and stable neighborhoods with assured and committed residents.”<sup>46</sup>

One of the most comprehensive and thorough examination of the broken windows theory to date is Robert Sampson and Stephen Raudenbush’s 1999 study. Their study grows out of the PHDCN and is based on systematic social observation: using trained observers who drove a sports utility vehicle at five miles per hour down every street in 196 Chicago census tracts, and randomly selecting 15,141 street sides, they were able to collect precise data on neighborhood disorder. With regard to the disorder-crime nexus, Sampson and Raudenbush found that disorder and predatory crime are only moderately correlated, but that, when antecedent neighborhood characteristics are taken into account, the connection between disorder and crime “vanished in 4 out of 5 tests—including homicide, arguably our best measure of violence.”<sup>47</sup> They nevertheless suggest that disorder may have indirect, neighborhood effects on crime by influencing “migration patterns, investment by businesses, and overall neighborhood viability.”<sup>48</sup>

On the basis of their extensive research, Sampson and Raudenbush conclude that “[a]ttacking public disorder through tough police tactics may thus be a politically popular but perhaps analytically weak strategy to reduce crime.”<sup>49</sup> As an alternative to the broken windows theory, Sampson and Raudenbush suggest that disorder is of the same etiology as crime—being, so often, forms of minor crime—and that both crime and disorder have the same antecedent conditions.

---

<sup>46</sup> Ralph B. Taylor, *Breaking Away from Broken Windows: Baltimore Neighborhoods and the Nationwide Fight against Crime, Guns, Fear, and Decline* 22 (Westview 2001) (concluding that a more integrated perspective, which combines the current results regarding incivilities and contemporary knowledge regarding the multiplicity of factors affecting neighborhoods over time, should be developed).

<sup>47</sup> Sampson and Raudenbush, 105 Am J Soc at 637 (cited in note 29) (offering results contradicting a strong version of the broken windows thesis but concluding that the role of disorder remained theoretically relevant for other purposes).

<sup>48</sup> Id at 637.

<sup>49</sup> Id at 638 (“A more subtle approach suggested by this article would look to how informal but collective efforts among residents to stem disorder may provide unanticipated benefits for increasing collective efficacy in the long run lowering crime.”) (internal citation omitted).

“Rather than conceive of disorder as a direct cause of crime, we view many elements of disorder as part and parcel of crime itself.”<sup>50</sup> Thus, “a reasonable hypothesis is that public disorder and predatory crimes are manifestations of the same explanatory process, albeit at different ends of a ‘seriousness’ continuum.”<sup>51</sup>

## 2. Studies of aggressive misdemeanor-arrest policing.

Another strand of research, focusing on studies of aggressive arrest policies, was also brought to bear on the broken windows hypothesis. Here, too, James Q. Wilson sparked the debate, primarily with his 1968 book *Varieties of Police Behavior*, and his research with Barbara Boland on the effects of police arrests on crime.<sup>52</sup> Wilson and Boland hypothesized that aggressive police patrols, involving increased stops and arrests, have a deterrent effect on crime.

A number of contributions ensued, both supporting and criticizing these findings, but, as Robert Sampson and Jacqueline Cohen suggested in 1988, the results were “mixed.”<sup>53</sup> There have been strong contributions to the literature, such as the 1999 study led by Anthony

---

<sup>50</sup> Id at 608.

<sup>51</sup> Id. Sampson and Raudenbush have a more recent study showing that neighborhood racial composition affects people’s perceptions of neighborhood disorder. Robert J. Sampson and Stephen W. Raudenbush, *Seeing Disorder: Neighborhood Stigma and the Social Construction of “Broken Windows,”* 67 Soc Psych Q 319 (2004). The study explores the grounds on which individuals form perceptions of disorder. Id at 337 (concluding that although observed disorder may predict perceived disorder to some degree, the racial and economic context affect an individual’s perceived disorder to a greater extent). For a study of disorder and youth crime in Canada, see John Hagan and Bill McCarthy, *Mean Streets: Youth Crime and Homelessness* 12 (Cambridge 1997):

[M]any contemporary studies point to positive relationships between living on the street and minor and more serious crime, including break and enter, robbery, assault, and drug- and sex-related offenses. We investigate an array of crimes and give particular attention to minor and more serious theft, prostitution, drug selling, and assault.

<sup>52</sup> See James Q. Wilson, *Varieties of Police Behavior: The Management of Law and Order in Eight Communities* 281–82 (Harvard 1968) (describing three styles of police management and concluding that a “legalistic” approach, emphasizing a high arrest rate, has gained popularity because it has been shown to deter criminals). See also James Q. Wilson and Barbara Boland, *The Effects of the Police on Crime: A Response to Jacob and Rich*, 16 L & Socy Rev 163, 168–69 (1981) (critiquing a study challenging the proposition that increased arrest rates will have a deterrent effect on crime); James Q. Wilson and Barbara Boland, *The Effect of the Police on Crime*, 12 L & Socy Rev 367, 380 (1978) (“Aggressiveness and a large number of patrol units, separately and in combination, will lead to a higher arrest ratio for robbery, and this higher ratio, in turn, leads to a lower robbery crime rate.”).

<sup>53</sup> Robert J. Sampson and Jacqueline Cohen, *Deterrent Effect of the Police on Crime: A Replication and Theoretical Extension*, 22 L & Socy Rev 163, 166 (1988) (discussing the results of a disorderly conduct enforcement experiment performed by members of the Newark, New Jersey, police department and finding that while the results could be seen to support the belief that proactive policing may reduce crime, they remained mixed).

Braga, titled *Problem-Oriented Policing in Violent Crime Places*.<sup>54</sup> But still, most of this research is unable to distinguish between the broken windows hypothesis and more traditional explanations of incapacitation and deterrence associated with increased police arrests, presence, contact, and surveillance. The problem is somewhat endemic to the design of these studies. As Sampson and Cohen conclude with regard to their own work, “It is true . . . that our analysis was not able to choose definitely between the two alternative scenarios.”<sup>55</sup>

In this vein, Jeffrey Fagan and Garth Davies test, in their 2003 research titled *Policing Guns: Order Maintenance and Crime Control in New York*, whether quality-of-life policing in New York City contributed to the reduction in lethal violence in the late 1990s. They analyze precinct crime rates from 1999 and try to determine whether these crime rates can be predicted by the amount of stop-and-frisk activity that occurred in the precinct in the preceding year. Based on their research, Fagan and Davies find that “[f]or both violence arrests broadly and homicide arrests specifically, there is no single category of citizen stops by police that predicts where crime will increase or decrease in the following year.”<sup>56</sup> When they examine homicide fatalities, they observe different effects by type of stop and by victim race. “Stops for violence are significant predictors of reductions in both gun homicide deaths and overall homicide deaths, but only among Hispanics.”<sup>57</sup> In contrast, for African Americans, no type of arrest predicts homicide victimization a year later; and for whites, the results are not reliable because of the low white-homicide victimization rate.

Why is it that there may be effects for Hispanics, but not for African Americans? Fagan and Davies suggest that it may have to do with what they call “stigma saturation” in black communities: when stigma is applied in ways that are perceived as too harsh and unfair, it may

---

<sup>54</sup> Anthony A. Braga, et al, *Problem-Oriented Policing in Violent Crime Places: A Randomized Controlled Experiment*, 37 *Criminology* 541, 571 (1999) (evaluating the effects of problem-oriented policing interventions on urban violent crime problems and suggesting that police efforts focused on “modifying the places, routine activities, and situations that promote violence may be effective in reducing violent behavior”). See also Anthony A. Braga, *Problem-Oriented Policing and Crime Prevention* 123 (Criminal Justice 2002) (“Many problem-oriented policing interventions are multifaceted and, as such, it is complicated to evaluate a single strategy within a varied bundle of tactics because it is difficult to isolate each response’s effects.”) (internal citation omitted).

<sup>55</sup> Sampson and Cohen, 22 *L. & Soc’y Rev.* at 185 (cited in note 53) (concluding that a model allowing for empirical determinations of both the direct effects of police aggressiveness and the arrest/offense ratio on crime proved elusive).

<sup>56</sup> Jeffrey Fagan and Garth Davies, *Policing Guns: Order Maintenance and Crime Control in New York*, in Bernard E. Harcourt, ed., *Guns, Crime, and Punishment in America* 191, 205 (NYU 2003) (offering empirical evidence showing that no single enforcement measure in 1998 proved to be a significant predictor of arrest rates the following year).

<sup>57</sup> *Id.*

have reverse effects. They write, “When legal control engenders resistance, opposition or defiance, the opportunity to leverage formal social control into informal social control is lost. The absence of crime control returns from OMP [order-maintenance policing] may reflect just such a dynamic among African Americans, who shouldered much of the burden of OMP.”<sup>58</sup>

The final and most recent contribution to this literature is Steve Levitt’s 2004 *Journal of Economic Perspectives* review essay, in which he argues that policing practices probably do not explain much of the crime drop in the 1990s because crime went down everywhere, even in places where police departments did not implement new policing strategies.<sup>59</sup> Instead, Levitt attributes the massive period effects on crime throughout the United States during the 1990s to some combination of increased imprisonment, increases in the number of police, the ebbing of the crack epidemic that started in many big cities in the mid 1980s, and the legalization of abortion in the United States during the early 1970s.<sup>60</sup>

## II. NEW YORK CITY’S EXPERIENCE

In this Part we discuss the most recent studies on broken windows policing in New York City, both the 2001 Kelling and Sousa study and the evidence presented by Corman and Mocan in their 2005 paper in the *Journal of Law and Economics*. We argue that the 2001 Kelling and Sousa analysis has limitations that ultimately render it uninformative about the causal effects of broken windows policing practices. We also show that the Corman and Mocan analysis cannot support the claim that broken windows policing activities are responsible for declines in crime.

### A. Kelling and Sousa (2001)

The study by George Kelling and William Sousa (hereafter KS) fits in the larger tradition of studies of aggressive arrest policies discussed earlier. The goal of their study is the “systematic attempt to statistically parse out the relative contributions of police actions, the economy, demographics, and changing drug use patterns on crime” in New York City.<sup>61</sup> The major problem with previous studies, they argue, is that those studies lack an adequate comparison group for New York

---

<sup>58</sup> Id at 210.

<sup>59</sup> Steven D. Levitt, *Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not*, 18 J Econ Perspectives 163, 177 (2004).

<sup>60</sup> Id at 184.

<sup>61</sup> Kelling and Sousa, *Do Police Matter?* at 1 (cited in note 6).

City: previous research has either used an unsuitable comparison, such as other cities, or failed to use any comparison at all. The key insight in this study, Kelling and Sousa suggest, is to simulate comparison groups by treating the city as seventy-five separate and comparable entities. “Rather than one entity,” they explain, “we view New York as 76 separate ‘cities,’ corresponding to the 76 police precincts.”<sup>62</sup>

The research design, then, is to statistically compare the relationship between violent crime and four explanatory variables—broken windows policing, economic indicators, young male population shifts, and the decline in crack cocaine consumption—in the seventy-five precincts of New York City. They find a strong negative relationship between precinct-level misdemeanor arrests and violent crime. In what follows we reexamine these New York City results using a wide variety of alternative statistical approaches. Our efforts to obtain, replicate, and extend their data are discussed in detail in the Appendix.

Replicating the KS results is complicated in part by the fact that in neither the KS Manhattan Institute report nor Sousa’s dissertation<sup>63</sup> do the authors spell out the exact estimating equations for their analysis. Nor does their Table 4, which presents their key results, show the number of observations used to generate their estimates (to give some sense for how the analysis is structured).<sup>64</sup> Nevertheless, from reading over the discussion in KS and in Sousa’s dissertation it would appear that they are estimating a two-level hierarchical linear growth model, of the sort discussed in Chapter 6 of Raudenbush and Bryk’s *Hierarchical Linear Models*.<sup>65</sup> If we let level 1 in this model represent time (subscripted by  $t$ ) and level 2 represent precincts (subscripted by  $i$ ), we believe that the two-level linear growth model that they are estimating is given by the following equations:

- (1)  $VC_{ti} = \pi_{0i} + \pi_{1i} A_t + \varepsilon_{ti}$ ;
- (2)  $\pi_{0i} = \alpha_{00} + \alpha_{01} MA_i + \alpha_{02} X_i + r_{0i}$ ; and
- (3)  $\pi_{1i} = \alpha_{10} + \alpha_{11} MA_i + \alpha_{12} X_i + r_{1i}$ ,

<sup>62</sup> Id at 4 & n 27 (explaining that in 1994, one precinct was divided into two, resulting in seventy-six precincts existing today but that to maintain consistency over the studied period the original seventy-five precincts were used).

<sup>63</sup> See generally, William H. Sousa, Jr., *Crime Reduction in New York City: The Impact of the N.Y.P.D.*, unpublished PhD dissertation, Rutgers University (2003), online at <http://wwwlib.umi.com/dissertations/fullcit/3088010> (visited Jan 19, 2006).

<sup>64</sup> Kelling and Sousa, *Do Police Matter?* at 9 (cited in note 6).

<sup>65</sup> Stephen W. Raudenbush and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods* 161 (Sage 2d ed 2002) (“The development of hierarchical linear models has created a powerful set of techniques for research on individual change. When applied with valid measurements from a multi-time-point design, these models afford an integrated approach for studying the structure and predictors of individual growth.”).

where

$VC_{ti}$  = violent crimes in precinct (i) in year (t);

$A_t$  = time (1989, 1990, . . . 1998);

$MA_i$  = precinct (i)'s average misdemeanor arrests over the sample period; and

$X_i$  = average value of other covariates for precinct (i) over the sample period.

The empirical setup that is being estimated by KS is easier to see by substituting equations (2) and (3) into (1) to get the reduced-form estimating equation (4):

$$(4) \quad VC_{ti} = \beta_1 + \beta_2 MA_i + \beta_3 A_t + \beta_4 MA_i * A_t + \varepsilon_{ti}.$$

The intuition behind what Kelling and Sousa are doing here is more straightforward than the statistical notation and equations might suggest: do police precincts with relatively higher levels of average misdemeanor arrests during the 1990s (MA) experience relatively larger declines in violent crime (VC) during this period? In statistical terms, the coefficient  $\beta_2$  will be positive if police activity (measured here by misdemeanor arrests) is relatively greater in higher-crime areas. Since crime rates were declining throughout the United States during the 1990s we expect the overall linear trend among all New York City precincts to be declining ( $\beta_3$  will be negative). The key empirical question of interest for the KS hypothesis is whether the estimate for the coefficient  $\beta_4$  in equation (4) is negative—that is, is the decline in violent crime larger in precincts with relatively higher numbers of average misdemeanor arrests?

We can replicate the key coefficient in their analysis ( $\beta_4$  or, equivalently,  $\alpha_{11}$ ) as shown in the first row of Table 1, where we estimate equation (4) measuring all of our variables in precinct counts (rather than per capita rates) and do not weight by precinct population. Note that as shown in Table 1, these estimates are not very sensitive to decisions about whether to weight by precinct population or not, or to work in per capita crime and arrest rates rather than counts. Note also that the coefficient and standard error for the effects of misdemeanor arrest rates on the time slope in violent crimes, which is the key estimate of interest, is identical to what is reported in KS Table 4,<sup>66</sup> although our point estimates for the intercept terms have a slightly different scaling.

<sup>66</sup> Kelling and Sousa, *Do Police Matter?* at 9 (cited in note 6).

TABLE 1  
Replicating Kelling and Sousa's Multilevel Model  
with a Reduced-Form Single Equation Model

| Model specification:<br>dependent variable = | Coefficient on<br>MA | Coefficient on<br>MA*A |
|----------------------------------------------|----------------------|------------------------|
| Crime counts,<br>not population weighted     | 72.68 (5.94)         | -.036 (.003)           |
| Crime counts,<br>population weighted         | 70.06 (13.20)        | -.035 (.007)           |
| Crime rates,<br>not population weighted      | 509.95 (0.27)        | -.255 (.0001)          |
| Crime rates,<br>population weighted          | 139.02 (76.56)       | -.070 (.038)           |

NOTES: Standard errors are in parentheses. Each row in Table 1 represents the results from estimating a separate regression of the form  $VC_{it} = \beta_1 + \beta_2 MA_{it} + \beta_3 A_{it} + \beta_4 MA_{it} * A_{it} + \varepsilon_{it}$  where VC = violent crimes for precinct (i) in year (t), MA = misdemeanor arrests for precinct (i) in year (t), and A = year (ranging from 1989 to 1998). See text for additional details.

Kelling and Sousa conclude from these results that the broken windows strategy is highly effective at crime fighting. The bottom line: "The average NYPD precinct during the ten-year period studied could expect to suffer one less violent crime for approximately every 28 additional misdemeanor arrests made."<sup>67</sup> This, Kelling and Sousa suggest, offers "the most-definitive possible answer to the question of whether police mattered in New York City during its intense crime-drop."<sup>68</sup>

Our conclusion from these results is somewhat different, and it points in the direction of mean reversion. Any study of the influences on American crime patterns during the past twenty years is complicated by the massive period effects that have generated dramatic year-to-year changes in crime across the country. The increase in crime rates was particularly dramatic from the mid 1980s through the early-to-mid 1990s, which is thought to have been driven largely by the growth in crack cocaine use and involvement of firearms in the new street markets for crack.<sup>69</sup> Using city-level data, Steven Raphael and

<sup>67</sup> Id.

<sup>68</sup> Id at 1.

<sup>69</sup> See Alfred Blumstein, *Youth Violence, Guns, and the Illicit-Drug Industry*, 86 J Crim L & Criminology 10, 10 (1995) (examining some empirical data reflecting changing crime patterns beginning in the mid 1980s and concluding that the illegal drug markets' recruitment of youths resulted in a dramatic growth in youth homicide). See also Philip J. Cook and John H. Laub, *After the Epidemic: Recent Trends in Youth Violence in the United States* 22 (NBER Working Paper 8571, Oct 2001), online at <http://www.nber.org/papers/w8571> (visited Jan 19, 2006) (dis-

one of us show that those cities that experienced the largest increases in crime during this period subsequently also experienced the largest crime drops.<sup>70</sup> A natural concern is that the same process may be at work at the neighborhood or police precinct level as well.

Figure 1 suggests that crime patterns across New York precincts that KS attributes to the effects of broken windows policing can be explained by mean reversion: broken windows policing (as measured by misdemeanor arrests) was conducted most intensively in New York within the city's most violent neighborhoods, which are the areas that experienced the largest increases in violent crime during the 1980s and the largest declines in violent crime during the 1990s. Panel A shows that at the start of the KS panel (1989) precincts with higher violent crime rates also have higher rates of misdemeanor arrests. That is, the regression line relating violent crime and misdemeanor arrests in 1989 has a positive slope, consistent with Kelling and Sousa's own findings (top panel of their Table 4).<sup>71</sup> Panel B shows that the most violent precincts in 1989 also experienced the largest *increase* in misdemeanor arrests from 1989 to 1998. Panel C shows that the neighborhoods with the highest violent crime rates in 1989 experienced the largest declines in such crimes from 1989 to 1998.

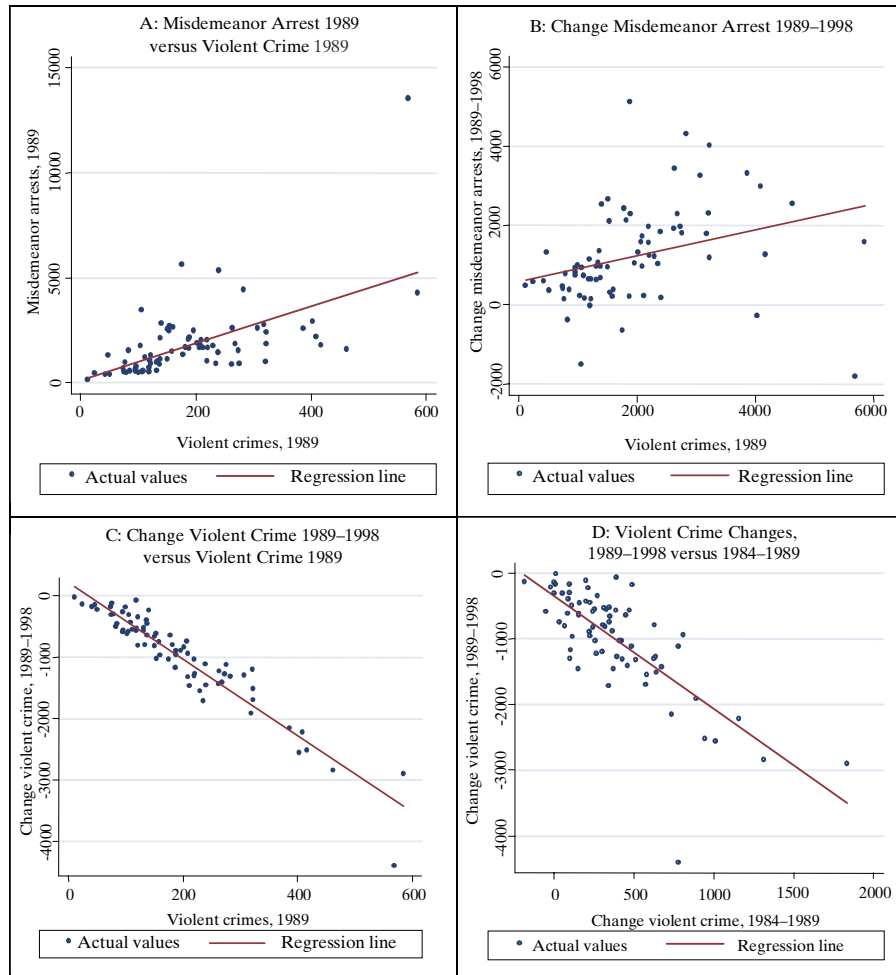
---

cusssing epidemics of youth violence in different time periods and concluding that the observed youth violence of the late 1980s was closely tied to the epidemic of crack cocaine).

<sup>70</sup> Raphael and Ludwig, *Prison Sentence Enhancements* at 267 (cited in note 16) ("To summarize, the large increase in homicide rates occurring during the late 1980s in Richmond coupled with the inverse relationship between earlier and later changes in homicide rates observed among other U.S. cities casts doubt on the validity of previous claims about the effects of Project Exile.").

<sup>71</sup> Kelling and Sousa, *Do Police Matter?* at 9 (cited in note 6).

FIGURE 1  
Misdemeanor Arrests and Violent Crime  
in NYC Precincts, 1989–1998



Why do precincts with unusually high initial crime rates experience unusually large declines in crime thereafter? Mean reversion is a good candidate. Panel D shows that, as is true with city-level crime data, those police precincts with the largest increases in crime during the crack epidemic have the largest declines thereafter. Most criminologists believe that this increase in violent crime was driven by the crack cocaine epidemic and attendant violence in the crack market, which began to ebb during the early 1990s. Hence those places where crack served to drive violent crime to unusually high levels at the height of the epidemic would be expected to experience the largest

subsequent declines as the influence on violence from crack use and distribution begin to wane.

The KS analysis seems particularly susceptible to confounding from mean reversion because their model basically relates *changes* in violent crimes (each precinct's linear trend in violent crime over the 1989–1998 period) to the *levels* of misdemeanor arrests (average arrests from 1989 to 1998). Put differently, their analysis throws away all of the over-time variation in misdemeanor arrests across precincts from 1989 to 1998, and simply relates variation in the linear trend in violent crime rates across precincts to variation in the average number of misdemeanor arrests over this period.<sup>72</sup> The level of misdemeanor arrests is strongly related to the initial level of violent crimes, as suggested by Figure 1, which may lead to a spurious association between misdemeanor arrests and violent crimes in their study.<sup>73</sup>

Table 2 presents the results of a more formal analysis that seems to implicate mean reversion:

<sup>72</sup> In this sense their two-level linear growth model is set up in a fashion analogous to Raudenbush and Bryk's 2002 example, which related changes in student's test scores measured four times each year over several years with the total hours of instruction the child received. See Raudenbush and Bryk, *Hierarchical Linear Models* at 167 (cited in note 65). But the time trend in the key treatment variable of interest in the policing example seems to matter much more than in the schooling example offered by Raudenbush and Bryk.

<sup>73</sup> The problem of relating levels against changes can be illustrated with a simple hypothetical example:

| Precinct | Year | MA  | VC  |
|----------|------|-----|-----|
| 1        | 1989 | 150 | 500 |
| 1        | 1990 | 100 | 400 |
| 1        | 1991 | 50  | 300 |
| 2        | 1989 | 75  | 500 |
| 2        | 1990 | 50  | 475 |
| 2        | 1991 | 25  | 450 |

Precinct 1 has a higher mean number of misdemeanor arrests over the sample period than does Precinct 2 (100 versus 50), and also experiences a larger decline in violent crimes per year (100 per year compared to only 25 in Precinct 2). The Kelling-Sousa model applied to these data would suggest a negative relationship between misdemeanor arrests (MA) and the time trend in violent crime (VC) across precincts—more misdemeanor arrests, less crime. However regressing changes against changes—the change over time in violent crimes against the change in misdemeanor arrests—would yield the opposite conclusion.

TABLE 2  
Effects of Model Specification and Mean Reversion  
in the Kelling-Sousa Analysis: Regressing  
Crime Changes against Arrest Levels

| Explanatory variables:                 | Model 1           | Model 2            | Model 3           | Model 4           | Model 5           | Model 6           |
|----------------------------------------|-------------------|--------------------|-------------------|-------------------|-------------------|-------------------|
| Average misdemeanor arrests, 1989–1998 | -.303**<br>(.035) | -.221**<br>(.023)  | -.079**<br>(.019) | -.082**<br>(.022) | -.101**<br>(.019) | -.031<br>(.024)   |
| Violent crime, 1989                    |                   |                    | -.546**<br>(.029) | -.524**<br>(.057) | -.528**<br>(.048) | -.576**<br>(.055) |
| Change violent crimes, 1984–1989       |                   | -1.338**<br>(.124) |                   | -.069<br>(.162)   | -.053<br>(.137)   | -.097<br>(.140)   |
| Change manpower, 1989–1998             |                   |                    |                   |                   | 4.070**<br>(.763) | 3.786**<br>(.944) |
| Other covariates?                      | N                 | N                  | N                 | N                 | N                 | Y                 |
| N                                      | 75                | 74                 | 74                | 74                | 74                | 74                |
| R-squared                              | .504              | .811               | .915              | .914              | .939              | .970              |

Dependent variable = Precinct change violent crimes, 1989–1998. Other covariates include change from 1989 to 1998 in poverty, racial and age composition of the population, percent households headed by females, public assistance, vacant housing. For models 2 through 6, we are using 1984 crime data for one precinct. \* = Statistically significant at 10 percent cut-off. \*\* = Statistically significant at 5 percent cut-off.

The first row of Table 2 presents estimates for the parameters in equation (5) below, where the change in violent crimes within a precinct for the period 1989 to 1998 is regressed against the average number of misdemeanor arrests within that precinct over the entire 1989 to 1998 period. This simple model is based on the same intuition as the hierarchical linear modeling (HLM) linear growth model of KS, although the key difference is that our dependent variable is the actual change in violent crimes from 1989 to 1998 for each precinct rather than each precinct's estimated linear trend in violent crimes over this period. (The choice by KS to fit a linear trend through these violent crime counts for each precinct is itself a bit puzzling given that Appendix Figure 1 in our paper and Figure 1 in their Manhattan Institute report<sup>74</sup> show a nonlinear trend in such crimes in New York over

<sup>74</sup> Kelling and Sousa, *Do Police Matter?* at 6 (cited in note 6).

this period, first increasing for a few years and then declining thereafter.) The average number of misdemeanor arrests within these precincts has a strong negative relationship with the change in violent crime rates over this period, as with the basic results presented by KS.

$$(5) \Delta VC_i = \lambda_1 + \lambda_2 MA_i + v_i$$

The remaining rows of Table 2 show that controlling for either the precinct's 1989 violent crimes or change from 1984 to 1989 in violent crimes reduces the coefficient on the average misdemeanor arrest variable by more than two-thirds. The reason is suggested by our Figure 1: the average number of misdemeanor arrests over the 1989–1998 period is highest in those precincts that experienced the largest increases in crime from 1984 to 1989 and had the largest number of violent crimes in 1989. Statistically relating the average number of misdemeanor arrests from 1989 to 1998 with the decline in violent crimes over this period without controlling for differences across precincts in the run-up in violent crime they experienced during the crack epidemic mistakenly attributes the influence of these initial conditions and subsequent mean reversion to the average number of misdemeanor arrests. Unfortunately, none of the proxies for crack, including the borough-level measure of cocaine-related hospital discharges used by Kelling and Sousa, seem to adequately capture the influence of crack markets and use on crime. For example, the cocaine proxy used by KS does not have a statistically significant relationship to violent crime rates in their own analysis,<sup>75</sup> nor is this variable statistically significant when included in our own models (and by implication does not change any of the other results shown in our Table 2, either).

In contrast to the weak explanatory power of the KS proxy for crack-related violence—admittedly an extremely difficult phenomenon to quantify—the final row of our Table 2 shows that controlling for the set of detailed precinct-level covariates in our dataset yields an estimated relationship between the change in violent crime and the 1989–1998 average number of misdemeanor arrests that is about 10 percent as large as the baseline estimate and no longer statistically significant. These covariates include measures of structural disadvantage (such as the percent of the precinct that is poor, receiving public assistance, or has less than a high school degree), demographics (percent of the precinct in their peak offending ages, percent of households headed by a female, percent black), measures of physical disorder-

<sup>75</sup> Id at 9.

der (percent of housing units that are vacant), and police manpower assigned to the precinct.<sup>76</sup>

$$(6) \Delta VC_i = \lambda_1 + \lambda_2 \Delta MA_i + v_{ti}$$

Now suppose we instead use the within-precinct over-time variation in the data by relating *changes* in violent crime rates from 1989 to 1998 to *changes* over this period in misdemeanor arrests, as in equation (6) above. The results from this analysis, shown in Table 3 below, suggest that if anything, increases in misdemeanor arrests are accompanied by *increases* in violent crime—more misdemeanor arrests, more crime. While the positive relationship between changes in misdemeanor arrests and changes in violent crime is somewhat sensitive to the model specification, there is no evidence from this first-difference model of a *negative* relationship between changes in misdemeanor arrests and violent crime. The expectation that violent crime should decline in response to an increase in misdemeanor arrests is the key empirical prediction of the argument that broken windows policing is effective. Although the Kelling-Sousa analysis does not directly test this prediction, our own analysis shown in Table 3 demonstrates that the data are not consistent with the idea that stepped-up zero-tolerance policing reduces crime.

---

<sup>76</sup> The police manpower variable is potentially problematic because some arrests within a precinct might be made by law enforcement officers who are officially assigned to different areas, although our results are not sensitive to excluding this variable. Adding just a control for the percent of the precinct's population that is black to the baseline model in the first column of Table 2 reduces the coefficient on average number of misdemeanor arrests from -0.30 to -0.28. Including the Kelling and Sousa measures of cocaine-related hospital discharges and borough-level unemployment rates has little effect on the results shown in Table 2.

TABLE 3  
The Effects of Model Specification and Mean Reversion in the  
Kelling-Sousa Analysis: Regressing Crime Changes  
Against Arrest Changes

| Explanatory variables:                 | Model 1         | Model 2            | Model 3           | Model 4           | Model 5           | Model 6            |
|----------------------------------------|-----------------|--------------------|-------------------|-------------------|-------------------|--------------------|
| Average misdemeanor arrests, 1989–1998 | -.086<br>(.074) | -.046<br>(.051)    | -.114**<br>(.022) | -.114**<br>(.022) | -.094**<br>(.025) | -.004<br>(.030)    |
| Violent crime, 1989                    |                 |                    | -.660**<br>(.023) | -.710**<br>(.039) | -.716**<br>(.039) | -.625**<br>(.041)  |
| Change violent crimes, 1984–1989       |                 | -1.762**<br>(.183) |                   | -.214<br>(.133)   | .243*<br>(.137)   | -.013<br>(.127)    |
| Change manpower, 1989–1998             |                 |                    |                   |                   | 1.412<br>(.963)   | 3.326**<br>(1.065) |
| Other covariates?                      | N               | N                  | N                 | N                 | N                 | Y                  |
| N                                      | 75              | 74                 | 74                | 74                | 74                | 74                 |
| R-squared                              | .018            | .561               | .924              | .926              | .928              | .969               |

Dependent variable = Precinct change violent crimes, 1989–1998. Other covariates include change from 1989 to 1998 in poverty, racial and age composition of the population, percent households headed by females, public assistance, and vacant housing. \* = Statistically significant at 10 percent cut-off. \*\* = Statistically significant at 5 percent cut-off.

#### B. Corman and Mocan (2005)

But even putting aside these precinct comparisons, for many observers the massive drop in New York City's crime rate during the 1990s—coincident with the onset of broken windows policing in the city—alone provides compelling proof of the efficacy of this policing strategy. Corman and Mocan's analysis provides a more formal version of this same insight, by analyzing monthly time-series data for New York City as a whole. Controlling for city-wide measures of New York's unemployment rate, real minimum wage, incarceration rate, police manpower, number of fourteen to sixteen year olds and lagged values of monthly crime rates, they find a negative relationship between city-wide misdemeanor arrest rates and city-wide robbery and motor vehicle theft rates. They do not find a relationship between the former and other types of crime. While Corman and Mocan's time series uses data from 1970 to 2000, graphs of their data suggest that the relationship between misdemeanor arrests and crime would ap-

pear to be driven by the unusually large increase in misdemeanor arrests that occurred in New York during the mid-to-late 1990s.<sup>77</sup>

What can we conclude about the causal effects on crime of broken windows policing—at least as measured by misdemeanor arrests? Research designs that rely on time-series data for a single jurisdiction (in their case, New York) typically provide weak power to rule out alternative explanations for the patterns observed in the data. For example, consider just one candidate counterexplanation, what we term the “Broken Yankees Hypothesis” (BYH). When the New York Yankees do well, violence should decline through the strengthened social ties that develop by the bonding that occurs among the city’s residents at local bars and restaurants, with much of the city’s attention focused on a single, shared goal. When the Yankees do poorly, residents may be less likely to aggregate together for a common purpose in communal settings, and the team’s poor performance may even spur dissension among New Yorkers over the causes of these failures.

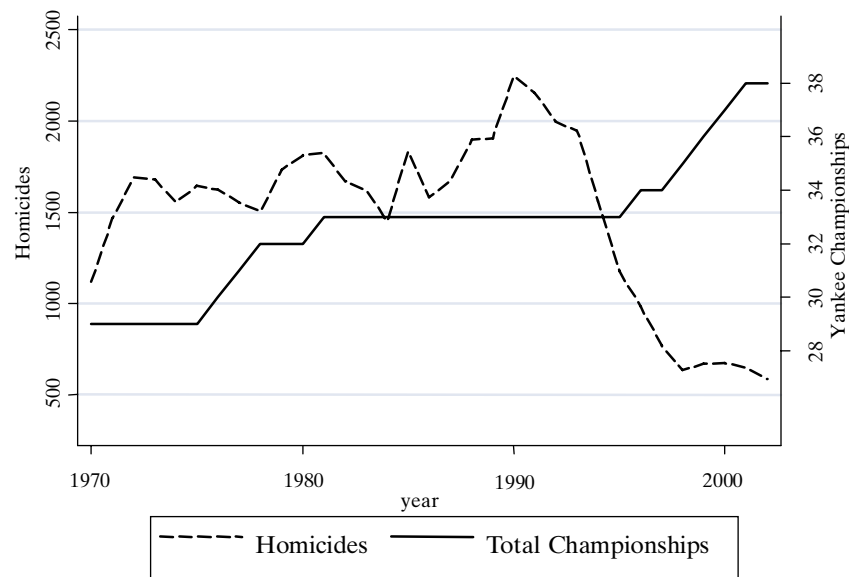
Although Corman and Mocan were not willing to share their monthly time-series data with us, we were able to construct on our own an annual time series for New York measuring crime rates and a reasonable proxy for the operational mechanism behind the Broken Yankees Hypothesis, defined as the cumulative number of World Series championships dating back to 1921.<sup>78</sup> Figure 2 provides what appears to be some empirical support for the BYH: the strong performance of Billy Martin’s Yankees teams during the late 1970s coincides with a drop in homicides, but even more striking is the massive decline in homicides that accompanies the consistent excellence of Joe Torre’s squads beginning in the late 1990s. A time-series regression of the homicide rate against the BYH index and lags of the murder variable frequently yields a negative and statistically significant coefficient (and even controlling for lagged values of robbery to proxy for other criminogenic characteristics), although we note that the magnitudes of the point estimate and standard error are somewhat sensitive to the choice of lag length.

---

<sup>77</sup> Hope Corman and Naci Mocan, *Carrots, Sticks and Broken Windows* 28 (NBER Working Paper 9061, July 2002), online at <http://www.nber.org/papers/w9061> (visited Jan 19, 2006) (showing that after hovering between 9,000 and 14,000 between 1982 and 1994, the number of misdemeanor arrests in New York City nearly doubled from 1994 to 2000). See also Corman and Mocan, 48 J L & Econ at 244 (cited in note 14).

<sup>78</sup> These data come from the New York Yankees website. Major League Baseball, *Yankees Championship Clubs* (New York Yankees 2005), online at <http://newyork.yankees.mlb.com/NASApp/mlb/nyy/history/championships.jsp> (visited Jan 19, 2006).

FIGURE 2  
NYC Homicides versus Yankee Championships



Although our simple empirical example is not intended to provide a rigorous test of the Broken Yankees Hypothesis, it does serve to highlight the vulnerability of single-city time-series findings to counterexplanations. An equally or perhaps even more plausible counterexplanation for New York City's crime pattern during the 1990s comes from the dramatic period effects that caused crime to decline almost everywhere throughout the United States during this period, even in cities that did not adopt innovative policing strategies.<sup>79</sup>

We have, in sum, tried to demonstrate that the correlations that have been reported in previous research between misdemeanor arrests and crime rates in New York City cannot provide evidence for a

<sup>79</sup> See Levitt, 18 *J Econ Perspectives* at 163 (cited in note 59). Levitt argues that crime declined throughout the United States during the 1990s due to some combination of more police, increased incarceration, the ebbing of the crack epidemic that is widely thought to have caused violent crimes to increase during the late 1980s, and legalization of abortion during the early 1970s. *Id.* at 186 ("Other factors often cited as important factors driving the decline do not appear to have played an important role: the strong economy, changing demographics, innovative policing strategies, gun laws and increased use of capital punishment."). While we find Levitt's explanation persuasive, accepting the specific bundle of causal factors implicated by Levitt is not crucial to our argument for a skeptical interpretation of Corman and Mocan's findings. One need only accept Levitt's observation that crime dropped everywhere over this period to accept the importance of common period effects in understanding crime drops during the 1990s.

*causal* effect of zero-tolerance policing on crime. It is true, of course, that disentangling this causal relationship is a difficult task given the wide variety of confounding factors that lead crime rates to vary so dramatically across areas over time. Nevertheless, from a public policy perspective, the faith that many policymakers place in the efficacy of broken windows policing is in the end just faith, rather than the result of convincing empirical evidence. In the next Part we try to overcome some of these inferential difficulties by examining the effects on crime of the basic social process that motivates zero-tolerance policing and that is at the heart of the broken windows hypothesis—disorder.

### III. EVIDENCE FROM THE MOVING TO OPPORTUNITY EXPERIMENT

Suppose that we could design the ideal social experiment to test the effects of disorder *alone* on criminal behavior. We would start with a sample of people who were at high risk for criminal offending and were living in very socially disordered communities. We would then randomly assign some of these families, but not others, to neighborhoods that were less disorderly—ideally, much less disorderly, so that the “treatment dose” that families experience from neighborhood moves would be large enough to yield statistically detectable impacts on behavior. In this idealized experiment we would then wish to follow participants for many years, measure their involvement in criminal activity in different ways (for example with both self reports and administrative arrest records) as well as characteristics of their neighborhoods, and be careful to minimize sample attrition.

The MTO experiment, launched in 1994 by HUD, conforms in most ways to the parameters of the ideal experiment described above, with the one exception being that MTO changes other neighborhood characteristics for program participants as well. In what follows, we provide a review of the effects of MTO on criminal offending by program participants about five years after random assignment, and discuss their implications for ongoing debates about broken windows policies.<sup>80</sup>

We show that MTO succeeds in moving families to neighborhoods that are characterized by much lower levels of both physical and social disorder—arguably a more relevant “treatment indicator” for measuring the broken windows hypothesis compared to more indirect policy levers such as misdemeanor arrests that may or may not

---

<sup>80</sup> These results have previously been reported in greater technical detail. See generally Kling, Ludwig, and Katz, 120 Q J Econ 87 (cited in note 19). See also generally Jens Ludwig, Jeffrey R. Kling, and Maria J. Hanratty, *Neighborhood Effects on Crime over the Life Cycle* (Georgetown University Public Policy Institute Working Paper, 2004) (on file with authors).

succeed in reducing disorder. However, we also show that the findings from MTO are not consistent with the idea that change in neighborhood disorder is enough to change criminal activity. At the very least, MTO helps bound the size of the effect on crime that could result from reducing disorder: such disorder impacts cannot be any larger than whatever pernicious effects on criminal behavior arise from some increase in neighborhood characteristics related to affluence.

#### A. Background on MTO

Sponsored by HUD, MTO has been in operation since 1994 in five cities: Baltimore, Boston, Chicago, Los Angeles, and New York. Eligibility for the program was restricted to low-income families with children, living in public or Section 8 project-based housing in selected high-poverty census tracts.<sup>81</sup> The approximately 4,600 families who volunteered for the program from 1994 to 1997 were randomly assigned into one of three groups. The Experimental group was offered the opportunity to relocate using a housing voucher that could only be used to lease a unit in census tracts with 1990 poverty rates of 10 percent or less.<sup>82</sup> Movers through MTO were required to stay in these tracts for at least one year. Experimental group families were also provided with mobility assistance and in some cases other counseling services as well. Families assigned to the Section 8 group were offered housing vouchers with no constraints under the MTO program design on where the vouchers could be redeemed. Families assigned to the Control group were offered no services under MTO, but did not lose access to social services to which they were otherwise entitled, such as public housing.

Because of random assignment, MTO yields three comparable groups of families living in very different kinds of neighborhoods during the postprogram period. This random assignment helps overcome the self-selection problem that is very likely to plague most previous

---

<sup>81</sup> See Orr, et al, *Moving to Opportunity Interim Impacts Evaluation* at i–ii (cited in note 17). Section 8 project-based housing might be thought of as essentially privately-operated public housing. Edgar O. Olsen, *Housing Programs for Low-Income Households*, in R. Moffit, ed, *Means-Tested Transfer Programs in the United States* 366, 371–74 (Chicago 2003) (explaining that HUD contracts with private providers to develop and manage housing projects that include units reserved for low-income families).

<sup>82</sup> Housing vouchers provide families with subsidies to live in private-market housing. The subsidy amount is typically defined as the difference between 30 percent of the household's income and the HUD-defined Fair Market Rent, which ranges between the fortieth and fiftieth percentiles of the local area rent distribution. See HUD, *Fair Market Rents: Increased Fair Market Rents and Higher Payment Standards for Certain Areas*, 65 Fed Reg 58870, 58870 (2000). For a general overview of the history and implementation of housing vouchers under Section 8, see Olsen, *Housing Programs for Low-Income Households* at 368–86 (cited in note 81).

studies of “neighborhood effects” in general or “broken windows” in particular.

The results summarized below from Kling, Ludwig and Katz (2005) and Ludwig and Kling (2005) measure the delinquency and criminal behavior of youth in MTO using two main sources: survey data and administrative arrest records. Adults were surveyed, but they were not asked about criminal behavior, so we can only measure adult criminal activity using official arrest records. Information on potential mediating processes that could lead to these outcomes comes from the surveys as well as administrative data on local-area crime rates.<sup>83</sup>

The families in the main survey sample enrolled in the MTO demonstration from 1994 to 1997. At the time of enrollment, the head of household completed a baseline survey that included information about the family as well as some specific information about each child. Descriptive statistics for the baseline characteristics of youth and adults are shown in Table 4 below. Overall about two-thirds of MTO participants are black, with the program populations in Chicago and Baltimore almost entirely black and an even mix between black and Hispanic in the other sites. MTO households are quite poor, with around three-quarters having been on welfare at baseline. One-quarter of household heads had their first child before the age of eighteen, and only a little more than half of all household heads had a GED or high school diploma. Around three-quarters of households report gangs and drugs as the first or second most important reason that they enrolled in the MTO program, while around one-half report access to better schools as one of their top two reasons. Eligibility for the MTO program was limited to families in public housing or Section 8 project-based housing located in some of the most disadvantaged census tracts in the five MTO cities and, for that matter, in the country as a whole.

---

<sup>83</sup> For more detail on these data sources see Kling, Ludwig, and Katz, 120 Q J Econ at App 1 (cited in note 19) (providing survey data, local area crime rate data, and administrative data on arrests).

2006]

*Broken Windows: New Evidence*

303

TABLE 4  
Baseline Descriptive Statistics for MTO Adult and Youth Samples

|                                 | FEMALES<br>Exp | S8     | Control | MALES<br>Exp | S8    | Control |
|---------------------------------|----------------|--------|---------|--------------|-------|---------|
| <b>ADULTS</b>                   |                |        |         |              |       |         |
| Black                           | .650           | .646   | .657    | .359         | .364  | .386    |
| Hispanic                        | .294           | .297   | .298    | .505         | .494  | .487    |
| MTO site:                       |                |        |         |              |       |         |
| Baltimore                       | .150           | .162   | .147    | .039         | .071  | .051    |
| Boston                          | .229           | .223   | .221    | .211         | .192* | .287    |
| Chicago                         | .209           | .209   | .210    | .149         | .128  | .131    |
| LA                              | .155           | .149   | .158    | .304         | .351  | .345    |
| NYC                             | .257           | .257   | .264    | .297**       | .259  | .185    |
| HH on<br>AFDC at<br>baseline    | .739           | .752   | .756    | .579         | .586  | .491    |
| Moved<br>because:               |                |        |         |              |       |         |
| Drugs, crime                    | .767           | .755   | .783    | .739         | .755  | .764    |
| Schools                         | .468           | .521** | .465    | .469         | .577  | .489    |
| Age at end<br>of 2001           | 38.96          | 39.40  | 39.13   | 43.00        | 43.39 | 44.84   |
| Any pre-<br>RA arrest           | .258           | .231   | .260    | .375         | .423  | .354    |
| Missing<br>admin arrest<br>data | .038           | .054   | .035    | .056         | .048  | .057    |
| N                               | 1,483          | 1,013  | 1,102   | 224          | 153   | 166     |
| <b>YOUTH</b>                    |                |        |         |              |       |         |
| Black                           | .647           | .606   | .640    | .609         | .605  | .612    |
| Hispanic                        | .296           | .318   | .304    | .329         | .333  | .339    |
| MTO site:                       |                |        |         |              |       |         |
| Baltimore                       | .168           | .138   | .140    | .151         | .154  | .139    |
| Boston                          | .187           | .192   | .216    | .166         | .200  | .189    |
| Chicago                         | .210           | .215   | .203    | .220         | .209  | .205    |
| LA                              | .165           | .185   | .199    | .195         | .189  | .196    |
| NYC                             | .270           | .271   | .242    | .269         | .248  | .270    |
| HH on<br>AFDC at<br>baseline    | .732           | .744   | .749    | .743         | .706  | .727    |
| Moved<br>because:               |                |        |         |              |       |         |
| Drugs, crime                    | .807           | .732   | .782    | .780         | .760  | .791    |
| Schools                         | .460           | .524   | .483    | .511         | .549  | .505    |
| Age at end<br>of 2001           | 19.05          | 18.90  | 18.90   | 19.02        | 18.86 | 18.96   |
| Any pre-<br>RA arrest           | .062           | .041   | .048    | .147         | .122  | .131    |
| Missing<br>admin arrest<br>data | .057           | .048   | .055    | .059         | .063  | .061    |
| N                               | 966            | 651    | 716     | 988          | 691   | 739     |

Source: Ludwig and Kling (2005). NOTES: \* = Difference with Control mean statistically significant at 10 percent cut-off. \*\* = Difference with Control mean statistically significant at 5 percent cut-off.

Consistent with random assignment of families to MTO groups, Table 4 shows that there are no statistically significant differences

across MTO groups in the fraction of male or female adults or youth who have ever been arrested prior to random assignment or for other baseline characteristics. These results, together with those presented elsewhere, suggest that MTO random assignment was in fact random.<sup>84</sup>

Of the families with youth in the survey sample (aged fifteen to twenty years old at the end of 2001), 44 percent of those in the Experimental group and 57 percent of those in the Section 8 group complied with the treatment (that is, relocated through MTO). These moves lead to substantial differences in neighborhood attributes across treatment groups, as seen in Table 5 below. Four years after random assignment, the average census tract poverty rate (from the 2000 Census) for families assigned to the Section 8 group was 18 percent lower than that of the Control group, while families assigned to the Experimental group had average census tract poverty rates 24 percent below those of Controls. Assignment to either the Experimental or Section 8 groups reduces local-area (police precinct) violent crime rates by 13–15 percent compared to Controls, with proportionally smaller effects on property crime rates. Given the changes in tract poverty rates induced by MTO, it is surprising that the program engenders so little residential integration with respect to race. The average family in all three MTO groups lives in a census tract where the large majority of residents are also members of racial or ethnic minorities.

---

<sup>84</sup> Id.

TABLE 5  
Effects of Moving to Opportunity Random Assignment on  
Community Disorder and Other Neighborhood Characteristics

|                                                                                    | Control<br>– all | Exp –<br>all | Exp –<br>movers | S8 –<br>all | S8 –<br>movers |
|------------------------------------------------------------------------------------|------------------|--------------|-----------------|-------------|----------------|
| Neighborhood characteristics, four years after randomization (all five MTO sites): |                  |              |                 |             |                |
| Avg tract poverty rate                                                             | 41.68            | 31.70        | 19.24           | 34.38       | 28.49          |
| % in tract with pov 0 – .2                                                         | 13.43            | 33.75        | 65.05           | 21.91       | 29.74          |
| % in tract with pov .2 – .4                                                        | 33.73            | 33.90        | 26.81           | 42.46       | 51.61          |
| % in tract with pov > .4                                                           | 52.83            | 32.35        | 8.14            | 35.63       | 18.65          |
| Avg tract black                                                                    | 53.93            | 53.27        | 41.29           | 52.05       | 50.79          |
| Avg tract minority                                                                 | 89.27            | 84.20        | 73.94           | 87.83       | 85.08          |
| Violent crime rate per 10,000                                                      | 235              | 204          | 128             | 200         | 201            |
| Property crime rate per 10,000                                                     | 513              | 491          | 373             | 463         | 508            |
| Adult survey reports on neighborhood in 2002 (all five sites)                      | .65              | .53          | .43             | .57         | .58            |
| Neighbors would not likely do something about truant children                      | .47              | .36          | .26             | .41         | .40            |
| Neighbors would not likely do something about spraying or graffiti                 | .48              | .38          | .19             | .40         | .32            |
| Problem in neighborhood with police not coming when called                         | .33              | .22          | .11             | .27         | .23            |
| Misdemeanor arrest rate, 4 years after randomization (NYC site only)               | 6838             | 5294         | 3587            | 5758        | 4428           |

Note: Panel on adult survey reports from Kling, Ludwig, and Katz (2005), for adults with youth ages fifteen to twenty-five at the end of 2001.

The bottom panel of Table 5 presents results from surveys of MTO adults conducted from four to seven years after random assignment about their perceptions of physical and social disorder within their neighborhoods, as well as the quality of local policing. Adults assigned to the Experimental or Section 8 groups are less likely than Controls to report that neighbors would fail to get involved if local youth were truant or engaging in delinquency (spray-painting graffiti). The next row shows that adults in the Experimental and Section 8 group also report less physical disorder as well compared to the re-

ports of adults in the Control group, as measured by the fraction that report that graffiti is a problem in the neighborhood.

A 2003 study prepared for HUD demonstrates that MTO reduces a wide variety of other self-reported measures of neighborhood social and physical disorder as well for both the Experimental and Section 8 groups relative to Controls, including 20–30 percent increases in the fraction who feel safe in their neighborhood at night, one-quarter reductions in the share who saw drugs in their neighborhood the preceding thirty days, 10–15 percent declines in the share who report problems with litter, trash, graffiti, or abandoned buildings in the neighborhood, 15–25 percent declines in the share who report problems with public drinking or groups of people hanging out in public spaces, and 10–25 percent increases in the share who are satisfied or very satisfied with their neighborhoods.<sup>85</sup>

The last row of Table 5 highlights the potential problems with the key explanatory variable used in the 2001 KS study, namely the police precinct misdemeanor arrest rate. These data are available for New York but not the other MTO sites. The final row of Table 5 shows assignment to either the Experimental or Section 8 groups substantially *reduces* the local misdemeanor arrest rate compared to the neighborhoods in which the Control group resides.<sup>86</sup> Yet the survey data reported by the MTO participants reveal that Experimental or Section 8 assignment also reduces social and physical disorder. This fact reinforces the notion that measures of zero-tolerance policing such as misdemeanor arrests may not be very informative about variation across neighborhoods in actual disorder.

MTO enables us to rigorously test what happens to individuals' criminal behavior when they move to neighborhoods characterized by what broken windows theory predicts should be of greatest relevance—order and disorder.<sup>87</sup> Of course as Table 5 shows, MTO also induces

---

<sup>85</sup> See Orr, et al, *Moving to Opportunity Interim Impacts Evaluation* at 66 (cited in note 17) (presenting data on MTO Housing and Neighborhood Safety and Conditions Outcomes).

<sup>86</sup> This finding is consistent with our analysis above demonstrating that the highest levels and largest increases in misdemeanor arrests in New York City during the 1990s were in the highest crime (and so presumably most disadvantaged) police precincts. See text accompanying notes 69–73.

<sup>87</sup> Ideally we would wish to complement the survey-based measures of social and physical disorder obtained from MTO adults with measures for systematic social observation (SSO) of the sort pioneered by the PHDCN research team. See generally Sampson, 105 Am J Soc 603 (cited in note 29) (describing research methods). Such data were not collected as part of the MTO evaluation for cost and other reasons, although fortunately PHDCN research shows that, at least for Chicago neighborhoods, measures of disorder from SSO and surveys are highly correlated. See id at 625. SSO measures of disorder are also highly correlated with neighborhood structural disadvantage. See id at 624. The fact that various measures of disorder and structural

changes in a variety of other characteristics of the communities in which program participants live, including lower crime rates, fewer low-income residents, and more residents with high levels of schooling or occupation in high-status jobs. Findings from MTO thus provide a test of the combined effects of reducing community disorder together with increasing neighborhood affluence, the sort of combined neighborhood changes that we would expect in normal circumstances: when government policies reduce neighborhood disorder, an important local amenity, we would expect gentrification to occur to some degree and so change the socioeconomic composition of the neighborhood somewhat.

#### B. Effects of MTO on Criminal Behavior

Analysis of arrest records and survey data suggests that moving to a less disadvantaged, less disorderly neighborhood does *not* on net reduce criminal behavior for MTO program participants. While some subgroups do respond to moves to less disorderly neighborhoods by reducing their involvement in criminal behavior, most notably female youth, these effects are offset by increases in antisocial behavior among other subgroups. Nothing in broken windows theory or most other models of neighborhood effects suggests that such influences on criminal behavior should be strongly contingent on people's demographic characteristics. Therefore, at the very least, broken windows is not a complete explanation for how communities influence criminal behavior, because even if the broken windows mechanism is at work for MTO participants, other behavioral processes seem to predominate for at least some subgroups. Moreover, for policy purposes what is most relevant is the impact of neighborhood disorder on the overall offending rate, and MTO provides fairly strong evidence that for at least this population there is no net reduction in crime or other antisocial behaviors.

The first row of Table 6, adapted from Ludwig and Kling (2005), summarizes the main MTO finding: when we pool youth and adults, using data for both males and females, and compare overall arrests across MTO groups, we find no statistically significant differences in arrest rates for people who live in neighborhoods with quite different levels of physical and social disorder. The intent-to-treat (ITT) estimates compare the average number of arrests for everyone assigned to the Experimental versus Control groups or Section 8 versus Control, regardless of whether the family has moved through the MTO

---

disadvantage are all highly correlated means that MTO provides a test for the causal effects of changing all of these neighborhood attributes simultaneously.

program.<sup>88</sup> The estimates for the effects of treatment-on-the-treated (TOT) are essentially equal to the ITT estimates divided by the fraction of families in the Experimental group (or Section 8 group, for the Section 8-Control estimate) that relocate through the MTO program.<sup>89</sup>

---

<sup>88</sup> These across-group differences are calculated with regression adjustment for a series of baseline survey characteristics such as household head race, age, educational attainment, and employment status, as well as indicators for prerandom assignment arrests. Because of random assignment, regression adjustment for these characteristics has little effect on the point estimates for the across-group differences but helps improve the precision of our estimates (that is, reduce the standard errors) by accounting for residual variation in the outcome measures of interest. We calculate robust standard errors that are adjusted for the clustering of adult and youth participants within the same households. The estimates also use weights to account for changes in the random assignment probabilities over time during the course of the MTO demonstration.

<sup>89</sup> See Howard Bloom, *Accounting for No-Shows in Experimental Design*, 8 Evaluation Rev 225, 225–46 (April 1984). The TOT estimate will be an unbiased estimate of the effects of treatment on the treated if random assignment is truly random, and if assignment to the treatment group has no effect on those who do not move through MTO. This second assumption may not be literally true, because the counseling services and search assistance offered to treatment families may influence later mobility patterns or other youth behaviors even among families that do not relocate through MTO. The disappointment of searching but failing to find an apartment may also affect nonmovers in the treatment groups. If the effects of treatment-group assignment are substantially smaller for those who do not move through MTO compared to those who do, our TOT estimates will approximate the effects of MTO moves on those who move through the MTO program. Mechanically, we calculate TOT estimates using two-stage least squares where we use indicators for random assignment outcomes as instruments for indicators for MTO treatment take-up.

TABLE 6  
Effects of MTO Random Assignment on Arrests of Youth and Adults

|                        |          | E-C               | E-C               | S8-C              | S8-C              |
|------------------------|----------|-------------------|-------------------|-------------------|-------------------|
|                        | Controls | ITT               | TOT               | ITT               | TOT               |
| <i>All crimes</i>      |          |                   |                   |                   |                   |
| All                    | 1.123    | 0.031<br>[0.053]  | 0.072<br>[0.120]  | 0.015<br>[0.060]  | 0.027<br>[0.104]  |
| Females                | 0.759    | -0.049<br>[0.047] | -0.111<br>[0.107] | -0.036<br>[0.051] | -0.062<br>[0.089] |
| Males                  | 1.994    | 0.225<br>[0.131]  | 0.513<br>[0.298]  | 0.136<br>[0.147]  | 0.237<br>[0.256]  |
| <i>Violent crimes</i>  |          |                   |                   |                   |                   |
| All                    | 0.322    | -0.017<br>[0.017] | -0.039<br>[0.040] | -0.001<br>[0.022] | -0.002<br>[0.038] |
| Females                | 0.759    | -0.024<br>[0.016] | -0.055<br>[0.035] | -0.029<br>[0.018] | -0.051<br>[0.031] |
| Males                  | 1.994    | 0<br>[0.042]      | 0<br>[0.096]      | 0.063<br>[0.054]  | 0.109<br>[0.094]  |
| <i>Property crimes</i> |          |                   |                   |                   |                   |
| All                    | 0.329    | 0.056<br>[0.026]  | 0.128<br>[0.060]  | 0.031<br>[0.027]  | 0.054<br>[0.048]  |
| Females                | 0.257    | 0.025<br>[0.030]  | 0.057<br>[0.068]  | 0.016<br>[0.030]  | 0.028<br>[0.052]  |
| Males                  | 0.502    | 0.131<br>[0.050]  | 0.3<br>[0.115]    | 0.066<br>[0.052]  | 0.115<br>[0.091]  |
| <i>Drug crimes</i>     |          |                   |                   |                   |                   |
| All                    | 0.26     | 0<br>[0.024]      | 0<br>[0.055]      | -0.024<br>[0.025] | -0.041<br>[0.044] |
| Females                | 0.136    | -0.032<br>[0.020] | -0.073<br>[0.045] | -0.018<br>[0.022] | -0.031<br>[0.037] |
| Males                  | 0.558    | 0.077<br>[0.063]  | 0.176<br>[0.143]  | -0.036<br>[0.065] | -0.062<br>[0.113] |
| <i>Other crimes</i>    |          |                   |                   |                   |                   |
| All                    | 0.212    | -0.008<br>[0.016] | -0.018<br>[0.036] | 0.009<br>[0.018]  | 0.016<br>[0.031]  |
| Females                | 0.136    | -0.018<br>[0.014] | -0.041<br>[0.032] | -0.005<br>[0.015] | -0.009<br>[0.026] |
| Males                  | 0.394    | 0.016<br>[0.039]  | 0.037<br>[0.090]  | 0.043<br>[0.048]  | 0.075<br>[0.083]  |

Source: Ludwig and Kling (2005). NOTES: Sample consists of 2,731 males and 6,402 females, which reflects a pooled sample of youth, ages fifteen to twenty-five years old at the end of 2001, plus MTO adults. The gender disparity in the sample arises because most MTO households are headed by a single female, and so there are far more female than male adults in the sample. The youth sample is gender balanced. Standard errors are in brackets.

The second row shows that for females the effects on arrests of assignment to the Experimental or Section 8 rather than the Control group are negative but not statistically significant, while for males the across-group differences are positive and not quite significant at the conventional cut-off level. The remaining panels of Table 6 disaggregate the results by crime type. For females, the treatment-control

group differences in arrests are negative (albeit not significant) for violent, drug and other crimes, but not for property crimes. Males assigned to the Experimental group experience more property-crime arrests than do those assigned to the Control group.

Heterogeneity in people's responses to moving to a less disorderly, less disadvantaged neighborhood arises with respect to age as well as gender. Figure 3 below from Ludwig and Kling (2005) shows average arrest rates for MTO participants in each of the three MTO groups by age at the end of 2001, where each panel shows results separately by crime type and gender. These results come from revising the intent-to-treat estimates with an interaction between the treatment indicator variables and a cubic polynomial in age, and then presenting the predicted values of arrests-by-age for each group implied by the parameter estimates. The eight panels of Figure 3 taken together suggest that on balance moving to a less disadvantaged, less disorderly neighborhood has more beneficial (or less detrimental) effects on younger compared to older MTO participants.<sup>90</sup> In national data most crime seems to be committed by adults, even though offending rates per year are higher for teens,<sup>91</sup> so the detrimental effects on adults are not as encouraging as one might like from a policy perspective.

---

<sup>90</sup> One concern with these results stems from the use of official arrest data, which capture the combined effects of the behavior of both MTO participants and local criminal justice agencies. Variation in the probability of arrest ( $P$ ) across neighborhoods will affect the likelihood that a criminal event ( $C$ ) results in arrest ( $A$ ), with  $A = P \times C$ . Above we showed that compared to adults assigned to the Control group, those in the Experimental or Section 8 group report that local police are more responsive to calls for service. If responsiveness of police to 911 calls is positively correlated with the probability of arrest, so that the probability of arrest is higher in more affluent areas, then our analysis of arrest data may understate any effects of the MTO Experimental and Section 8 treatments that reduce criminal behavior and overstate any effects that lead to an increase in criminal offending.

<sup>91</sup> For example, in 1998, 81.3 percent of all people arrested in the United States for any crime were aged eighteen or older at the time; the figures for violent and property crime equal 82.8 and 65.2 percent, respectively. See United States Department of Justice and Federal Bureau of Investigation, *Uniform Crime Reports for the United States: 1997* 232–33 (GPO 1998), online at [http://www.fbi.gov/ucr/Cius\\_97/97crime/97crime4.pdf](http://www.fbi.gov/ucr/Cius_97/97crime/97crime4.pdf) (visited Jan 19, 2006) (reporting the number and rates of arrests by region). For evidence of differential offending rates by age, see Figure 3.

2006]

*Broken Windows: New Evidence*

311

FIGURE 3  
MTO Treatment Effects on Lifetime Arrests  
by Age and Gender—Males

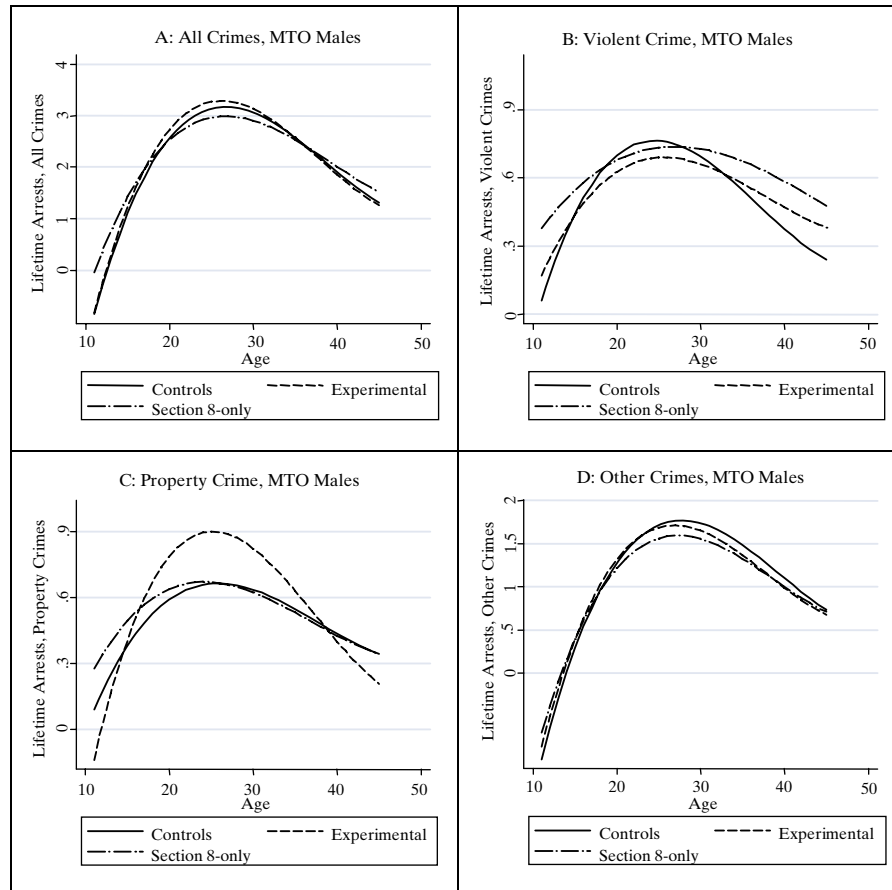
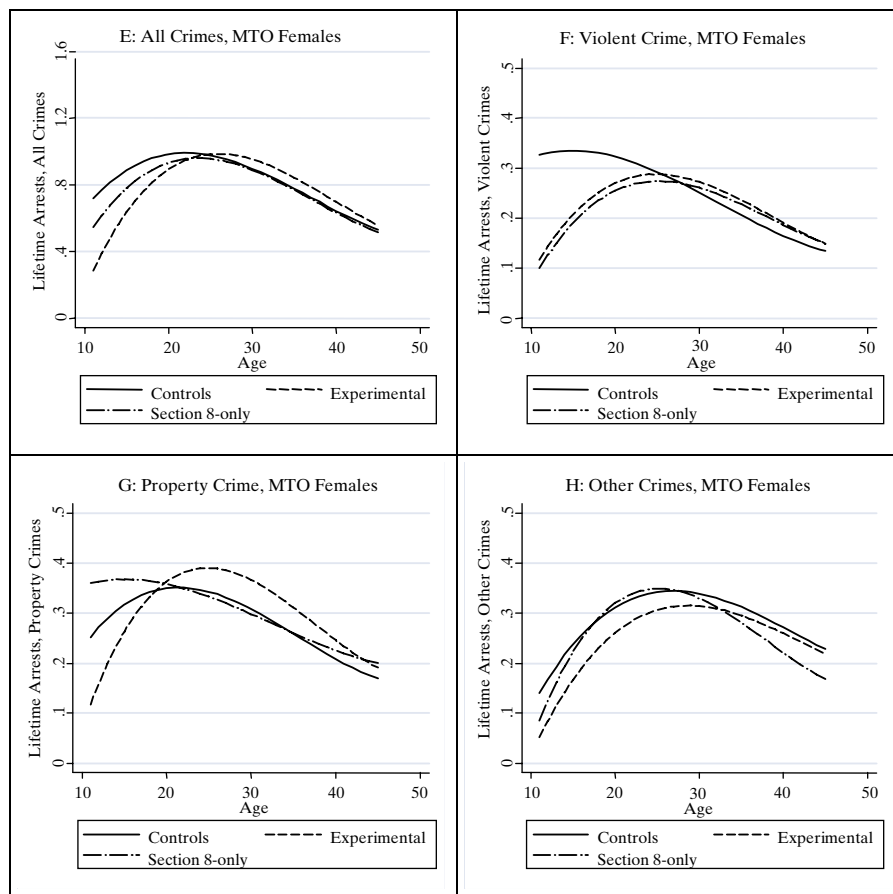


FIGURE 3 (continued)  
MTO Treatment Effects on Lifetime  
Arrests by Age and Gender—Females



Additional evidence to suggest that moving to a less disorderly, less disadvantaged community does not on net reduce criminal behavior comes from the self-reported survey data collected for program participants. Survey data on youth reveal no statistically significant differences across groups (for either males or females) in self-reported arrests or delinquency, and an increase in self-reported problem behaviors among males in the Experimental compared to the Control group.<sup>92</sup>

<sup>92</sup> See Kling, Ludwig, and Katz, 128 Q J Econ at 116–17 (cited in note 19). Comparing the control group's mean self-reported arrest rate with what is implied by the administrative records suggest that the former are susceptible to considerable underreporting. Whether this is also true for the behavior problems index, which reveals a positive Experimental-Control difference for male youth, is not clear. Of course, misreporting would have to be systematically different across groups in order to affect the estimate for across-group differences in behavior problems.

The sharp gender difference in youth responses to moving to a less disorderly, less disadvantaged neighborhood do not appear to be driven by different responses by males and females to the stress and disruption of moving per se, in part because in the first few years after random assignment experimental males experience fewer violent-crime arrests compared to controls.<sup>93</sup> The gender difference in effects—also found in recent MTO research on education, substance use, mental health, and physical health<sup>94</sup>—seems to reflect differences in how males and females respond to similar neighborhoods. Boys and girls in the same randomly assigned treatment groups move into similar types of neighborhoods, and within families, brothers and sisters respond differentially to the same mobility patterns.<sup>95</sup>

Note that one complication for interpretation of these results comes from the fact that the MTO analysis relies on administrative ar-

---

<sup>93</sup> Previous studies of the Baltimore, Boston, and New York sites that use the exogenous variation in neighborhoods induced by MTO within individual demonstration sites on balance yield evidence consistent with the view that moving to less distressed communities reduces anti-social behavior by youth, at least in the short run (one to three years from random assignment). In the Boston site, boys in the Experimental and Section 8 groups exhibit about one-third fewer problem behaviors compared to controls in the short run. See Lawrence F. Katz, Jeffrey R. Kling, and Jeffrey B. Liebman, *Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment*, 116 Q J Econ 607, 649 (2001) (offering statistical data illustrating the short-run impacts of MTO on measures of child health and behavior problems). For the Baltimore site, official arrest data suggest that teens in both treatment groups are less likely than controls to be arrested for violent crimes. These short-run impacts are large for both boys and girls, but not statistically significant when disaggregated by gender. See Jens Ludwig, Greg J. Duncan, and Paul Hirschfield, *Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment*, 116 Q J Econ 655, 671 (2001) (offering statistical data illustrating that “the results are similar when we stratify the sample by gender, although the experimental-treatment effects for girls are smaller both in absolute and proportional terms than those for boys”). Short-term survey data from the New York site reveal no statistically significant differences across groups in teen delinquency or substance use. See Tama Leventhal and Jeanne Brooks-Gunn, *Moving to Opportunity: An Experimental Study of Neighborhood Effects on Mental Health*, 93 Am J Pub Health 1576, 1579 (2003) (focusing on the short-term impact of the MTO program in New York City). Five-year data for MTO reveal that there were short-term declines in violent criminal offending for males in the Experimental versus Control groups in every site except for New York, which then dissipated over time, which suggests that changes over time in the effects of neighborhood mobility, rather than idiosyncrasies of the Boston or Baltimore sites, is the way to reconcile the short-term and medium-term results from MTO.

<sup>94</sup> See Jeffrey R. Kling and Jeffrey B. Liebman, *Experimental Analysis of Neighborhood Effects on Youth* 39 (Kennedy School of Government Working Paper RWP04-034, Aug 2004), online at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=600596](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=600596) (visited Jan 19, 2006) (“In sum, we reject the hypothesis that neighborhoods had only limited effects on these youth. We have identified some important beneficial effects of moving out of high poverty neighborhoods on the outcomes of female teenage youth, and we have established that similar benefits did not accrue to males.”).

<sup>95</sup> See Kling, Ludwig, and Katz, 120 Q J Econ at 116 (cited in note 19) (finding that although females participating in MTO experienced a large reduction in lifetime violent and property crime arrests, men participating in the program experienced a proportionally smaller reduction in such arrests).

rest data, and so in principle may confound variation across neighborhoods in criminal activity with variation in the propensity of police to make arrests or victims to report crimes to police. One (admittedly imperfect) test of this possibility comes from focusing on just the most serious crimes, which will presumably be most likely to be reported to police by victims and to result in arrest when a suspect is identified. When we replicate our results using just violent crimes excluding assaults (most of which are simple assaults for which there may be non-trivial variation across areas in victim reporting and police propensity to arrest rather than warn suspects), we obtain qualitatively similar findings to those reported above, with a full-sample Experimental-Control difference equal to about 1.5 percent of the Control group's mean arrest rate and a standard error of 12 percent of the Control mean.

The findings from MTO suggest either that declines in community disorder do not translate into reductions in individual criminal behavior or, at the very least, that any effects on criminal activity from less disorder are outweighed by the countervailing effects from increased neighborhood socioeconomic status.<sup>96</sup> These results would seem to suggest that any policy intervention that reduces disorder may not reduce people's criminal behavior if such changes are also accompanied by gentrification that alters the composition of neighborhoods in a fashion analogous to what the Experimental or Section 8 families experience in MTO.

### CONCLUSION

When Wilson and Kelling proposed the theory of broken windows in the early 1980s many academic researchers were skeptical about the ability of police activities to reduce crime. But since that time, a new body of empirical literature has, in our view convincingly, demonstrated that increased police spending does indeed reduce crime,<sup>97</sup> and that targeting police resources against the highest-crime "hot spots" can also help prevent criminal activity.<sup>98</sup> Outside of per-

---

<sup>96</sup> One candidate explanation for the possibility of such countervailing effects is the possibility that moving to a more affluent community reduces people's relative social standing within the community. Erzo Luttmer has demonstrated that having lower relative earnings compared to one's neighbors reduces happiness holding one's own income constant, which could in principle translate into increased antisocial behavior. See generally Erzo F.P. Luttmer, *Neighbors as Negatives: Relative Earnings and Well-Being* (NBER Working Paper 10667, Aug 2004), online at <http://www.nber.org/papers/W10667> (visited Jan 19, 2006).

<sup>97</sup> See Levitt, 18 J Econ Perspectives at 177 (cited in note 59).

<sup>98</sup> See Lawrence W. Sherman, *Fair and Effective Policing*, in James Q. Wilson and Joan Petersilia, eds, *Crime: Public Policies for Crime Control* 383, 389 (Institute for Contemporary Studies 2002) (describing a policing strategy centered on applying increased police presence to

haps a few remaining university sociology departments and some Berkeley coffee shops, the notion that “police matter” is (or at least should be) widely accepted. The key scientific and policy question behind the Kelling and Sousa analysis is thus whether asking police to focus on minor disorder crimes, as in broken windows policing, yields more pronounced reductions in violent crime than does having police focus on violent crimes directly. Our analysis provides no empirical evidence to support the view that shifting police towards minor disorder offenses would improve the efficiency of police spending and reduce violent crime.

We have set out in this Article, not only to assess the best available evidence for the broken windows theory—George Kelling and William Sousa’s 2001 study—but also to rethink the research design most appropriate to studying the broken windows hypothesis. We demonstrate that the pattern of crime changes across New York City precincts during the 1990s that Kelling and Sousa attribute to broken windows policing is equally consistent with mean reversion: those precincts that received the most intensive broken windows policing are the ones with the largest increases and levels of crime during the city’s crack epidemic. Consistent with findings elsewhere from city-level data,<sup>99</sup> jurisdictions with the greatest increases in crime during this period tend to experience the largest subsequent declines as well. The data from the MTO experiment reveal that moving to a less disorderly, less disadvantaged community on balance does not appear to reduce criminal behavior among the MTO program population. If disorder does affect crime, any such effects are small enough to be dominated by whatever pernicious effects on people’s criminal behavior may arise from increases in neighborhood socioeconomic status, as would be expected to occur to some degree in normal circumstances as neighborhoods with declines in disorder begin to gentrify.

When asked in January 2004 whether the broken windows theory had ever been empirically verified, James Q. Wilson reportedly told the *New York Times*: “People have not understood that this was a speculation.”<sup>100</sup> The theory was not based on empirical data, Wilson emphasized. “We made an assumption that a deteriorating quality of life caused the crime rate to go up.”<sup>101</sup> As to whether that assumption is right, Wilson stated: “I still to this day do not know if improving order

---

randomly selected hot spot intersections in Minneapolis, which decreased all reported crime by about 13 percent).

<sup>99</sup> See Raphael and Ludwig, *Prison Sentence Enhancements* at 265 (cited in note 16).

<sup>100</sup> Dan Hurley, *Scientists at Work—Fenton Earls; On Crime as Science (a Neighbor at a Time)*, *NY Times* F1 (Jan 6, 2004).

<sup>101</sup> Patricia Cohen, *Oops, Sorry: Seems that My Pie Chart is Half-Baked*, *NY Times* B7 (Apr 8, 2000).

will or will not reduce crime.”<sup>102</sup> As Wilson noted in a different interview, “God knows what the truth is.”<sup>103</sup>

Yet understanding the ability of a broken windows policy to affect disorder and crime is important for both legal and scientific purposes. The notion that broken windows policing might reduce crime is plausible because many of the behavioral mechanisms underlying this policing strategy are at least in principle consistent with existing models of social contagion.<sup>104</sup> Since the Almighty has so far resisted the temptation to publish in scholarly journals, our results help answer Wilson’s question in the interim. Our bottom line is that there appears to be no good evidence that broken windows policing reduces crime, nor evidence that changing the desired intermediate output of broken windows policing—disorder itself—is sufficient to affect changes in criminal behavior.

---

<sup>102</sup> Hurley, *Scientists at Work*, NY Times at F1 (cited in note 100).

<sup>103</sup> Cohen, *Oops, Sorry*, NY Times at B7 (cited in note 101).

<sup>104</sup> See, for example, Cook and Goss, *A Selective Review of the Social-Contagion Literature* at 3–4 (cited in note 36).

## APPENDIX: NEW YORK CITY DATA

As noted earlier, Kelling and Sousa refused to share their data with us. Fortunately we have been able to obtain the same New York City Police Department (NYPD) crime and arrest data used by Kelling and Sousa as their key dependent and explanatory variables. To measure broken windows policing, KS used precinct-level reports of total misdemeanor arrests. To measure violent crime, KS used precinct-level reports of four violent offenses (murder, rape, felonious assault, and robbery). In all cases, KS used data from 1989 to 1998. We have these data from 1989 through 2000, and so have the option of examining whether the results are sensitive to the inclusion of additional years worth of precinct-level information. We also have precinct-level reports for other types of crime, including property offenses, which enables us to explore the pattern of broken windows policing effects across crime types.

One challenge for the KS study, and for ours as well, is that data on important potential confounding factors are not readily available for New York City at the precinct level. To proxy the effect of the crack epidemic, they use borough-level reports of hospital discharges for cocaine-related episodes. To proxy the number of young males, KS uses precinct-level school enrollment data. To measure unemployment, they use borough-level gross unemployment data. Whether data measured at the level of New York's five boroughs adequately captures variation in social and policy conditions across the city's seventy-six separate precincts is an open question. Moreover, the hospital discharge data by its nature cannot distinguish between the prevalence of crack use and powdered cocaine consumption. The standard concern in the case of poorly measured explanatory variables is attenuation—bias towards zero in the coefficients for these covariates. Some evidence for this concern comes from the fact that the control variables for young males and borough cocaine consumption used by Kelling and Sousa have limited explanatory power in their model.<sup>105</sup>

We have also obtained the measures used by KS to capture variation across precincts in the drug problem and economic conditions. Specifically, we have obtained borough-level data on the number of unemployed people from the New York State Department of Labor. We have also obtained data on hospital discharges for drug-related causes from the New York State Department of Health, Bureau of Biometrics.

---

<sup>105</sup> Kelling and Sousa, *Do Police Matter?* at 9 (cited in note 6).

In addition, however, we attempt to improve upon the KS dataset in part by incorporating census tract-level measures of sociodemographic characteristics, taken from the 1990 and 2000 decennial censuses. (Data for the intercensal years are linearly interpolated.) Because census tract and police precinct boundaries do not perfectly overlap in New York City, we have geocoded both tract and precinct boundaries, and then aggregated tracts up to the precinct level by assuming that the population of tracts that cross precinct boundaries are distributed across precincts proportionately to the tract's land area.<sup>106</sup> We use these census data to calculate measures of each precinct's age distribution, poverty rate, female-headed households, fraction of adults with different levels of educational attainment, median income, and welfare receipt. To measure physical signs of disorder we also control for the fraction of housing units in the precinct that are vacant. Put differently, compared to the data used by Kelling and Sousa, our dataset includes a much richer set of sociodemographic covariates measured at the precinct level rather than some much larger unit of analysis.

Finally, we also incorporate into our dataset a measure of the number of police officers assigned to each precinct in each year by the NYPD. One important conceptual concern with the KS study is whether its key explanatory variable of interest—the misdemeanor arrest rate—captures the effects of changes in how police resources are deployed or instead simply reflects increased police presence. This counterexplanation for the KS findings is of some concern because, as Kelling and Sousa note, from 1994 to 1999 the size of the NYPD force increased by about one-third.<sup>107</sup>

Descriptive statistics from our dataset on the key dependent and explanatory variables closely match those reported by KS and by Sousa's doctoral dissertation. For example, in Sousa's Table 5-2, the mean number of misdemeanor arrests per precinct for the 1989–1998 period is 2,247, with a standard deviation of 1,968; in our dataset the mean is equal to 2,245 with a standard deviation of 1,958. Appendix Table 1 repeats this comparison for 1989, 1993, and 1998, and again shows that our figures and theirs are quite close.

<sup>106</sup> Suppose for example that census tract 1 lies entirely within precinct A, tract 2 lies entirely within precinct B, but 25 percent of the land area of tract 3 is in precinct A while 75 percent of the land area of tract 3 is within precinct B. Let  $X_i$  be some population characteristic for tract (i), such as percent poor, and let  $P_i$  represent the population of tract (i). In this case we calculate percent population poor in precinct A as  $(P_1 \times X_1 + (0.25)P_3 \times X_3) / (P_1 + (0.25)P_3)$ .

<sup>107</sup> See Kelling and Sousa, *Do Police Matter?* at 19 (cited in note 6).

APPENDIX TABLE 1

## Average Misdemeanor Arrests per Precinct, Selected Years

|                         | 1989  | 1993  | 1998  |
|-------------------------|-------|-------|-------|
| Harcourt-Ludwig dataset | 1,754 | 1,795 | 3,034 |
| Kelling-Sousa dataset   | 1,811 | 1,779 | 3,034 |

Notes: The Kelling-Sousa figures are taken from Table 2 of their Manhattan Institute report. These figures are mean number of misdemeanor arrests per precinct, calculated without weighting by precinct population.

Finally, while KS does not report the mean violent crime rate for their dataset over the entire 1989–1998 period (the sum of murder, rape, robbery, and felonious assault), their Figure 1A reports the total number of violent crimes for New York City as a whole by year. In our dataset these figures equal 144,375 in 1989, 131,310 in 1993, 97,170 in 1995, and 70,725 in 1998. Each of these numbers, and the overall trend shown in the top panel of our Appendix Figure 1, match closely the numbers represented in their Figure 1A.

## APPENDIX FIGURE 1

Violent Crime Counts in New York City by Year,  
Harcourt-Ludwig Dataset