Evidence of an inverse relationship between proactive policing and major crime

Christopher M. Sullivan[1*] & Zachary P. O'Keeffe[2]

[1]*Louisiana State University*

[2]*University of Michigan*

Governments employ police to prevent criminal acts. But it remains in dispute whether high rates of police stops, criminal summonses, and aggressive low-level arrests reduce serious crime[1–7]. Police officers target their efforts at areas where crime is anticipated and/or where they expect enforcement will be most effective. Simultaneously, citizens decide to comply with the law or commit crime based in part on police deployment and enforcement strategies. In other words, policing and crime are endogenous to unobservable strategic interaction, which frustrates causal analysis. Here, we resolve these challenges and present new evidence that proactive policing—which involves systematic and aggressive enforcement of low-level violations—is inversely related to reports of major crime. We examine a political shock that caused the New York Police Department (NYPD) to effectively halt proactive policing in late 2014 and early 2015. Analysing several years of unique data obtained from the NYPD, we find that civilian complaints of major crimes (such as burglary, felony assault, and grand larceny) decreased during and shortly after sharp reductions in proactive policing. The results challenge prevailing scholarship as well as conventional wisdom on authority and legal compliance, as they imply that aggressively enforcing minor legal statutes incites more severe

criminal acts.

In the last few decades, proactive policing has become a centrepiece of "new policing" strategies across the globe[8,9]. The logic, commonly associated with the broader theory of Order Maintenance Policing (also known as Broken Windows), is that rather than wait for citizens to report criminal conduct, law enforcement should proactively patrol communities, maintaining order through systematic and aggressive low-level policing[1,10,11]. According to proponents, increasing police stops, quality-of-life summonses, and low-level arrests deters more serious criminal activity by signalling that the area is being monitored and that deviance will not be tolerated[12,13]. As a corollary, following a phenomenon termed the "Ferguson effect," disengaging from proactive policing emboldens criminals, precipitating spikes in serious crime[14].

But while elected officials commonly justify proactive policing by pointing to the enforcement of legal statutes, the strategy's efficacy continues to be debated[5,15,16]. A serious concern is that proactive policing diverts finite resources and attention away from investigative units, including detectives working to track down serial offenders and break up criminal networks[8,17]. Proactive policing also disrupts communal life, which can drain social control of group-level violence[18]. Citizens are arrested, unauthorised markets are disrupted, and people lose their jobs, all of which create more localised stress on individuals already living on the edge[19,20]. Such strains are imposed directly through proactive policing, and thus are independent from subsequent judgements of guilt or innocence[21]. Inconsistency in aggressive low-level policing across community groups undermines police legitimacy, which erodes cooperation with law enforcement[11,20]. The cumulative effect in-

creases “legal cynicism”—individual reliance on extra-legal sanctions and informal institutions of violence as a replacement for police[22,23]. Reflecting these mechanisms, we propose that sharply reducing proactive policing in areas where it had been deployed pervasively may actually improve compliance with legal authority, reducing acts of major crime.

To assess these claims, our study analyses an aberration in NYPD strategy, in which police sharply limited foot patrols, criminal summonses, and low-level arrests in a manner unrelated to the city’s underlying crime rate. In the midst of a political fight between Mayor de Blasio, anti-police brutality protesters, and the city’s police unions, the NYPD held a work “slowdown” for approximately seven weeks in late 2014 and early 2015. Within New York City, the most proximate cause of protests against the NYPD was the strangling death of Eric Garner in Staten Island. While there was considerable fallout from the incident itself, the conflict intensified when a grand jury declined to indict the involved officers on December 4, 2014. Thousands of protesters marched across the Brooklyn Bridge, while others blocked portions of the West Side Highway as well as the Lincoln and Holland Tunnels. Then, two weeks after the non-indictment decision, two NYPD officers, Wenjian Liu and Rafael Ramos, were fatally shot by an anti-police extremist.

Because they are legally prohibited from striking, NYPD officers coordinated a work-to-rule strike. Officers were ordered to respond to calls only in pairs, leave their squad cars only if they felt compelled, and perform only the most necessary duties. The act was a symbolic show of strength to demonstrate the city’s dependence on the NYPD. Officers continued to respond to community calls for service, but refrained from proactive policing by refusing to get out of their vehicles to

issue summonses or arrest people for petit-crimes and misdemeanours.

Emblematic of the slowdown's effects (and the change from proactive to responsive policing), zero summonses were issued for quality-of-life violations on New Year's Eve 2014, while just the week before, two officers were fatally shot responding to a reported robbery. Eventually, under pressure from the media as well as growing demands for city revenue, Commissioner Bratton conceded to the "self-initiated" slowdown in proactive policing, before publicly ordering his officers to return to work by January 16.

The change in tactics appears particularly stark when compared to the aggressive strategy of proactive policing the NYPD pursued during the preceding decades. Correspondence between the introduction of proactive policing in New York and the city's historic drop in major crime has been heralded as prima facie evidence of the strategy's effectiveness[13]. As a result, cities across the globe adopted the NYPD's protocols and practices, which suggests that not only are proactive policing strategies presumed to deter major crime in New York City, but these policies are widely thought to work in other contexts as well[5,24].

If, as would seem to be the case, the slowdown was unrelated to the city's underlying crime rate, this makes for a unique natural experiment to identify the causal effects of changing police practices. While Garner was being arrested for a misdemeanour offence, and the killings of Liu and Ramos were both homicides, these three crimes neither reflect nor predict citywide (nor precinct-wide) crime. And while anti-police brutality protests and the ensuing political conflict were tied to policing practices across the country, it is difficult to argue that the protests were caused by NYC's

crime rate.

To assess the slowdown's effects, we filed a series of Freedom of Information requests soliciting a comprehensive set of NYPD CompStat reports from 2013–2016 (see Supplementary Fig. 1 for an example). CompStat (short for computer statistics) was introduced in New York as part of a series of reforms to target proactive policing at "hot spots" in which crime was most concentrated[5,24]. The reports document weekly activity in each NYPD precinct. Based upon findings from earlier research, we are confident that CompStat data (1) represent the best available source of disaggregated information on police behaviour and crime, and (2) correlate strongly with the underlying reality (see further discussion in the Methods section)[4,24–26]. Perhaps the best evidence of their validity comes from the fact that the NYPD uses CompStat reports to allocate police resources and develop strategy in real time[27].

Examining citywide time-series, we find new evidence of the timing of the NYPD slowdown, as well as preliminary indications of its effects (Fig. 1). Several policing measures are considered. *Criminal summonses* includes charges issued for summary Penal Law Violations (i.e., quality-of-life violations, including, most commonly, public consumption of alcohol and disorderly conduct, but not ticketable parking fines or moving violations). *Stop, question, and frisks (SQFs)* are temporary street detentions and searches of individuals for contraband. Use of SQFs dropped precipitously to a new baseline in anticipation of the judgement in *Floyd vs. City of New York*, which ordered a series of reforms to prevent unconstitutional racial profiling. *Non-major crime arrests* includes arrests for all crimes and misdemeanours, excluding the NYPD's "seven major

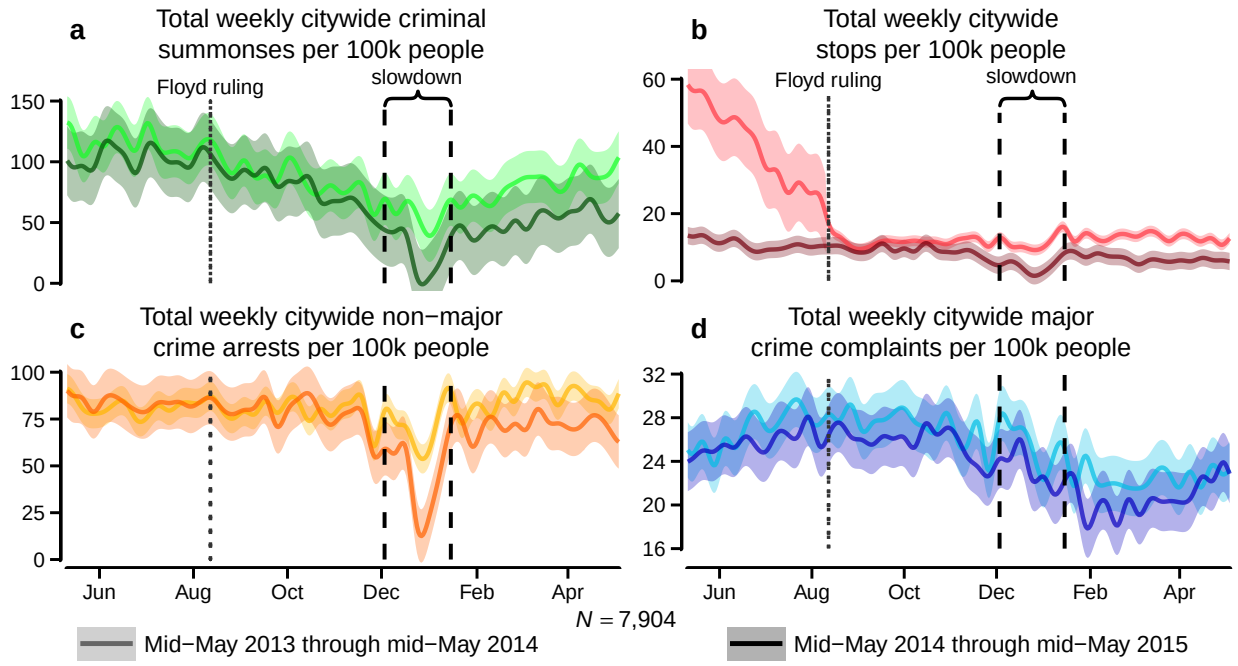


Figure 1. Temporal variation in policing and crime complaints in NYC. Titles refer to y-axes; x-axis is time; unit is one week. Lines are natural cubic splines fit through weekly data points. Series run from week 20 of 2013 (2014) to week 19 of 2014 (2015). Criminal summonses are misdemeanour and summary offences. Major crimes are murder, rape, robbery, felony assault, burglary, grand larceny, and grand theft auto; non-major crime arrests are for all other crimes. Within the first series for stop-question-and-frisks, separate standard deviations are calculated for the weeks before and after the August 12, 2013 *Floyd v. City of New York* ruling.

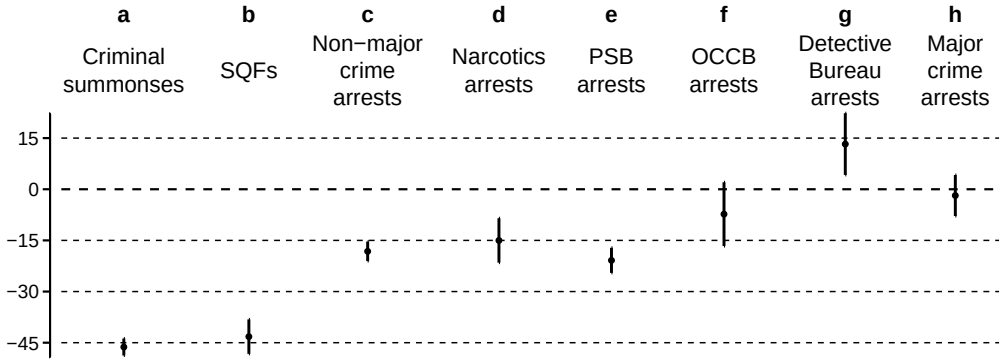
crimes"—murder, rape, robbery, felony assault, burglary, grand larceny, and grand theft auto. It includes arrests made by members of the precinct as well as officers from the Transit and Housing Departments and two specialised bureaus: the Organised Crime Control Bureau (OCCB) and the Detective Bureau. According to annual NYPD statistics, misdemeanour arrests represented 92% of all non-major crime arrests in 2014[12].

Our indicator of legal compliance, *Major crime complaints*, measures civilian reports of any of the "seven major crimes" indexed by the NYPD. We focus on major crime complaints for several reasons. First, the premise behind proactive policing is that increasing police stops, criminal summonses, and low-level arrests will prevent these types of major crimes. As expressed by two of proactive policing's chief architects, "A neighbourhood where minor offenses go unchallenged soon becomes a breeding ground for more serious criminal activity and, ultimately, for violence"[13]. Second, the NYPD pays particular attention to these offences and tracks them consistently across time and space[24]. Indicative of the measures' validity, the NYPD employs the same index of major crime complaints when assessing tactical effectiveness[27]. Third, focusing on major crime complaints is relatively standard within the literature, largely because these statistics are the most reliable across time and space[5]. Research auditing the NYPD's major crime complaints data validates the statistics: patterns found in independent sources of crime data, including victims' surveys, coroners' reports, and insurance losses, appear identical to major crime complaints[24].

Our analyses identify the effects of the 2014–2015 NYPD slowdown using a cross-sectional weekly time-series of proactive policing and major crime complaints in 76 NYPD precincts. Our

identification strategy uses difference-in-differences (DiD) to compare police and criminal behaviour before, during, and after the slowdown to similar patterns observed during the same period the year before. For our primary analyses, we examine the period from mid-January 2013 through mid-January 2015 ($N = 7{,}904$). In our DiD design, the *Treatment series* includes precinct–weeks from mid-January 2014 to mid-January 2015. The *Control series* is the same, but for 2013 to 2014. Drawing on the evidence above, our study defines the *Treatment window* (i.e., the slowdown) as occurring from December 1 through January 19. Our base specification uses negative binomial regression (see also Supplementary Tables 1–3, Supplementary Figs. 2–4). We report results from replications using Poisson, ordinary least squares, and interrupted time series specifications (Fig. 3, Supplementary Tables 4–9, Supplementary Figs. 5–7). In the analyses, we control for a variety of demographic characteristics, measures of police capacity and strategy, elements of concentrated disadvantage, season and weather indicators, time trends, and spatial-temporal lags of our dependent variables. Details on our measurement and identification strategy are contained in the Methods section.

To describe the effects, we rely primarily on the *ATT weekly % change*, which is a derivative of the average treatment effect on the treated (ATT). We calculate the *ATT weekly % change* as the average (per precinct–week) percentage change in the outcome caused by the slowdown during the slowdown weeks, compared to the predicted outcome had the slowdown not occurred[28] (see also Supplementary Table 1 and Supplementary Fig. 1). It is similar to the average marginal effect (AME), which is the average per precinct–week predicted change across the time-series. To provide substantive context for interpreting these results, we also report the NYPD's average rate



	a	b	c	d	e	f	g	h
	Criminal summonses	SQFs	Non−major crime arrests	Narcotics arrests	PSB arrests	OCCB arrests	Detective Bureau arrests	Major crime arrests
Control series treatment window mean	64.541	12.883	80.115	10.045	53.767	8.352	9.039	10.299
Treatment series treatment window mean	30.226	5.103	52.521	6.449	34.288	5.472	9.118	8.859
ATT weekly % change	−46.266	−43.226	−18.201	−15.014	−20.849	−7.303	13.252	−1.838
ATT	−24.169	−3.721	−11.216	−1.072	−8.646	−0.418	1.057	−0.163
ATT standard error	0.666	0.222	0.902	0.237	0.778	0.273	0.368	0.272
Average marginal effect	−47.082	−22.140	−16.339	−1.760	−12.185	−0.785	1.384	−0.199
AME standard error	2.164	3.811	1.470	0.441	1.280	0.527	0.423	0.337
AME *p*	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.136	0.001	0.555
N	7,904	7,904	7,904	7,904	7,904	7,904	7,904	7,904

Figure 2. Effects of slowdown on police behaviour. Column headings indicate outcome variables. Control (treatment) series is week 4 of 2013 (2014) to week 3 of 2014 (2015). Treatment window is week 49 of 2013 (2014) to week 3 of 2014 (2015). Series window means = mean observations of the DV during the treatment window of the series. ATT = mean predicted precinct–weekly change in the DV during the slowdown. Averaged predicted % change is 100 × ATT/mean of the predicted counterfactual values. AME = mean of *Series* × *window* partial effects. Delta method standard errors clustered by precinct. Lines are 95% CIs. AME p-values use two-tailed F-tests.

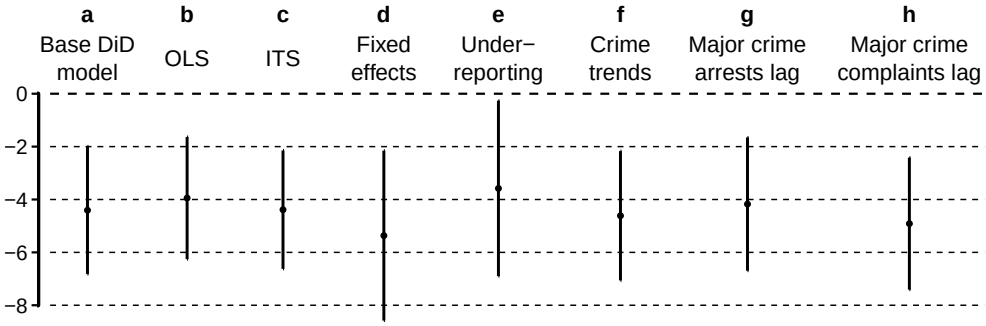
of police behaviours observed for each precinct–week during the treatment window, as well as the same time period a year before.

We first estimate the slowdown's effects on police behaviour (Fig. 2, Supplementary Table 1, Supplementary Fig. 2). Following the procedures of a recent NYPD assessment of OMP, our approach, "acknowledges that disorder reduction may not always require issuing summonses or making misdemeanor arrests, and may include other police activities like... situational crime prevention or problem-oriented policing strategies," while limiting analyses of proactive policing to, "focus exclusively on quality-of-life enforcement as a crime reduction tactic rather than these other forms of disorder reduction"[12]. We find that, compared to other policing tactics, *Criminal summonses* and *SQFs* decreased most precipitously during the slowdown, supporting earlier claims that the slowdown particularly affected low-level policing. *Non-major crime arrests* also declined significantly and by substantively meaningful amounts. Because CompStat data do not allow the study to exclude felony offences and violent crimes other than the seven major crimes from non-major crime arrests, we consider additional evidence locating the effects of the slowdown on proactive policing. *Narcotics arrests*, which includes all charges relating to illegal drugs, dropped significantly during the slowdown. Alongside these measures, we consider arrests made by each precinct's Patrol Services Bureau (PSB), OCCB, and Detective Bureau, conditioning our estimates on precinct-wide trends to locate any unique changes affecting the different bureaus. While the PSB engaged in significantly fewer arrests during the slowdown, the OCCB did not experience any significant bureau-level decline in arrests. Replications examining arrests by officers in the Housing Bureau and Transit Bureau also returned nonsignificant results. In sharp contrast to

this trend, evidence shows that arrests by the Detective Bureau increased significantly during the slowdown. This result is highly relevant to one of our theoretical mechanisms, since the Detective Bureau is charged with intensive investigations, rather than proactive policing. Further confirming that the slowdown's effects were localised to proactive policing, we find no evidence that *Major crime arrests* were significantly affected by the slowdown when we condition our estimates on *Major crime complaints*.

Having established that the slowdown significantly reduced proactive policing, we next estimate the slowdown's effect on *Major crime complaints* (Fig. 3, Supplementary Table 2, Supplementary Fig. 3). Contradicting arguments that systematically decreasing proactive policing should correspond to increased crime (i.e., the Ferguson effect), our results reveal that civilian complaints of major crimes declined by approximately 2–6% during the slowdown. Following these estimates, the decline in major crime caused by the cessation of proactive policing corresponds roughly to the relative decline in crime that earlier research attributed to the effects of mass incarceration[29]. Replicating the analysis using alternative model specifications, including ordinary least squares and interrupted time-series specifications, produced substantively identical results (Fig. 3, Supplementary Tables 5 and 8, Supplementary Fig. 6).

One might worry that underreporting during the slowdown may be confounding our estimates of declining major crime complaints. Concerns of underreporting do not nullify the identified decline in major crime complaints, but they do complicate a strict causal interpretation of our results. Perhaps officers were less likely to learn of crimes because they were staying in their



	a	b	c	d	e	f	g	h
Control series treatment window mean	28.246	28.246		28.246	28.246	28.246	28.246	28.246
Treatment series treatment window mean	25.555	25.555		25.555	25.555	25.555	25.555	25.555
ATT weekly % change	−4.407	−3.945	−4.386	−5.367	−3.580	−4.615	−4.174	−4.914
ATT	−1.178	−1.050	−1.171	−1.448	−0.948	−1.237	−1.113	−1.320
ATT standard error	0.329	0.314	0.305	0.442	0.448	0.334	0.343	0.342
Average marginal effect	−1.218	−1.050	−1.203	−1.480	−0.986	−1.276	−1.154	−1.359
AME standard error	0.322	0.314	0.318	0.453	0.457	0.330	0.336	0.337
AME *p*	< 0.001	< 0.001	< 0.001	0.001	0.031	< 0.001	< 0.001	< 0.001
N	7,904	7,904	11,856	7,904	7,904	7,904	7,904	7,904

Figure 3. Effects of slowdown on crime complaints. Outcome variable = # of weekly precinct major crime complaints. All models use NB2 regression, except model **b**, which uses OLS. All models use DiD except model **c**, which uses ITS. Control (treatment) series for DiD models is week 4 of 2013 (2014) to week 3 of 2014 (2015). Treatment window for DiD models is week 49 of 2013 (2014) to week 3 of 2014 (2015). The ITS intervention week is 49 of 2014; the post-intervention period begins week 4 of 2015. **c–e** include precinct fixed-effects. **c** includes month dummies. **e** includes weekly precinct total of complains of misdemeanours and violations. **f** controls for the % change in weekly precinct major crime complaints between 2012 and 2011, and 2013 and 2012. **g** includes a one-week lag of major crime arrests. **h** includes a one-week lag of major crime complaints.

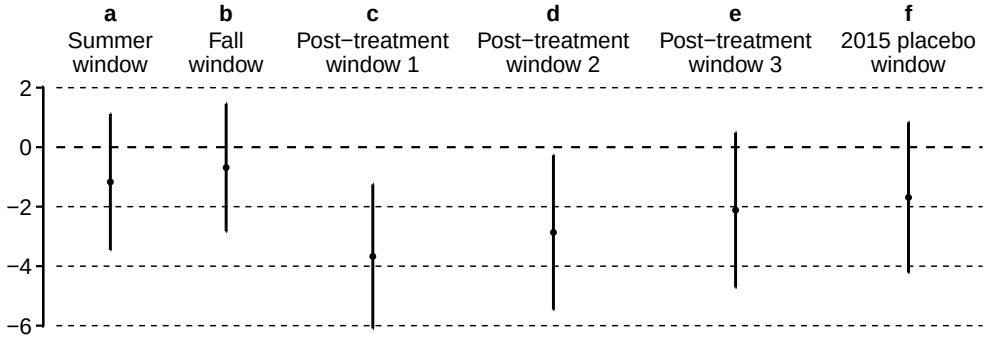
squad cars, rather than patrolling the streets and speaking with victims about their experiences. Or trust in police may have fallen due to tensions between protesters and police. Recent findings show that high-profile cases of police violence suppress police-related 911 calls[22]. Anecdotal evidence also suggests that trust in police was down during this period, though trust had been declining since the summer. And there is evidence that calls for NYPD service are significantly lower in areas with the highest rates of police stops and police use of force, further complicating questions about underreporting[30]. Individuals may be less likely to report crimes when they think they are going to be stopped, questioned, and potentially arrested in the process[20]. To the extent that the slowdown reduced the willingness of police officers to stop bystanders and/or arrest people for petty offences, the study could underestimate the decline in crime induced by the slowdown.

In our analyses, we examine how crime underreporting may bias the results. We employ precinct fixed-effects to address time-invariant sources of underreporting, such as communities' varying histories of police distrust. We then model time-variant sources of underreporting biases, such as those caused by the killing of Eric Garner and/or the heightened conflict between protesters and police. Model (e) in Fig. 3 controls for the number of community complaints reported in each precinct–week for misdemeanours and violations. Assuming time-variant sources of underreporting are correlated across crime types, this model is robust to slowdown-induced underreporting bias. While we cannot entirely rule out the effects of underreporting, our results show that crime complaints decreased, rather than increased, during a slowdown in proactive policing, contrary to deterrence theory. Additional tests show the results are robust to specifications including controls for long-term trends in crime (Fig. 3f), lagged *Major crime arrests* (Fig. 3g), and lagged *Major*

crime complaints (Fig. 3h). We report results from more robustness checks in Supplementary Fig. 8.

We also examined how the slowdown affected the different crimes constituting *Major crime complaints* (Supplementary Fig. 9). While no category showed statistically significant increases during the slowdown, four complaint categories—murder, rape, robbery, and grand theft auto—return statistically insignificant results, which we attribute to the relatively small number and high variance of such crimes. Robbery, the most common of the four insignificant categories, falls closest to statistical significance, but estimates appear highly sensitive to model specification. In light of earlier evidence, it is surprising that we find no robust increase in robbery complaints. One highly influential study finds the strongest evidence supporting OMP exists in a "significant albeit modest association of disorder and officially measured robbery"[16]. And a recent analysis examining similar quasi-experimental conditions shows small increases in larcenies and robberies during the 1996–1997 NYPD labour negotiations strike[31]. Our results belie these findings, as they show no statistically significant increase in complaints of any of the seven major crimes. Instead, evidence shows the decline in major crime complaints identified during the slowdown was most affected by statistically significant reductions in three high-volume categories: complaints of felony assault, burglary, and grand larceny. Each week during the 2014–2015 slowdown, we estimate that 28 fewer felony assaults, 41 fewer burglaries, and 37 fewer acts of grand larceny were reported.

Our analyses identify the timing and duration of the decline in major crime complaints by



	a	b	c	d	e	f
Control series treatment window mean	31.258	30.474	24.553	24.494	26.785	28.246
Treatment series treatment window mean	29.318	28.791	21.743	22.594	25.446	26.058
ATT weekly % change	−1.170	−0.685	−3.672	−2.865	−2.115	−1.690
ATT	−0.348	−0.199	−0.827	−0.665	−0.550	−0.448
ATT standard error	0.348	0.319	0.278	0.307	0.345	0.342
Average marginal effect	−0.323	−0.189	−1.011	−0.802	−0.586	−0.458
AME standard error	0.323	0.303	0.377	0.382	0.348	0.363
AME *p*	0.316	0.532	0.007	0.036	0.092	0.208
N	7,220	7,220	6,840	5,928	6,080	7,904

Figure 4. Alternate slowdown specifications for major crime complaints. Outcome variable = # of weekly precinct major crime complaints. Models are use DiD and NB2 regression. Control (treatment) series by model are: **a–b** week 1 of 2013 (2014) to week 48 of 2013 (2014); **c** weeks 10–48 of 2013 (2014) and weeks 4–9 of 2014 (2015); **d** weeks 17–48 of 2013 (2014) and weeks 10–16 of 2014 (2015); **e** weeks 25–48 of 2013 (2014) and weeks 4–24 of 2014 (2015); **f** week 3 of 2013 (2015) to week 3 of 2014 (2016). Treatment windows by model are: **a** week 29 of 2013 (2014) to week 35 of 2013 (2014); **b** week 37 of 2013 (2014) to week 43 of 2013 (2014); **c** week 4 of 2014 (2015) to week 9 of 2014 (2015); **d** week 10 of 2014 (2015) to week 16 of 2014 (2015); **e** week 17 of 2014 (2015) to week 25 of 2014 (2015); **f** week 49 of 2013 (2015) to week 3 of 2014 (2016).

replicating the analysis using different operational definitions for the *Treatment series*, *Control series*, and *Treatment window* (Fig. 4, Supplementary Table 3, Supplementary Fig. 4). The findings refute arguments that the decline in major crime complaints could have been affected by other factors emerging prior to the slowdown. No significant change in major crime complaints occurred following the death of Eric Garner (in July 2014) or in the months leading up to the slowdown. Additional tests confirm the timing of declines in major crime complaints aligns with the slowdown (Supplementary Fig. 11).

We also test whether the slowdown's effect on crime complaints extended past its publicly announced end. Results from post-treatment analyses show that statistically significant reductions in major crime complaints occurred seven and even fourteen weeks after sharp declines in proactive policing. While the study cannot address a principal concern of the law enforcement community—that reductions in proactive policing could increase criminality years later—it demonstrates substantial short-term reductions in crime that should prompt reflection on the mechanisms linking proactive policing to deterrence. Wilson and Kelling, for example, suggest the benefits of proactive policing could be observed "in a few years or even a few months"[10]. Other studies point to the fact that crime rates remain plastic and highly volatile as evidence that persistent proactive policing caused New York City's crime decline, rather than structural factors like demography[24]. As expressed by the NYPD, "Current crime levels don't stay down by themselves. . . crime is actively managed in New York City everyday."[13] Further research will need to examine additional long-term effects. Within the short term, we estimate that the slowdown resulted in nearly 1,420 fewer major crimes complaints. This estimate extrapolates from the ATTs for the seven weeks of the

slowdown, plus the 14 weeks of the two significant post-treatment windows. Tests of subsequent windows in the spring of 2015 return insignificant results, indicating that, as NYPD tactics returned to normal, the city's crime rate eventually reverted to its pre-treatment baseline. The insignificant results of a placebo test using a window spanning the seven weeks after the killing of Freddie Gray in April 2015 (Fig. 4e) supports our conclusion that the results were not solely induced by the effects of police-related violence on underreporting[22]. Finally, results from a placebo test estimating the counterfactual scenario in which the slowdown took place during the subsequent year (2015–2016) (Fig. 4f) prove insignificant, confirming we have not misidentified our causal effect.

Findings from our study warrant a reconsideration of the assumptions guiding scholarship and practice related to enforcement and legal compliance. In their efforts to increase civilian compliance, certain policing tactics may inadvertently contribute to serious criminal activity. The implications for understanding policing in a democratic society should not be understated. It is well established that proactive policing is deployed disproportionately across communities, and that areas with high concentrations of poverty and persons of colour are more likely to be targeted[8]. Our results imply not only that these tactics fail at their stated objective of reducing major legal violations, but also that the initial deployment of proactive policing can inspire additional crimes that later provide justification for further increasing police stops, summonses, etc. The vicious feedback between proactive policing and major crime can exacerbate political and economic inequality across communities[32]. Absent reliable evidence of proactive policing's effectiveness, it is time to consider how proactive policing reform might reduce crime and increase well-being in the most heavily policed communities.

Methods

Data For benchmarking purposes, each CompStat reports data both for the current year as well as for the same 7-day range in the previous year. Thus, the 2015 CompStats include 2014 data, and the 2014 CompStats include 2013 data, with weeks matched by their calendar start and end days. The 2014 data contained in the 2014 and 2015 CompStats do not perfectly align, however. Because of when the 52nd week of the previous year finished, week 1 of 2014 begins on December 30, 2013, and week 1 of 2015 begins on December 29, 2014. As a result, the weekly 2014 totals from the 2015 CompStats are off by one day compared to the weekly 2014 totals from the 2014 CompStats. While the choice of how to cut the data does not meaningfully impact our results, we constructed our data in the following way. Necessarily, 2013 data are taken from the 2014 CompStats, and 2015 data from the 2015 CompStats. But because there are two observations for each week in 2014 (one from the 2014 reports, one from the 2015 reports), we are forced to adopt a rule for which values to use. Because our treatment and control series span multiple years by approximately three weeks in the beginning of January, we reasoned the best criterion to use to subset the data is to maintain internal consistency within each series. To accomplish this, we used only the 2014 CompStats for all weeks measured as part of the control series, and only the 2015 CompStats for all weeks of the treatment series.

For several reasons, we are confident the results are not affected by the one-day difference in the series. First, since the *Series* $\times$ *Treatment window* interaction effect is estimated by averaging over a seven-week period, days contained within the five middle weeks overlap completely, leaving

only two weeks that are off by a day. Second, we replicated the analyses by averaging the 2014 weeks from the 2014 and 2015 reports. This approach yields comparable results, but because only a single data point is available for each week in 2013 and 2015, we prefer to maintain the data's internal consistency, rather than introduce another manipulation.

Since our data come from the NYPD, it is worth considering potential sources of bias in police reporting. Concerns have been raised about police data being influenced by the officers tasked with collecting statistics, as well as their superiors[25]. Still, we feel confident in the CompStat data for several reasons. First, police data are often strongly preferable to alternative sources. Because police records contain a more extensive listing of activity, they are often used to identify the form and extent of bias in other data sources. Second, to minimise the biases associated with human error, the NYPD requires officers to apply a "strict interpretation bias". When reporting a crime complaint, an officer must enter the incident based upon the most serious crime described by the claimant, regardless of whether the officer believes the perpetrator can be tried or arrested for that offence. This procedure was put into place under the theory that strict interpretation bias would increase the willingness of individuals to come forward with crime complaints. As a result, the majority of errors in the categorisation of a crime should lead to upgrading, rather than downgrading, the criminal classification[25].

Any remaining bias from manipulation by police officers would predispose the study towards identifying an escalation in major crime complaints. Prior to the slowdown, precinct commanders' interest lay in demonstrating continuing declines in crime. Professional incentives reversed during

and after the slowdown insofar as commanders wished to demonstrate the necessity of the police force and the effectiveness of their policing strategies.

With regards to police protocol, there were three important changes in NYPD procedure during our time-series worthy of mention. First, on October 31, 2013, an appellate court ruled on *Floyd vs City of New York*, ordering NYC to eliminate racial profiling in the NYPD's stop-and-frisk encounters. We display the corresponding sharp decline in these encounters in Fig. 1. Our analyses control for the effects of the *Floyd* decision by including precinct–week counts of SQFs in models of other police and criminal behaviours, and a dummy for the SQF model itself. Second, in July 2014, the Brooklyn District Attorney Ken Thompson declared that his office would no longer prosecute marijuana possession under certain conditions. Third, on November 19, 2014, NYC formally decriminalised marijuana possession, making it a summons rather than an arrestable offence. While these three procedural changes surely impacted policing practices, their causes are unrelated to the slowdown, and thus we should not expect them to impact our causal estimation. Indeed, our analyses in Fig. 4 and Supplementary Fig. 11 show that the timing of changing patterns of compliance corresponds to the period of the slowdown, rather than these earlier procedural changes.

With regards to data availability, we encountered missingness in only two situations. The first was with our measures of policing strength and strategy, which are from 2007 and 2013, and thus predate the formation of the 121st precinct, which became fully operational in July 2013. To address this, we imputed values for this precinct using data from the 120th and 122nd precincts,

which were split to form the 121st. We weight these variables' values for all three precincts proportionately based upon geographic coverage.

The second site of missing data results from the fact that the NYPD has thus far failed to turn over CompStat reports from two weeks in January 2016. In spite of the NYPD's recalcitrance, we have no reason to suspect that the missing weeks (weeks 2–3 of 2016) impact the results. To empirically demonstrate that our results are not affected by missing data, we take a conservative approach when imputing data for the missing values. In the final column of Fig. 4 we fill in the missing 2016 data with the two weeks from the actual slowdown (in January 2015). We believe this is a better modelling strategy than multiple imputation, which can introduce bias when applied to non-linear models[33]. Imputing the missing data using the actual slowdown values also presents a harder test for demonstrating insignificance in the placebo treatment as compared to last observation carried forward. In the first four weeks of the slowdown, rates of major crime complaints were nearly 20% lower as compared to the same period the following year. Replicating the 2015 placebo tests without the imputed weeks produces comparable nonsignificant results (Supplementary Fig. 8e).

Modelling Strategy Our econometric model is represented in equation 1. The DiD model estimates changes in police behaviour or civilian crime complaints ($\mathbf{Y}$) as a function of the slowdown (δ) and a variety of covariates ($\mathbf{X}$). The technique controls for any systematic, unobservable differences between the control and treatment series (γ), as well as trends common to both series during the treatment window (λ)[28,34]. $\mathbf{S}$ and $\mathbf{T}$ are dichotomous indicators of the *Series* and *Treatment*

window, respectively.

$$\mathrm{E}[\mathbf{Y}|\mathbf{S},\mathbf{T},\mathbf{X}] = \Phi(\alpha + \gamma\mathbf{S} + \lambda\mathbf{T} + \delta(\mathbf{S}\mathbf{T}^T) + \boldsymbol{\beta}\mathbf{X}) \tag{1}$$

A critical requirement of the difference-in-differences modelling strategy is the "parallel trends" assumption. To reliably estimate differences during the treatment window, the data must follow the same pattern outside the window. Fig. 1 confirms that the control series indeed provides a reliable baseline from which to measure any changes induced by the slowdown.

We estimate the models using a negative binomial specification (Φ) because all outcome variables are overdispersed count data, as revealed by two types of overdispersion tests. For the base model using major crime complaints as the outcome variable, a likelihood ratio test comparing a Poisson vs negative binomial specification produces a χ^2 value of = 1512 ($p < .001$). Results are comparable for all other models. Ordinary least squares is even less appropriate than Poisson, as the dependent variables are neither normally distributed nor interval. Furthermore, because observations within precincts are not independent, and hence their errors are correlated, we calculate robust standard errors clustered by precinct.

While the slowdown in policing is arguably independent from precinct-level covariates, we include a number of controls in case these variables influence the precincts' responsiveness. Including controls like population and other demographic characteristics helps normalise the variance in the dependent variables across precincts. Because it lacks a residential population, all analyses exclude the Central Park Precinct. We use the most recent demographic data, which is

taken from the 5-year American Community Survey (ACS). Using the ACS, we identified each precinct's *Population* in 2014, as well as the crime-prone age group *% Aged 15–24*.

We also include a number of key indicators of concentrated disadvantage. Using data from the ACS, we generated precinct-level measures of *Average family income*, *% Persons of colour* (PoC), *% Unemployed*, and several household-level measures, including *% Households on public assistance*, *% Female-headed households with children*, *% Occupied housing units rented*, and *% Households vacant*. Because these factors loaded poorly on a single dimension as well as on two dimensions, all analyses with covariates incorporate these variables individually (see Supplementary Tables 1–3). In Supplementary Fig. 8c, we report a replication using our measure of concentrated disadvantage, which is defined as the mean of the standardised (i.e., centred at 0 and scaled such that standard deviations are equal to 1) values of *% Persons of colour*, *% Unemployed*, *% Household on public assistance*, and *% Female-headed households with children*. The model accordingly does not include these constituent variables individually to avoid collinearity.

Our models also control for precinct-level variation in policing capacity and behaviour. In addition to the total precinct–week *SQFs* mentioned earlier, using data from Rengifo and Fowler[35], we construct per capita precinct-level variables of the number of officers assigned to a precinct in 2007 (*Officers per 100k people*) and complaints registered against the precinct with the community complaint review board in 2013 (*CCRB per 100k people*). The CCRB data also provide an indicator of the distribution of complaints across racial groups, which we measure as *% CCRB persons of colour/% persons of colour*.

We further include three weather related controls, each of which are weekly averages of daily measures for New York City from the National Weather Service: *Mean temperature*, *Total rain accumulation*, and *Total snow/sleet accumulation*. To account for temporal autocorrelation and geographic spill-over effects, all models include a one-week spatial lag of the dependent variable. To construct this, we identified adjacent, contiguous precincts for each precinct, and calculated the mean of the previous week's values. Because NYC is composed of multiple islands connected by bridges and tunnels, we deem this more appropriate than using an inverse distance weighted measure. The choice does not meaningfully alter the results.

Lastly, while Fig. 1 lends support to the parallel trends assumption necessary to DiD, we include two additional controls for temporal variance in policing and criminal behaviour. First, to adjust for the ongoing downward trend in crime, we include a *Time counter*, which counts the number of weeks since the first week of the time-series, starting at 1. Second, alongside the three weather-related variables, we include dummy variables for *Summer*, *Fall*, and *Winter* to help control for seasonal effects.

We estimate our causal effects as shown in equation 2:

$$\begin{aligned} \tau(\mathbf{S}=\mathbf{1},\mathbf{T}=\mathbf{1},\mathbf{X}) &= \mathrm{E}[\mathbf{Y}^1|\mathbf{S}=\mathbf{1},\mathbf{T}=\mathbf{1},\mathbf{X}] - \mathrm{E}[\mathbf{Y}^0|\mathbf{T}=\mathbf{1},\mathbf{S}=\mathbf{1},\mathbf{X}] \\ &= \Phi(\alpha+\gamma+\lambda+\delta+\boldsymbol{\beta}\mathbf{X}) - \Phi(\alpha+\gamma+\delta+\boldsymbol{\beta}\mathbf{X}) \end{aligned} \tag{2}$$

Our procedure for calculating causal effects is as follows: We generate average marginal effects ($\bar{\tau}$, or AME) by averaging the precinct–week differences between the predicted value with

the *Series* $\times$ *Treatment window* set to 1 vs. set to 0 (i.e., the difference between the value predicted for a given precinct–week observation had it occurred during the slowdown vs. had it not, all else equal)[28]. Average treatment effects on the treated (ATT) are calculated in a similar vein, but use only predicted values for the weeks in the treatment series treatment window (e.g., the actual slowdown weeks in our base model), as opposed to the whole time series. The ATT is converted to predicted weekly percentage change by dividing it by the mean predicted counterfactual and multiplying be 100. Figs. 2–4 graphically present the ATT percentage and corresponding 95% confidence intervals with delta method standard errors clustered by precinct, as well as report the raw ATTs and AMEs, as well as their standard errors. Statistical significance for average marginal effects are determined using two-tailed F tests. More detailed results for the models in Figs. 2–4 can be found in Supplementary Tables 1–3.

Fully understanding the estimated effects in the negative binomial DiD models requires considering the non-linear parameterisation of the model[28]. While γ and λ control for series and treatment effects, these coefficients cannot be presumed to equal zero in the cross-difference of equation 2. In Supplementary Figs. 2–4, we estimate the differential effects of the slowdown taking into account the non-linearity of the model by presenting the secondary, or product-term induced, interaction effects and clustered delta standard errors for each precinct–week observation. While we follow Bowen in believing the secondary interaction effect to be the estimate of interest, the results using the total interaction remain the same, because the structural, or model-inherent, effect is close to 0 in all cases[36].

Finally, while we believe DiD is the best modelling approach given our data and the nature of criminality and policing, we also ran interrupted time-series (ITS) models using the entire time-series. In the ITS analyses, we replicate the modelling approach adopted by Chandrasekher, with the addition of our precinct-level control variables[31]. Results from the base specification comparing ITS to the DiD estimates are presented in Fig. 3. Results from a full replication of all models using ITS instead are displayed in Supplementary Figs. 5–7. We contend, however, that DiD is more appropriate primarily for three reasons. First, with ITS, the modeller must specify the functional form of the proposed trends before, during, and after the "treatment". The most common assumption is that trends are linear. In Fig. 3 we do not include such additional trend shifts for simplicity, but doing so does not alter the results. In Supplementary Fig. 10 we show that the predicted values from a specification in which we include slowdown and post-slowdown trend shifts results in counterfactual predictions during the slowdown that closely mirror those of the base DiD model from Fig. 3a. Second, ITS is especially sensitive to seasonal effects, and while there are different ways to control for them, none are perfect. Third, and most importantly, the treatment window overlaps very closely with (meteorological and astronomical) winter, exacerbating the previously mentioned issue, especially because it is a time of depressed crime in general. DiD does not require imposing as much structure, and controls for cyclical trends by design. Regardless, the ITS results are essentially the same.

1. Kubrin, C., Messner, S., Deane, G., McGeever, K. & Stucky, T. Proactive Policing and Robbery Rates Across US Cities. *Criminology* **48,** 57–97 (2010).

2. Braga, A., Welsh, B. & Schnell, C. Can policing disorder reduce crime? A systematic review and meta-analysis. *Journal of Research in Crime and Delinquency* **52,** 567–588 (2015).

3. Cerdá, M. *et al.* Misdemeanor policing, physical disorder, and gun-related homicide: a spatial analytic test of "broken-windows" theory. *Epidemiology* **20,** 533–541 (2009).

4. MacDonald, J., Fagan, J. & Geller, A. The Effects of Local Police Surges on Crime and Arrests in New York City. *PLoS ONE* **11,** e0157223 (2016).

5. Harcourt, B. *Illusion of order: The false promise of broken windows policing* (Harvard University Press, 2005).

6. Weisburd, D., Wooditch, A., Weisburd, S. & Yang, S. Do Stop, Question, and Frisk Practices Deter Crime? *Criminology & Public Policy* **15,** 31–56 (2016).

7. Rosenfeld, R. & Fornango, R. The Relationship Between Crime and Stop, Question, and Frisk Rates in New York City Neighborhoods. *Justice Quarterly,* 1–21 (2017).

8. Fagan, J., Geller, A., Davies, G. & West, B. Street stops and broken windows revisited: The demography and logic of proactive policing in a safe and changing city in *Race, Ethnicity, and Policing: New and Essential Readings* (eds Rice, S. K. & White, M. D.) 309–348 (New York University Press, 2010).

9. Fagan, J., Braga, A. A., Brunson, R. K. & Pattavina, A. Stops and Stares: Street Stops, Race, and Surveillance in the New Policing. *Fordham Urban Law Journal* **In press** (2017).

10. Kelling, G. L. & Wilson, J. Q. Broken Windows: The police and neighborhood safety. *The Atlantic* (Mar. 1982).

11. Weisburd, D. Does Hot Spots Policing Inevitably Lead to Unfair and Abusive Police Practices, or Can We Maximize Both Fairness and Effectiveness in the New Proactive Policing. *University of Chicago Legal Forum* **2016,** 661–689 (2016).

12. The New York City Police Department. *Broken Windows and Quality-of-Life Policing in New York City* tech. rep. (Department of Investigation, City of New York, 2015).

13. Bratton, W. & Kelling, G. Why We Need Broken Windows Policing. *City Journal* ***Winter,*** (2015).

14. Pyrooz, D., Decker, S., Wolfe, S. & Shjarback, J. Was there a Ferguson Effect on crime rates in large US cities? *Journal of Criminal Justice* **46,** 1–8 (2016).

15. Fagan, J. & Davies, G. Street stops and broken windows: Terry, race and disorder in New York City. *Fordham Urban Law Journal* **28,** 457–497 (2000).

16. Sampson, R. & Raudenbush, S. Systematic social observation of public spaces: A new look at disorder in urban Neighborhoods. *American Journal of Sociology* **105,** 603–651 (1999).

17. Leovy, J. *Ghettoside* (Susan Mallery, 2015).

18. Hinkle, J. & Weisburd, D. The Irony of Broken Windows Policing: A Micro-place Study of the Relationship between Disorder, Focused Police Crackdowns and Fear of Crime. *Journal of Criminal Justice* **36,** 503–512 (2008).

19. Geller, A., Fagan, J., Tyler, T. & Link, B. Aggressive policing and the mental health of young urban men. *American Journal of Public Health* **104,** 2321–2327 (2014).

20. Tyler, T., Jackson, J. & Mentovich, A. The consequences of being an object of suspicion: Potential pitfalls of proactive police contact. *Journal of Empirical Legal Studies* **12,** 602–636 (2015).

21. Kohler-Hausmann, I. Managerial justice and mass misdemeanors. *Stanford Law Review* **66** (2014).

22. Desmond, M., Papachristos, A. & Kirk, D. Police Violence and Citizen Crime Reporting in the Black Community. *American Sociological Review* **81,** 857–876 (2016).

23. Kirk, D. & Papachristos, A. Cultural Mechanisms and the Persistence of Neighborhood Violence. *American Journal of Psychology* **116,** 1190–1233 (2011).

24. Zimring, F. E. *The City That Became Safe: New York's Lessons for Urban Crime and Its Control* (Oxford University Press, New York, 2012).

25. Kelley, D. & McCarthy, S. *"The report of the crime reporting review committee to Commissioner Raymond W. Kelly concerning CompStat auditing"* (New York Police Department, 2013).

26. Rosenfeld, R. & Fornango, R. The impact of police stops on precinct robbery and burglary rates in New York City, 2003–2010. *Justice Quarterly* **31,** 96–122 (2014).

27. Eure, P. *An Analysis of Quality-of-Life Summonses, Quality-of-Life Misdemeanor Arrests, and Felony Crime in New York City, 2010–2015* tech. rep. (Department of Investigation, City of New York, 2016).

28. Puhani, P. The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. *Economics Letters* **115,** 85–87 (2012).

29. Roeder, O., Eisen, L.-B. & Bowling, J. What Caused the Crime Decline? *Brennan Center for Justice* (Feb. 2015).

30. Lerman, A. & Weaver, V. Staying out of sight? Concentrated policing and local political action. *The ANNALS of the American Academy of Political and Social Science* **651,** 202–219 (2014).

31. Chandrasekher, A. The Effect of Police Slowdowns on Crime. *American Law and Economics Review* **18,** 385–437 (2016).

32. Lerman, A. & Weaver, V. *Arresting citizenship: The democratic consequences of American crime control* (University of Chicago Press, 2014).

33. Li, X., Mehrotra, D. & Barnard, J. Analysis of incomplete longitudinal binary data using multiple imputation. *Statistics in Medicine* **25,** 2107–2124 (2006).

34. Ai, C. & Norton, E. Interaction terms in logit and probit models. *Economics Letters* **80,** 123–129 (2003).

35. Rengifo, A. & Fowler, K. Stop, Question, and Complain: Citizen Grievances Against the NYPD and the Opacity of Police Stops Across New York City Precincts, 2007–2013. *Journal of Urban Health* **93,** 32–41 (2016).

36. Bowen, H. Moderating Hypotheses in Limited Dependent Variable and Other Nonlinear Models: Secondary Versus Total Interactions. *Journal of Management* **38,** 860–889 (2012).

Supplementary Information is available in the online version of the paper.

Author Contributions C.M.S. developed the original study concept. C.M.S. and Z.P.O. gathered and analysed the data, and drafted and revised the manuscript.

Author Information The authors declare that they have no competing financial interests. Correspondence and requests for materials should be addressed to C.M.S. (csullivan@lsu.edu).