Causal Peer Effects in Police Misconduct

Edika G. Quispe-Torreblanca*

Neil Stewart[§]

University of Warwick, UK

Quispe-Torreblanca, E. G. & Stewart, N. Causal Peer Effects in Police Misconduct. *Nature Human Behaviour* (in press).

[*] University of Warwick, Warwick Business School; University of Oxford, Said Business School. Email: Edika.Quispe-Torreblanca@sbs.ox.ac.uk.
[§] University of Warwick, Warwick Business School. Email: neil.stewart@wbs.ac.uk.

*Correspondence to Edika.Quispe-Torreblanca@sbs.ox.ac.uk, http://orcid.org/0000-0002-0974-0705

Causal Peer Effects in Police Misconduct

We estimate causal peer effects in police misconduct using data from about 35,000 officers and staff from London's Metropolitan Police Service for the period 2011-2014. We use instrumental variable techniques and exploit the variation in peer misconduct that results when officers switch peer groups. We find that a 10% increase in prior peer misconduct increases an officer's later misconduct by 8%. As the police are empowered to enforce the law and protect individual liberties, integrity and fairness in policing are essential for establishing and maintaining legitimacy and public consent[1-5]. Understanding the antecedents of misconduct will help develop interventions that reduce misconduct.

Uncertainty still exists on the influence of peers on unethical behaviour in real-world settings including policing. Previous research is limited to short-term or cross-sectional studies, which prevents inference about causality in peer misconduct. Our data, however, follows employees over time as they change roles and identifies their peers and their supervisors, allowing us to estimate reliable causal relationships. We estimate how officers are affected by the misconduct cases of their peers. Our estimation of these peer effects complements the existing literature, in which there is much work on how individual deviances predict misconduct and how organizational, social, and situational factors affect misconduct. Our estimation also provides new insights outside the American context, where most of the literature to date is concentrated.

The study of individual deviance within police is appealing because of the long-established fact that the majority of incidents of corruption, brutality or excessive use of force are accounted by a handful of officers or "rotten apples". For example, in the US, the Christopher Commission that investigated the Los Angeles Police Department found that, over the period 1987 to 1991, 5% of the officers (of nearly 6000) were responsible for 20% of

all reports of excessive use of force[6]. In the UK, in 1997 the then Commissioner of the Metropolitan Police Service Sir Paul Condon famously stated that there were up to 100-250 seriously corrupt officers in the Service (then, of about 27,000 officers)[7,8]. That a few officers are responsible for much of the misconduct raises two possibilities: First, identifying and removing, or otherwise preventing, misconduct from this small number of officers would have a large effect. Second, and more worryingly, in the presence of strong peer effects, when the bad apples are not identified and disciplined, corruption can become pervasive and organized.

Research focused on individual deviances shows that complaint-prone officers are more likely to be non-white[9-11], male, less experienced[11-14] and less educated[10,15]. Research also shows that black officers have an earlier onset of misconduct, prior military service appears to delay the onset, and neither education nor academy performance affected the timing of onset[16]. Recent work has also sought to understand the relationship between personality and misconduct. Donner and Jennings[17], for instance, have shown that low self-control is a key predictor of engagement in general misconduct, particularly related to physical and verbal abuse. In the same vein, Pogarsky and Piquero[18] found that impulsivity mediates the influence of legal and extra-legal sanctions on the decision to commit hypothetical acts of misconduct.

In contrast to the individual deviance view, research on organizational correlates of police misconduct is sparse. Some case studies have documented evidence of the influence of the police departments' characteristics, such as size, bureaucracy and professionalism on the decision to arrest (for a review see Dunham & Alpert[19]). More recent evidence has shown that officers who perceive fairness in managerial practices are less likely to justify noble-cause corruption or adhere to the code of silence that protects bad cops[20]. Some consideration has also been given to situational variables. For instance, the possibility of arrest at police-

citizen encounters escalates with the mere presence of supervisors[21,22] and officers use greater levels of force against suspects encountered in high-crime and disadvantaged neighbourhoods[23].

The understanding of deviance behaviour should not neglect social aspects. People making decisions inside organizations are constrained by authority rules and regulations, but are also constrained by social norms, cultural expectations, and considerable peer-group pressures. Kohlberg's research on moral reasoning[24] has shown that, unlike childhood (when children were more concerned about the physical consequences of their actions, i.e., punishments and rewards, and when elements of reciprocity and fairness started to be incorporated pragmatically), moral reasoning in adolescence and adulthood is typically determined by beliefs about what others will think is right or wrong. In this level of moral thinking (termed as 'conventional' by the author) the individuals try to conform to the natural or accepted behaviour. For a discussion on how colleagues influence organizational ethics, see Treviño et al.[25]

Compelling evidence for the existence of peer effects has already been documented in other settings: For example, Mas and Moretti[26] found that the productivity of cashiers in a supermarket chain increases with the effort of co-workers who face them, Zimmerman demonstrated that first-year college students in the middle of the SAT distribution who share a room with students in the bottom of the distribution do worse in grades[27], and Trogdon, Nonnemaker, and Pais provided evidence that weight gain spreads through peer networks[28]. Herbst and Mas[29] provide a meta-analysis of peer effects in co-worker productivity: across studies they find an increase in a co-worker's productivity causes an effect about 12% of the size in their peers. Herbst and Mas also show a consistency between effect sizes in the field and from laboratory experiments.

But although peer influences have been subject to analysis in various domains via both lab and field studies, surprisingly, much uncertainty still exists on the influence of peers in police ethics and integrity. The police misconduct literature already suggests an association, but the evidence falls short of supporting a causal link. For example, officers assigned to the same workgroup tend to share occupational attitudes due to their interactions and exposure to similar environments[30]. This shows correlation in attitudes, but not a causal link. In the Philadelphia Police Department, officers who thought that their peers considered the use of excessive force as less serious were more likely to have citizen complaints, as were officers who anticipated more minor punishment for theft[31]. Using the officers' judgments of their peers' attitudes, while ignoring the dynamics of peer group networks, allows again only a correlational but not a causal claim. In the Dallas Police Department, one quarter of the variation in trainees' subsequent allegations of misconduct was attributed to field training officers in a multilevel analysis nesting trainees with their field training officers[32]. Nevertheless, this multilevel analysis is likely to be driven by common variance elements that are typical in nested structures and thus do not reflect causal relationships—again, because unobserved, and so omitted, shocks occurring among trainees' groups who share a common environment can mask peer effects.

Estimating social learning is challenging as individuals from a peer group affect their peer group as much as the peer group affects them. In addition to this reflection problem, peer groups are not necessarily randomly sorted, as high-performance workers could be allocated to a high-performance peer group, and so workers from the same peer group might likely share common unobserved characteristics. Moreover, members of a group might show similar misconduct because they are subject to similar shocks[33]. In our econometric approach, we address these issues using the instrumental variable estimation technique. We exploit the variation in peer quality that results after workers change line managers and switch peer

groups. Misconduct of the new peers acquired following the change is instrumented with prior events of misconduct of their new peers' peers, allowing us to estimate the causal effect between peers.

We should note that by examining peer effects, we do not intend to engage in the debate of which specific mechanisms are driving these effects. Nor do our data allow us to distinguish between the mechanisms by which peer effects are mediated. For example, we will not discriminate between social influences motivated by learning about what behaviour is best to follow given the individuals' own needs or motivated by pure peer pressure and social conformity. In fact, due to the difficulty to discriminate between these mechanisms, most research in the peer effects literature have focused on measuring the magnitude of peer effects only and have overlooked the mechanisms that may be generating the peer effects.

Our data covers four years of allegations of misconduct, from 2011 to 2014, for 49,403 officers and staff. For analysis we required line manager history (from which we can infer peer groups), at least one peer, and demographic information. This was true for 35,924 officers and staff. Of these, 14,915 had records of one or more complaint during the period 2011 to 2014. However, most of them (54%) received only 2 or fewer complaints in this four-year interval (see Supplementary Figure 1). Note that this is a very common pattern in police departments[34], suggesting that misconduct is not systemic and apparently only a minority of officers (or roles) are complaint-prone.

Allegations of misconduct are classified in seven categories: failures in duty, malpractice, discriminatory behaviour, oppressive behaviour, incivility, traffic, and other allegations (we merged these last two). Their distributions in Table 1 reveals that for both members of police staff and police officers, the most recurrent allegations consist of cases of failures in duty, which can be, for instance, unjustified use of the relevant power,

unauthorised entry on search, failure to inform detained persons of their rights and entitlements, failure to maintain proper custody/property records, interviewing oppressively or in inappropriate circumstances, among other cases.

[Insert Table 1 about here]

The possible sanctions following misconduct are formal actions, unsatisfactory performance procedures, management actions, retirement, or resignation (though most complaints end in no sanction). Formal actions involve written warnings, while unsatisfactory performance procedures entail the organizational procedures designed to deal with unsatisfactory performance and attendance. Management actions refer to any action that can be locally resolved to handle the allegation of misconduct. They consist of, for example, the establishment of an improvement plan and the clarification of expectations for future conduct. Observe in Table 1 that very few cases received a formal disciplinary action. Furthermore, over 50% of allegations against members of police staff and about 90% of the allegations against police officers had no subsequent actions taken. Most of these allegations were instances in which, following investigation and based upon the available evidence, there was no case to answer concerning the allegation. It can then be argued that the allegations documented might over represent real events of misconduct. Nonetheless, research has shown that allegations are difficult to prove because of the relative lack of physical evidence and the absence of witnesses and, thus, cases deemed unsubstantiated do not necessarily imply the absence of police misconduct[16,35]. Note that the use of all allegations, irrespectively of their outcomes, is the usual approach adopted in the literature.

Supplementary Table 1 shows how the types of allegations correlate within individuals. People with alleged failures in duty seem to also exhibit, to some extent, some form of incivility and oppressive or discriminatory behaviour.

We test whether workers' peers' misconduct might affect the recurrence of workers' misconduct events. Peer groups were defined by linking officers and staff assigned to the same line manager. Our outcome is a binary variable, y_{it}, that equals one if worker i had an event of misconduct during quarter t. Our independent variable of interest is the proportion of peers of i in $t-1$ receiving reports of misconduct in $t-1$, $Peer\ y_{i\,(t-1)}$. Since officers who patrol together or are in certain units together have a higher likelihood of being involved in reports of misconduct that might not be their fault, to prevent overestimating the effects of peers' misconduct, we consider as events of peer misconduct only those episodes in which i had no same-day concurrent allegations of misconduct. That is, allegations against peers and allegations against the target officer i correspond to different cases and were reported on different dates. W is a vector of control variables that include demographic characteristics, such as gender, length of service, employee's business group, employee type, and employee performance; and additional controls for annual and seasonal effects.

$$P[y_{it} = 1] = Peer\ y_{i\,(t-1)} + W'\beta \qquad (1)$$

Empirically there are three challenges for the identification of peer effects[33,36,37]. First, due to non-random assignment into groups, individuals with similar characteristics may end up in the same group. Then what looks like peer effects could actually be due to common characteristics of the individuals themselves and not due to their peers. Without random assignment, the influence of individual's characteristics cannot be identified separately from the influence of their peer's characteristics. The second challenge is that, even when random assignment had been possible, individuals in the same group share similar environments and, thus, there could be unobservable institutional factors affecting the group members' performance simultaneously. These two threats are referred to in the literature as correlated effects and do not correspond to any social phenomenon between peers. Third, we would

expect peer effects to be bi-directional. This means that peer effects are, in part, a property of the target individual and are not exogenous to the individual. This reverse causality problem holds even if we had random assignment into groups.

To address these challenges, we proceed as follows. To absorb the effect of unobservable institutional factors affecting the likelihood to misbehave either because some workers are exposed to particular stressful environments or high crime areas, or because workers sharing some background characteristics preferred to join specific business groups, our econometric specification includes dummy variable controls for the business groups the employees belong to. These business groups consist of: Territorial Police (divided in Boroughs North, Boroughs South, Boroughs West, Central, Criminal Justice & Crime, and Westminster), Specialist Crime and Operations, Specialist Operations, and Other Business Group (which aggregate the groups Career Transition, Deputy Commissioners Portfolio, Directorate of Resources, Met HQ, National Functions and Shared Support Services). Our regressions also include quarter and year dummies to account for any seasonal fluctuation in crime.

To deal with individual heterogeneity, we also include controls for gender, years of length of service, employee type and police ranks, and police performance. Performance scores are reported on an annual basis in Performance Development Reviews and evaluate competencies in operational effectiveness, organizational influence, and resource management. To alleviate the concerns of simultaneity bias, note that we estimate the effect of lagged peer outcomes on misconduct. More importantly, to deal with endogenous worker sorting into peer groups and potential correlated effects unaccounted by our set of controls, we use instrumental variable techniques and estimate a linear probability model using two-step Generalized Method of Moments (GMM) estimators. Our identification strategy exploits the variation in peers that is experienced by workers who switch peer groups.

Figure 1 illustrates the procedure followed. The top panel shows the hypothetical composition of peer groups for three different line managers across the quarters in one year, from $t-3$ to t. We are interested in modelling the risk of misconduct of individual i (denoted as 'T', for target individual, from now on) at time t. 'T' is allocated to a new line manager, Line Manager 2, in quarter $t-1$ and encounters new peers, 'D', 'E', 'F', 'G', and 'H'. First, we look at his new peers and select those that were also recently allocated to Line Manager 2 (i.e., 'H'). Second, for the identified peer 'H', we observe his existing peers in $t-2$ ('I', 'J', and 'K') and compute the proportion of these existing peers who had reports of misconduct in $t-2$ (P1). Likewise, we also observe his existing peers in $t-3$ (again, 'I', 'J' and 'K') and compute the proportion of these existing peers who had reports of misconduct in $t-3$ (P2). These two measures P1 and P2 are used as instruments of $Peer\ y_{i\,(t-1)}$ in Equation 1. Note that the construction of our instruments ignores the behaviour of any worker that was under the supervision of Line Manager 2 during $t-2$ and $t-3$, such as workers 'D', 'E', 'F', and 'G', since due to potential non-random sorting these workers might share some background characteristics with 'T'.

Valid instruments satisfy two properties. The instrument must be (1) relevant: the instrument must be correlated strongly with the endogenous variable $Peer\ y_{i\,(t-1)}$. The instrument must satisfy the (2) exclusion restriction: the instrument must affect the outcome variable, y_{it}, only through its effect on the endogenous variable. That is, the instrument should not affect independently the outcome variable y_{it}. The exclusion restriction required for identification implies that misconduct of the peers of 'H' in $t-2$ and $t-3$ (i.e., misconduct of 'I', 'J', and 'K') should not affect the current behaviour of 'T' except through their impact on 'H' in $t-1$. If 'H' had not been allocated to Line Manager 2, the behaviour of 'I', 'J', and 'K' should not affect the behaviour of the target officer 'T'. Accordingly, to construct our instruments we discard in the first part of our procedure any new peer of 'T' in

$t-1$ that had at least one peer that worked along 'T' during quarters $t-3$ to t. This strategy satisfies the exclusion restriction since only the peers of peers who had no evidence of direct contact with 'T' during the past year are used in the construction of the instruments. Note that 'I', 'J' and 'K' satisfy this criterion.

In the bottom panel of Figure 1, we consider the case in which 'T' experiences new peers but does not change line manager. Following the same procedure, we select 'H' and observe the behaviour of his peers in $t-2$ and $t-3$ to construct the instruments. In our examples, only 'H' was selected in the first step; however, when more than one peer in $t-1$ satisfy the criteria imposed, we compute for each of these peers the two measures of peers of peers conduct described (P1 and P2) and average these measures across them. Thus, we use $\overline{P1}$ and $\overline{P2}$ as instruments of $Peer\ y_{i\,(t-1)}$.

Observations that satisfy our criteria for identification are not prevalent in the data and, thus, our estimation of peer effects is restricted to a sample of 80,632 quarter observations (24% of the total quarter observations of the data) from 30,627 individuals. A summary of the average composition (by quarter) of the sample used is shown in Supplementary Table 2. The left column displays the average composition per quarter for the whole sample. The right column restricts the sample to those observations in which an individual faces a change of peers. Supplementary Figure 2 shows the distribution of the number of peers for each of these samples. There is not apparent evidence of a disproportionate selection of particular groups of individuals, which means our estimates of peer effects should generalise to the wider population of all officers and civilian staff.

[Insert Figure 1 about here]

We have outlined above how the instrumental variable estimation approach is critical to addressing the three challenges for identification of the causal effect of peer misconduct.

To have an initial approximation of the direction and magnitude of peer effects on misconduct, in the Supplementary Information Section we present the estimates from linear probability panel data models—including both fixed and random effects—that cover all individuals in our data (see Supplementary Table 3). These panel models do not correct for endogeneity. While these panel models can be applied to the whole data set, they do not address the three challenges to estimating the casual effect. We find that the panel models show significant but small effects of peer misconduct. But our instrumental variable approach reveals that the panel models greatly underestimate the causal effect of peer misconduct.

Table 2 presents the estimates using our instrumental variable approach. The first variable, the proportion of peers in $t-1$ with misconduct, is instrumented using the proportion of peers of 'H' with misconduct from Figure 1 (i.e., using the proportion of 'I', 'J', and 'K' with misconduct). In Model 1, we present the estimates from a two-step efficient GMM estimator (results from the first stage are presented in Supplementary Table 5). Due to the instrumenting of our endogenous variable, 75% of the observations are lost; however, as described earlier in Supplementary Table 2, the remaining sample is structurally similar to the whole sample. Since in this remaining sample more than half of the individuals (15,038 out of 30,627) have only 2 or 3 quarter observations, we are unable to apply panel data estimators. However, the SEs of our GMM estimates are robust to arbitrary within-individual correlations. The coefficient of 0.768 ($t(30626)=4.91$, $p<0.001$, 95% CI[0.461 - 1.075]) in Model 1 for the instrumented proportion of peers at $t-1$ with misconduct means that a 10-percentage point change in the proportion of peers with misconduct would *cause* an increase of 7.68 percentage points in the target misconduct.

In Model 1, the estimates for the control variables are in line with the findings of other studies in the literature: male workers, police officers and less experienced employees

are prone to receive more allegations of misconduct. We also see expected signs for a positive effect of previous employee performance reviews.

At the bottom of the Model-1 Column of Table 2, we test the validity of our instruments. To be valid, they should satisfy two requirements: they must be correlated with the endogenous variable $Peer\ y_{i\,(t-1)}$ and orthogonal to the error process. At the bottom of Table 2, we report the first-stage Kleibergen-Paap F statistics for week identification that examines the joint significance of both instruments in determining the endogenous variable. With a value of 97.75, sufficiently larger than 10, the threshold suggested by Staiger and Stock[38] to prevent biases by using weak instruments, the first-stage F-statistic confirms that our instruments are strong. We also report the Kleibergen-Paap LM test statistic for under identification which is robust in the presence of heteroscedasticity and clustering in errors. Rejection of the null indicates that our model is identified—that is, that our instruments are relevant. To evaluate the validity of the instruments, we also report the J-statistic of Hansen[39] that tests the null hypothesis of orthogonality of the instruments and the error process which shows that our instruments are exogenous.

In Model 2, we use an alternative estimator, an instrumental variable probit estimator, which also alleviates endogeneity concerns, but it is appropriate for binary dependent variables and continuous endogenous covariates. The resulting estimates provide further statistical support for the presence of peer effects. At the bottom of the column for Model 2, we also report the χ^2statistics of the Wald test of endogeneity of the instrumented variable, which rejects the null hypothesis that $Peer\ y_{i\,(t-1)}$ is exogenous.

Coefficients from Model 2 do not represent marginal effects as coefficients from Model 1 do. In order to ease the comparison of both models, Figure 2 illustrates the extent of

the peer effects from Model 2. Reassuringly, the peer effects are close in magnitude to those obtained by GMM in Model 1.

[Insert Table 2 about here]

[Insert Figure 2 about here]

Under the concern that our estimation of peer effects might still reflect correlated effects due to unobservable events not accounted by our controls or endogeneity due to disregarded indirect interactions between individual i and the peers of peers used in the constructions of our instruments, we perform the following falsification test. Observe in the top panel of Figure 1 that the behaviours of individuals 'I', 'J' and 'K' are expected to influence the conduct of 'T' during quarter t through a single and unique channel, 'H'. However, during quarter t former peers of 'T' (i.e., 'A', 'B' and 'C') who remained under the direction of Line Manager 1 and, consequently, had no direct contact with 'H' should not be affected by any sort of misconduct of 'I', 'J' or 'K' that took place during quarter $t-2$ or $t-3$. Thus, our falsification test consists of replacing the dependent variable y_{it} by the proportion of former peers of i who receive allegations of misconduct during quarter t. These peers are those who worked along i during quarter $t-2$ (the period immediately preceding the movement of i into a new peer group). The control variables are analogous to those used in Table 2. They include the proportion of male peers, the proportions of peers for each rank, business group and performance rating, the average length of service and the usual year and seasonal controls.

Models 1-3 of Table 3 present the results of this falsification test. Models 1-3 are fitting the misconduct of former peers of the target, who should be unaffected by our instruments. The sample size for our falsification test is smaller because it is restricted to those quarter observations in which individuals change line managers (illustrated in the top

panel of Figure 1). Model 4 of Table 3 is fitting misconduct of the target, and here we should replicate our headline peer effect from Model 1 of Table 2, but on the smaller sample size.

[Insert Table 3 about here]

The peer effects for Models 1-3 of Table 3 are much lower, imprecise, and not statistically different from zero, as we expected from the falsification test. Model 4 of Table 3 produced estimates very like those found in Table 2 Model 1, replicating our headline peer effect within the smaller sample. The specification tests confirm the validity of the instruments in all models, as informed by the Hansen *J*-statistics and *F*-statistics, except for Model 3. Yet, any possible endogeneity problem that remains unsolved in Models 1-3 would induce some upward bias in the estimated peer effects these columns display. However, these peer effects are of small and non-significant size. Regarding the effect of the control variables, across the different specifications they exhibit the expected signs and comparable sizes.

In the Supplementary Information, we do additional robustness checks. To further control for placement in high crime areas, we repeat our main analysis and add fixed effects for geographical locations at a higher level. In terms of geographic policing, we add 32 dummy variables distinguishing 32 Borough Operational Command Units. We also control for specific groups of Territorial Policing (TP): TP - Central, TP - Westminster, and 6 subgroups that are part of the TP - Criminal Justice & Crime (Met Detention, Met Prosecutions, RTPC - Roads and Transport Policing Com, TP - Capability and Support, TP Crime Recording Investigation Bureau, and TP Crime Recording Investigation Bureau). The pattern of estimates is as before, the coefficients on the proportion of peers with misconduct are positive and indicate that an increase in 10-percentage points in the proportion of peers with misconduct raises the likelihood of misconduct by 7.31 percentage points (B=0.731,

t(30626)=4.29, p<0.001, 95% CI[0.397 - 1.066]) (Supplementary Table 7, Model 1). The coefficient estimates on the control variables are keeping with those reported in Table 2.

We also evaluate whether our results vary after controlling for the supervisors' performance. We repeat our analysis and include dummies for the performance scores of supervisors in the preceding year, quarter $t-4$. These performance scores are reported on annual basis in Performance Development Reviews and evaluate competences in operational effectiveness, organizational influence, and resource management. Our estimates remain consistent after the addition of these controls. The results suggest a correlation between misconduct and supervisors' performance such that individuals under the supervision of 'Competent Above Standard' or 'Exceptional' supervisors have 3.9 percentage points less incidence of misconduct than individuals under the supervision of 'Competent but with Development Required' or 'Not Yet Competent' supervisors (B=-0.039, t(32182)=-1.75 p=0.080, 95% CI[-0.083 - 0.005]), Supplementary Table 7, Model 3). Since supervisor effects are endogenous—individuals and their supervisors influence each other simultaneously—this association is merely correlational rather than causal.

We also investigate whether our effects are consistent by analysing separately the sample of individuals who moved to a new different peer group (as described at the top of Figure 1) and the sample that remain in their current peer group but have new incoming peers (as described at the bottom of Figure 1), while retaining the complete set of exhaustive geographical controls. In Supplementary Table 9, peer effects in both samples are of comparable size and the analysis on each sample individually pass the diagnostic tests for both the relevance (LM test statistic for under identification - Kleibergen-Paap) and the validity (Hansen's J test) of our instruments.

Finally, we studied whether peer effects interact with the peer group size. While controlling for the peer group size (Supplementary Table 11, Column 1) increases the point estimate of peer effects from 7.68 to 7.76 percentage points (point estimate at the median of the peer group size, 7 peers, following a 10% increase in prior peer misconduct, B=0.776, z=4.09, p<0.001, 95% CI[0.404 - 1.148]), both peer group size and its interaction with peers' misconduct have negligible and non-significant effects on misconduct.

Thus far, we have quantified sizable peer effects after analyzing peer groups dynamics. There remains the question of why officers and staff are moved from one group to another. In Supplementary Table 13, we study which factors correlate with transfers or line manager changes. While past misconduct and past performance appear to correlate with changes of line managers, this correlation is relatively small. On one hand, misconduct in the preceding semester increases the likelihood of switching line managers by only 4.6 percentage points (B=0.046, z=11.85, p<0.001, 95% CI[0.038 - 0.053)], Model 1, Suppementary Table 11). On the other hand, employees with 'Exceptional' or 'Competent Above Standard' performance are 3.6 percentage points less likely to switch line managers in comparison with employees with poor performance (rated as 'Not Yet Competent' or 'Competent but Development Required', B=-0.036, z=-2.85, p=0.004, 95% CI[-0.062 - -0.011], Model 2, Suppementary Table 13). Rather than past behavior, switching peer groups appear to be driven more by demographic characteristics. Inspectors and police sergeants, workers with few years of experience, and workers placed in particular geographic locations have a higher likelihood of moving peer groups, Model 3, Supplementary Table 13.

Before turning to the main conclusions of the study, we must highlight a few caveats. Obviously, the major constraint of our study is the assumption that complaints filed against officers are accurate proxies of misconduct events. Yet, these complaints could either over or under represent real misconduct cases. For example, fellow officers, as opposed to citizens,

fail to report misconduct due to their cultural rules of integrity. Informally, the "Code" discourages them from reporting misconduct of their peers[20,40]. On the other hand, citizen allegations of misconduct may be discouraged when there is fear of retaliation or a low confidence in the complaint process[41]. Our data, however, do not allow us to distinguish the source of the complaints. Moreover, most of the allegations reported were unsubstantiated because of the relative lack of physical evidence and the absence of witnesses, which makes the cases difficult to probe. However, the absence of evidence does not necessarily imply the absence of police misconduct. In research of this nature, we are limited to the analysis of reported cases of misconduct taking them as factual. We note, however, that the study of allegations of misconduct is the usual approach adopted by the related literature and so no study in this domain has been immune to this constraint.

There is also concern about whether the frequency of complaints mirrors officers' productivity. There is evidence suggesting that more proactive officers, officers placed in areas with high crime rates, and officers that due to their patrol assignment are more likely to be in contact with citizens, are prone to receive citizen's allegations of misconduct[14,41]. Unfortunately, we were not able to control for the officers' arrest activity. However, to the extent that some degree of arrest activity might be associated to characteristics that might have remained relatively stable over the four-year interval of available data, such as rank hierarchy or the assignment to different police units, we do capture the effects of individual productivity.

In conclusion, we demonstrate that deviant behaviour can be spread through socialization: a 10-percentage points increase in the fraction of peers with misconduct would raise the incidence of misconduct by an absolute 8%. These results are consistent when an officer switches to an entirely new group or when he receives new members to his current peer group. Perhaps officers' beliefs about what is acceptable and unacceptable behaviour

become more permissive when officers become part of closely connected groups with deviant behaviour. Following, Ashforth and Anand[42], because life is lived in concrete settings, localized social cultures tend to be highly salient, and the individual's commitment to ethics may relax under the press of local circumstances. By process of socialization, officers may learn to accept unethical practices. Moreover, local groups often provide accounts to rationalize or neutralize the guilt that individuals engaging in misconduct might otherwise feel, such as denial of the victim, denial of injury, denial of responsibility, and refocusing attention, among other accounts.

We should note that our results do not imply (or deny) the possibility that these effects occurred because officers learned from each other which behaviour is best to follow to satisfy their own interests or, instead perhaps, because they were corrupted by the pure peer pressure of their colleagues. Nor can we engage in the discussion about which mechanisms have driven these peer influences. Nevertheless, it is quite reasonable to speculate that a large portion of these effects reveal evidence of social conformity. Notice that extensive qualitative research highlights that police culture is typically imbedded in unwritten rules and protected by a code of silence and extreme group loyalty[43]. Recent findings provided by Hough et al.[44], after examining cases of alleged misconduct involving chief police officers in England and Wales over a six-year period, up to 2013, suggest that, throughout their careers, police officers felt under pressure to not step outside the norm. The ethical climate, promoted by a typical command-and-control style of management, is alleged to lack ethical values or, even worse, to sustain the wrong kinds of values. The command-and-control style of management appears to encourage close mutually supportive and inward-looking networks that favor homogeneity, preclude difference and even accept or tolerate bullying behavior. Hough et al's findings suggest that officers involved in misconduct are part of groups in which there is little to no stigma associated with misconduct.

Our peer effect results are to some extent consistent with the work of Chappell and Piquero[31], Getty, Worrall and Morris[32], and Ingram, Paoline and Terrill[30], who suggested that peer effects are important determinants of misconduct based on correlational studies, and lend also support to differential association theory, according to which criminal behaviour can be learnt through long, frequent and intense interactions with individuals holding attitudes that encourage criminal activity[45].

Beyond quantifying the magnitude of peer effects, our research has important policy implications. We have provided robust evidence that misconduct spreads between peers. It is unlikely that officers will have incentives to attempt to eliminate misconduct if there is no stigma associated with misconduct among their peers. Our results suggest that moving a bad cop to alternative locations will increase the risk of spreading misconduct. Thus, deterrence of police misconduct requires additional actions beyond the mere transfer of officers to other units. Studying which policy actions (ethical training, clear ethical standards, stronger sanctions, etc.) are more effective in preventing or discouraging misconduct represents an important arena for future research.

In addition to identifying sizable peer effects, we also replicate the individual differences that are associated with misconduct. In consistency with earlier research, we found that certain demographic characteristics are consistently present in individuals with higher risk of misconduct, such as few years of experience, poor ratings of past performance, male gender, or certain employee types (like police sergeant and constable).

Although it seems intuitive that individuals' experience and the social context in which they operate can influence their behaviour, our research provides compelling evidence for this intuition in police misconduct research.

Methods

Data Sources

Our study uses four databases maintained by the Metropolitan Police Service. The first dataset contains demographic information for 13,558 civilian staff and 35,845 police officers in active service at the end of the first quarter of 2015. This information includes gender, employee types and roles, length of service and business groups. The Met comprises four business groups: Specialist Operations, Met Operations (or Specialist Crime & Operations), Professionalism, and Frontline or Territorial Policing. Territorial Policing data is divided into 32 Borough Operational Command Units. These business groups are supported by civilian staffed support departments, which provide personnel, finance and legal services.

The second dataset includes daily records of allegations of misconduct filed against civilian staff and police officers from the second quarter of 2010 to the first quarter of 2015. Each record contains fields for the date of the incident, the nature of the allegations and the complaint's final disposition (if any). Allegations include citizen complaints and internal complaints filed by supervisors or other officers, however the records do not distinguish between these two sources. The third dataset comprises the individuals' performance scores reported on annual basis in Performance Development Reviews from 2011 to 2014. Scores are given on specific categories: operational effectiveness, organizational influence, resource management, and final overall rating of performance. Final scores position individuals as 'Not Yet Competent', 'Competent but Development Required', 'Competent at Required Standard', 'Competent Above Standard' and 'Exceptional'. The fourth dataset contains semestral records of employees and their line managers from 2011 to 2015.

The final panel of data, obtained by merging these data sources, has repeated quarterly observations nested within each of the individuals. It comprises 35,924 people

(31.7% were civil staff; 64.7%, males; and 13.6%, from black and minority ethnic groups) for the period 2011 to 2014. In this final panel of data, we were able to identify the work groups of individuals by linking officers assigned to the same supervisor in a given quarter. The median team size is eight.

Supervisors are in charge of familiarizing their team about their roles, responsibilities and local policing aims. Supervisors are also in charge of addressing underperformance among their team. Team members are socially more cohesive and evaluated under alike ethical standards. Our study evaluates the effects of peers' misconduct under this definition of peer groups.

Statistical methods

We use instrumental variable techniques and test peer effects in a linear probability model using two-step GMM estimators. Our identification strategy exploits the variation in peers that is experienced by officers who switch peer groups. Figure 1 illustrates the procedure followed for the construction of our instruments.

Code availability

Analyses were conducted in R 3.4.5 and Stata 13.1. All code is available in the public repository https://github.com/edikaQT/misconduct_peer_effects.

Data availability

The data that support the findings of this study are not publicly available. If you would like to view and reproduce our results, please contact E.G.Q.-T. to organize a supervised visit to our local network.

Author contributions

Data were provided by the Metropolitan Police Service to N.S. The concept for the paper was developed jointly by the authors. E.G.Q.-T. designed and completed all the analysis and wrote the manuscript of the paper. Both authors revised the manuscript and approved the final version.

Competing interests

The authors declare no competing interests.

Acknowledgments

This work was supported by Economic and Social Research Council grants ES/K002201/1, ES/P008976/1, ES/N018192/1, and Leverhulme Trust grant RP2012-V-022. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1 Goldsmith, A. Police reform and the problem of trust. *Theoretical criminology* **9**, 443-470 (2005).
2 Rosenbaum, D. P. Special issue on police integrity: an introduction. *Policing: An International Journal of Police Strategies & Management* **39** (2016).
3 Bayley, D. H. Law enforcement and the rule of law: Is there a tradeoff? *Criminology & Public Policy* **2**, 133-154 (2002).
4 Walker, S. E. & Archbold, C. A. *The new world of police accountability*. (Sage Publications, 2013).
5 Murphy, K., Hinds, L. & Fleming, J. Encouraging public cooperation and support for police. *Policing & Society* **18**, 136-155 (2008).
6 Christopher, W. *Report of the independent commission on the Los Angeles Police Department*. (Diane Publishing, 1991).
7 Gillard, M. & Flynn, L. *Untouchables: Dirty cops, bent justice and racism in Scotland Yard*. (A&C Black, 2012).
8 *UK Government Select Committee on Home Affairs*, <https://publications.parliament.uk/pa/cm199798/cmselect/cmhaff/258-i/ha0103.htm> (1997).
9 Harris, C. J. Problem officers? Analyzing problem behavior patterns from a large cohort. *Journal of Criminal Justice* **38**, 216-225 (2010).
10 Kane, R. J. & White, M. D. Bad Cops: A Study of Career-Ending Misconduct Among New York City Police Officers. *Criminology & Public Policy* **8**, 737-769 (2009).
11 Lersch, K. M. & Mieczkowski, T. Who are the problem-prone officers? An analysis of citizen complaints. *American journal of police* **15**, 23-44 (1996).
12 Brandl, S. G., Stroshine, M. S. & Frank, J. Who are the complaint-prone officers?: An examination of the relationship between police officers' attributes, arrest activity, assignment, and citizens' complaints about excessive force. *Journal of criminal justice* **29**, 521-529 (2001).
13 McElvain, J. P. & Kposowa, A. J. Police officer characteristics and internal affairs investigations for use of force allegations. *Journal of Criminal Justice* **32**, 265-279 (2004).
14 Harris, C. J. Exploring the relationship between experience and problem behaviors: A longitudinal analysis of officers from a large cohort. *Police Quarterly* **12**, 192-213 (2009).
15 Kappeler, V. E., Sapp, A. D. & Carter, D. L. Police officer higher education, citizen complaints and departmental rule violations. *Am. J. Police* **11**, 37 (1992).
16 Harris, C. J. The onset of police misconduct. *Policing: An International Journal of Police Strategies & Management* **37**, 285-304 (2014).
17 Donner, C. M. & Jennings, W. G. Low self-control and police deviance: Applying Gottfredson and Hirschi's general theory to officer misconduct. *Police Quarterly* **17**, 203-225 (2014).
18 Pogarsky, G. & Piquero, A. R. Studying the reach of deterrence: Can deterrence theory help explain police misconduct? *Journal of Criminal Justice* **32**, 371-386 (2004).
19 Dunham, R. G. & Alpert, G. P. *Critical issues in policing : contemporary readings*. (2015).
20 Wolfe, S. E. & Piquero, A. R. Organizational justice and police misconduct. *Criminal justice and behavior* **38**, 332-353 (2011).
21 Engel, R. S. The effects of supervisory styles on patrol officer behavior. *Police Quarterly* **3**, 262-293 (2000).
22 Engel, R. S. How police supervisory styles influence patrol officer behavior. *Critical issues in policing: Contemporary readings* **6** (2003).
23 Terrill, W. & Reisig, M. D. Neighborhood context and police use of force. *Journal of research in crime and delinquency* **40**, 291-321 (2003).
24 Kohlberg, L. *Stage and sequence: The cognitive-developmental approach to socialization*. (Rand McNally, 1969).
25 Treviño, L. K., Den Nieuwenboer, N. A. & Kish-Gephart, J. J. (Un) ethical behavior in organizations. *Annual review of psychology* **65**, 635-660 (2014).
26 Mas, A. & Moretti, E. Peers at work. *American Economic Review* **99**, 112-145 (2009).
27 Zimmerman, D. J. Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics* **85**, 9-23 (2003).
28 Trogdon, J. G., Nonnemaker, J. & Pais, J. Peer effects in adolescent overweight. *Journal of Health Economics* **27**, 1388-1399 (2008).
29 Herbst, D. & Mas, A. Peer effects on worker output in the laboratory generalize to the field. *Science* **350**, 545-549 (2015).
30 Ingram, J. R., Paoline III, E. A. & Terrill, W. A multilevel framework for understanding police culture: The role of the workgroup. *Criminology* **51**, 365-397 (2013).

31 Chappell, A. T. & Piquero, A. R. Applying social learning theory to police misconduct. *Deviant Behavior* **25**, 89-108 (2004).

32 Getty, R. M., Worrall, J. L. & Morris, R. G. How far from the tree does the apple fall? Field training officers, their trainees, and allegations of misconduct. *Crime & Delinquency* **62**, 821-839 (2016).

33 Manski, C. F. Identification of endogenous social effects: The reflection problem. *The review of economic studies* **60**, 531-542 (1993).

34 Dunham, R. G. & Alpert, G. P. *Critical issues in policing: Contemporary readings*. (Waveland Press, 2015).

35 Prenzler, T. & Ransley, J. *Police reform: Building integrity*. (Hawkins Press, 2002).

36 Angrist, J. D. The perils of peer effects. *Labour Economics* **30**, 98-108 (2014).

37 Manski, C. F. Economic analysis of social interactions. *Journal of economic perspectives* **14**, 115-136 (2000).

38 Staiger, D. & Stock, J. H. Instrumental Variables Regression with Weak Instruments. *Econometrica* **65**, 557-586, doi:10.2307/2171753 (1997).

39 Hansen, L. P. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029-1054 (1982).

40 Klockars, C. B., Ivkovich, S. K., Harver, W. E. & Haberfeld, M. R. *The measurement of police integrity*. (Washington, DC: National Institute of Justice, 1997).

41 Lersch, K. M. Are citizen complaints just another measure of officer productivity? An analysis of citizen complaints and officer activity measures. *Police Practice and Research* **3**, 135-147 (2002).

42 Ashforth, B. E. & Anand, V. The normalization of corruption in organizations. *Research in organizational behavior* **25**, 1-52 (2003).

43 Loree, D. 1-31 (Royal Canadian Mounted Police, 2006).

44 Hough, M., May, T., Hales, G. & Belur, J. Misconduct by police leaders in England and Wales: an exploratory study. *Policing and Society* **28**, 541-552 (2018).

45 Akers, R. L. *Criminological theories: Introduction and evaluation*. (Routledge, 2013).

Figure Legends

Figure 1. The identification strategy for peer effects. Each column represents the peer groups under the direction of three different line managers over time. ‘T’ is the target individual under study. The double line frames highlight the groups that ‘T’ belongs to at each time. In time $t-1$, ‘T’ experiences a different peer group, either because he switches line manager (top panel) or because new workers are assigned to his group (bottom panel). In both cases, the behaviour of ‘I’, ‘J’, and ‘K’, who are the peers of worker ‘H’ in $t-2$ and $t-3$, are used as instruments of the peers of ‘T’ in $t-1$. Observe that ‘I’, ‘J’ and ‘K’ had no direct contact with ‘T’ during the past year (i.e., $t-3$ to t) and so this strategy satisfies the exclusion restriction required for identification.

Figure 2. Fitted probability of misconduct at t conditional on the proportion of peers exhibiting events of misconduct in $t-1$. Peer effects are based on estimates of Model 2, IVPROBIT, of Table 2 (N=80,612 observations). Estimates are at the mean base levels of covariates. The shaded area represents 95% confidence intervals.

Tables

Table 1. *The Distribution of Allegations Against Civil Staff and Police Officers by Disciplinary Outcome*

	Action Type											
	Civilian Staff						Police					
Allegation Type	No Action	Management Action	Formal Action	UPP	Retired / Resigned	Total	No Action	Management Action	Formal Action	UPP	Retired / Resigned	Total
Failures in duty	1,020 52.44%	479 24.63%	444 22.83%	2 0.10%	0 0.00%	1,945	22,628 86.47%	2,849 10.89%	633 2.42%	59 0.23%	0 0.00%	26,169
Malpractice	59 41.55%	19 13.38%	64 45.07%	0 0.00%	0 0.00%	142	2,610 88.99%	246 8.39%	73 2.49%	4 0.14%	0 0.00%	2,933
Discriminatory behaviour	108 70.59%	35 22.88%	10 6.54%	0 0.00%	0 0.00%	153	2,938 91.81%	210 6.56%	48 1.50%	4 0.13%	0 0.00%	3,200
Oppressive behaviour	35 79.55%	7 15.91%	2 4.55%	0 0.00%	0 0.00%	44	2,962 91.93%	231 7.17%	24 0.74%	5 0.16%	0 0.00%	3,222
Incivility	394 45.50%	408 47.11%	63 7.27%	1 0.12%	0 0.00%	866	4,955 85.01%	764 13.11%	105 1.80%	4 0.07%	1 0.02%	5,829
Other	94 30.92%	48 15.79%	162 53.29%	0 0.00%	0 0.00%	304	815 77.18%	166 15.72%	75 7.10%	0 0.00%	0 0.00%	1,056
Total	1,710 49.51%	996 28.84%	745 21.57%	3 0.09%	0 0.00%	3,454	36,908 87.03%	4,466 10.53%	958 2.26%	76 0.18%	1 0.00%	42,409

Note. Allegations recorder against 1,994 civil Staff and 12,921 police officer over the period 2011 to 2014. UPP refers to 'Unsatisfactory Performance Procedure'. Other allegations include traffic allegations. Most formal actions (88.90%) were taken based on substantiated allegations, while only 3.35% management actions and 0.05% no actions were linked to substantiated allegations.

Table 2. *Peer Effects on the Likelihood of Misconduct*

	Individuals experiencing new peers	
	(1)	(2)
VARIABLES	GMM	IV PROBIT
Prop. of peers in $t-1$ with misconduct	0.768***	5.426***
	[0.461 - 1.075]	[4.048 - 6.803]
Gender (reference: Females)		
Male	0.017***	0.140***
	[0.013 - 0.020]	[0.102 - 0.178]
Employee type (reference: Civil Staff)		
Police Constable	0.017***	0.210***
	[0.009 - 0.025]	[0.128 - 0.291]
Police Sergeant	0.022***	0.270***
	[0.013 - 0.031]	[0.178 - 0.361]
Inspector	0.019***	0.254***
	[0.008 - 0.031]	[0.145 - 0.364]
Chief Inspector, Superintendent, Chief Superintendent	0.016*	0.143
	[0.004 - 0.028]	[-0.036 - 0.323]
Business Group (reference: Territorial Police (TP) - Boroughs East)		
TP - Boroughs North	-0.001	-0.005
	[-0.010 - 0.008]	[-0.057 - 0.047]
TP - Boroughs South	0.007	0.038
	[-0.001 - 0.015]	[-0.009 - 0.085]
TP - Boroughs West	0.002	0.013
	[-0.007 - 0.010]	[-0.037 - 0.064]
TP - Central	-0.033**	
	[-0.056 - -0.010]	
TP - Criminal Justice & Crime	0.006	0.066*
	[-0.003 - 0.015]	[0.010 - 0.121]
TP - Westminster	0.006	0.034
	[-0.007 - 0.018]	[-0.039 - 0.108]
Specialist Crime and Operations	-0.006	-0.053
	[-0.016 - 0.004]	[-0.127 - 0.022]
Specialist Operations	-0.013~	-0.163**
	[-0.027 - 0.001]	[-0.281 - -0.045]
Other Business Group	-0.001	-0.204**
	[-0.015 - 0.014]	[-0.354 - -0.055]
Length of service		
Length of service (10 years)	-0.013*	-0.057
	[-0.025 - -0.001]	[-0.152 - 0.038]
Length of service (10 years)^2	0.002	-0.003
	[-0.001 - 0.005]	[-0.029 - 0.023]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)		
Exceptional + Competent (above standard)	-0.038**	-0.285***
	[-0.062 - -0.015]	[-0.428 - -0.142]
Competent (at required standard)	-0.035**	-0.256***
	[-0.059 - -0.012]	[-0.392 - -0.120]
Constant	0.038*	-1.606***
	[0.002 - 0.074]	[-1.753 - -1.459]
Observations	80,632	80,612
Number of individuals	30,627	30,617

	Individuals experiencing new peers	
	(1)	(2)
VARIABLES	GMM	IV PROBIT
LM test statistic for under identification (Kleibergen-Paap)	199.3	
P-value of under identification LM statistic	<0.001	
F statistic for weak identification (Kleibergen-Paap)	97.75	
Hansen Statistic	0.520	
Degrees freedom of Hansen Statistic	1	
P-value Hansen Statistic	0.471	
Wald test of endogeneity, $\chi^2(1)$		29.97
Exogeneity test Wald *p*-value		<0.001
Quarter FEs	YES	YES
Year FEs	YES	YES

Note. All models estimate the probability of an event of misconduct in quarter t conditional on a set of covariates. The dependent variable is a dummy equal to one when at least one event of misconduct is reported in quarter t. The independent variable of interest is the proportion of peers in quarter $t-1$ with reported cases of misconduct. Our identification strategy exploits the variation in peer groups experienced by the individuals during the period 2011-2014. We use instrumental variable techniques for the identification of peer effects. Column 1 presents a 2-step GMM linear model and Column 2, an IV PROBIT model. Two instruments are used for identification: the average proportion of peers of peers with incidence of misconduct in $t-2$ and, likewise, the average proportion of peers of peers with incidence of misconduct in $t-3$. The models include dummy controls to account for seasonal variation in the report of misconduct events: Quarter FE and year FE correspond to quarter dummies and year dummies. The first stage results of Column 1 are displayed in Supplementary Table 5. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Table 3. *Estimated Likelihood of Misconduct, Peer Effects: Falsification Test*

	DV: Prop. of former peers in $t-2$ with cases of misconduct in t				DV: Misconduct in t
	(1)	(2)	(3)		(4)
VARIABLES	GMM	GMM	GMM	VARIABLES	GMM
Prop. of peers in $t-1$ with misconduct	0.162	0.156	0.132	Prop. of peers in $t-1$ with misconduct	0.802*
	[-0.105 - 0.428]	[-0.144 - 0.456]	[-0.171 - 0.436]		[0.124 - 1.480]
Gender (reference: Prop. of Females)				Gender (reference: Females)	0.014**
Prop. of Males	0.018***	0.018***	0.018***	Male	[0.005 - 0.022]
	[0.011 - 0.024]	[0.011 - 0.026]	[0.010 - 0.025]		
Employee type (reference: Prop. of Civil Staff)				Employee type (reference: Civil Staff)	
Prop. of Police Constable	0.027***	0.029***	0.030***	Police Constable	0.015~
	[0.019 - 0.036]	[0.018 - 0.039]	[0.019 - 0.040]		[-0.001 - 0.031]
Prop. of Police Sergeant	0.027***	0.031***	0.034***	Police Sergeant	0.026***
	[0.018 - 0.036]	[0.020 - 0.042]	[0.022 - 0.045]		[0.012 - 0.041]
Prop. of Inspector	0.015**	0.021**	0.025***	Inspector	0.019*
	[0.004 - 0.026]	[0.007 - 0.034]	[0.011 - 0.038]		[0.001 - 0.038]
Prop. of Chief Inspector, Superintendent, Chief Superintendent	0.025**	0.022*	0.026*	Chief Inspector, Superintendent, Chief Superintendent	0.017~
	[0.008 - 0.042]	[0.002 - 0.042]	[0.006 - 0.046]		[-0.003 - 0.036]
Prop. of Special Constabulary	-0.049***	-0.068***	-0.067***	Special Constabulary	-
	[-0.061 - -0.037]	[-0.094 - -0.042]	[-0.092 - -0.042]		
Business Group (reference: Prop. in TP - Boroughs East)				Business Group (reference: TP - Boroughs East)	
Prop. in TP - Boroughs North	-0.004	-0.009*	-0.009~	TP - Boroughs North	-0.001
	[-0.011 - 0.004]	[-0.018 - -0.000]	[-0.018 - 0.000]		[-0.018 - 0.015]
Prop. in TP - Boroughs South	0.006~	0.005	0.005	TP - Boroughs South	0.001
	[-0.001 - 0.013]	[-0.003 - 0.013]	[-0.003 - 0.014]		[-0.014 - 0.016]
Prop. in TP - Boroughs West	-0.006	-0.007	-0.007	TP - Boroughs West	0.001
	[-0.013 - 0.001]	[-0.016 - 0.003]	[-0.016 - 0.002]		[-0.015 - 0.017]

	DV: Prop. of former peers in $t-2$ with cases of misconduct in t				DV: Misconduct in t
	(1)	(2)	(3)		(4)
VARIABLES	GMM	GMM	GMM	VARIABLES	GMM
Prop. in TP - Central	-0.011	-0.026	-0.022	TP - Central	-0.029*
	[-0.048 - 0.026]	[-0.078 - 0.027]	[-0.075 - 0.032]		[-0.054 - -0.004]
Prop. in TP - Criminal Justice & Crime	-0.006	-0.009~	-0.009~	TP - Criminal Justice & Crime	0.011
	[-0.015 - 0.004]	[-0.019 - 0.001]	[-0.019 - 0.001]		[-0.008 - 0.029]
Prop. in TP - Westminster	0.007	0.007	0.007	TP - Westminster	0.020
	[-0.005 - 0.019]	[-0.008 - 0.021]	[-0.007 - 0.021]		[-0.005 - 0.045]
Prop. in Specialist Crime and Operations	-0.028***	-0.030***	-0.029***	Specialist Crime and Operations	-0.002
	[-0.037 - -0.019]	[-0.040 - -0.020]	[-0.039 - -0.019]		[-0.022 - 0.018]
Prop. in Specialist Operations	-0.047***	-0.049***	-0.048***	Specialist Operations	-0.005
	[-0.060 - -0.034]	[-0.064 - -0.034]	[-0.063 - -0.033]		[-0.034 - 0.024]
Prop. in Other Business Group	-0.035***	-0.037***	-0.035***	Other Business Group	0.001
	[-0.047 - -0.022]	[-0.051 - -0.023]	[-0.049 - -0.022]		[-0.028 - 0.031]
Length of service				Length of service	
Average Length of service (10 years)	-0.026***	-0.034***	-0.039***	Length of service (10 years)	-0.022*
	[-0.038 - -0.015]	[-0.048 - -0.020]	[-0.053 - -0.025]		[-0.042 - -0.002]
Average Length of service (10 years)^2	0.003~	0.006**	0.007**	Length of service (10 years)^2	0.004
	[-0.000 - 0.007]	[0.001 - 0.010]	[0.002 - 0.011]		[-0.001 - 0.010]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)				Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)	
Prop. of Exceptional + Competent (above standard)		-0.079***	-0.080***	Exceptional + Competent (above standard)	-0.048*
		[-0.125 - -0.034]	[-0.125 - -0.035]		[-0.089 - -0.006]
Prop. of Competent (at required standard)		-0.072**	-0.071**	Competent (at required standard)	-0.041~
		[-0.117 - -0.026]	[-0.116 - -0.025]		[-0.083 - 0.000]
Constant	0.057***	0.136***	0.146***	Constant	0.054

	DV: Prop. of former peers in $t-2$ with cases of misconduct in t				DV: Misconduct in t
	(1)	(2)	(3)		(4)
VARIABLES	GMM	GMM	GMM	VARIABLES	GMM
	[0.037 - 0.076]	[0.088 - 0.184]	[0.098 - 0.194]		[-0.011 - 0.118]
Observations	27,040	19,796	19,796	Observations	20,374
Number of individuals	18,506	14,111	14,111	Number of individuals	14,401
LM test statistic for underidentification (Kleibergen-Paap)	52.51	40.70	39.53	LM test statistic for under identification (Kleibergen-Paap)	35.16
P-value of under identification LM statistic	<0.001	<0.001	<0.001	*P*-value of under identification LM statistic	<0.001
F statistic for weak identification (Kleibergen-Paap)	25.39	19.62	19.09	F statistic for weak identification (Kleibergen-Paap)	16.98
Hansen Statistic	3.135	3.486	3.889	Hansen Statistic	0.0304
Degrees freedom of Hansen Statistic	1	1	1	Degrees freedom of Hansen Statistic	1
P-value Hansen Statistic	0.077	0.062	0.049	*P*-value Hansen Statistic	0.862
Quarter FEs	NO	NO	YES	Quarter FEs	YES
Year FEs	NO	NO	YES	Year FEs	YES

Note. All models apply instrumental variable techniques for the identification of peer effects. Models 1 to 3 are part of a falsification test that study the behaviour of former peers of an individual i who moves to a different peer group in $t-1$. The outcome constitutes the proportion of these peers who had reports of misconduct at time t. The independent variable of interest is the proportion of peers of i in $t-1$ presenting incidence of misconduct. This variable is instrumented by two measures of conduct of peers of peers of i. By construction, these two measures are expected to have no influence on the outcome variable of these models. Model 4 is presented for comparative purposes and uses the standard outcome variable of the study, misconduct of i at time t. 95% confidence intervals in parentheses. Significance levels: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, ~ $p<0.10$.

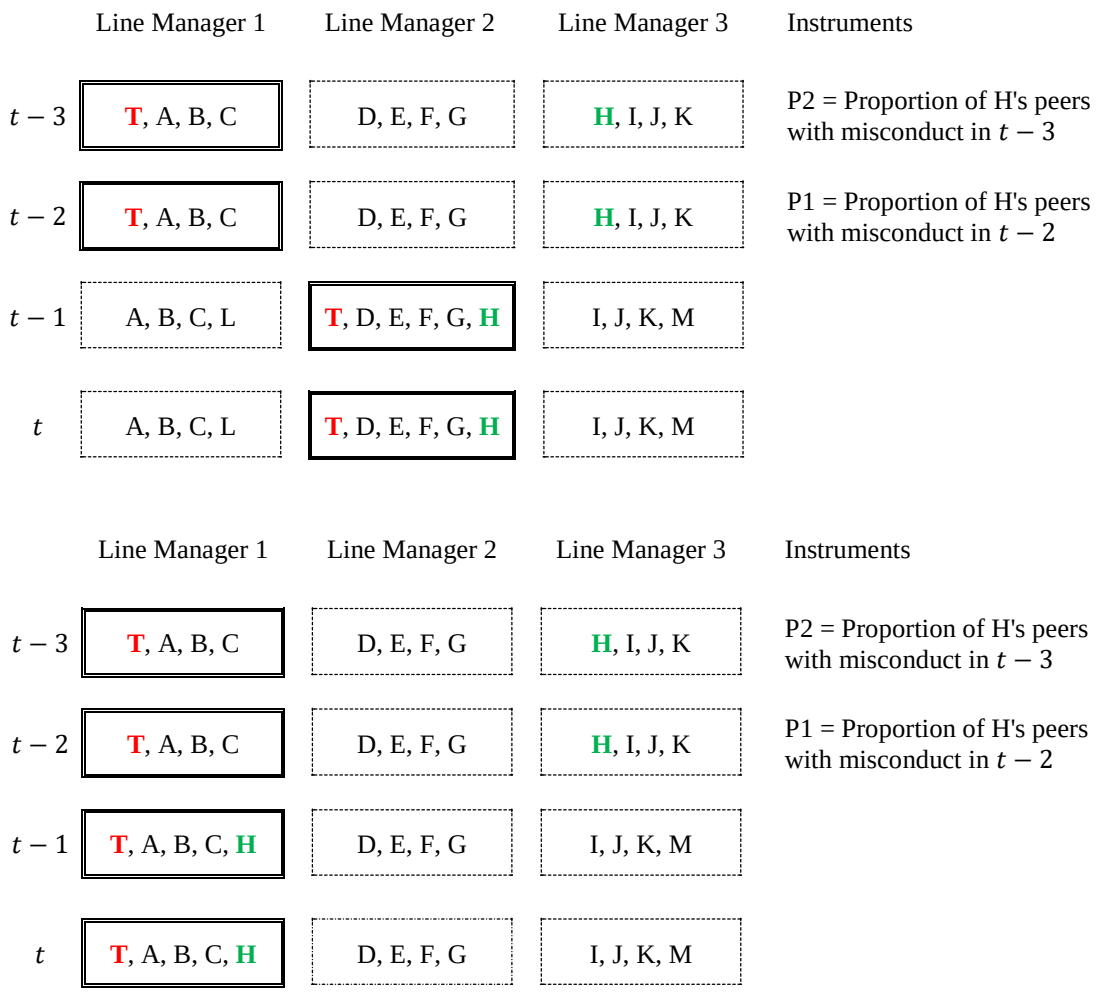


Figure 1. The identification strategy for peer effects. Each column represents the peer groups under the direction of three different line managers over time. 'T' is the target individual under study. The double line frames highlight the groups that 'T' belongs to at each time. In time $t-1$, 'T' experiences a different peer group, either because he switches line manager (top panel) or because new workers are assigned to his group (bottom panel). In both cases, the behaviour of 'I', 'J', and 'K', who are the peers of worker 'H' in $t-2$ and $t-3$, are used as instruments of the peers of 'T' in $t-1$. Observe that 'I', 'J' and 'K' had no direct contact with 'T' during the past year (i.e., $t-3$ to t) and so this strategy satisfies the exclusion restriction required for identification.

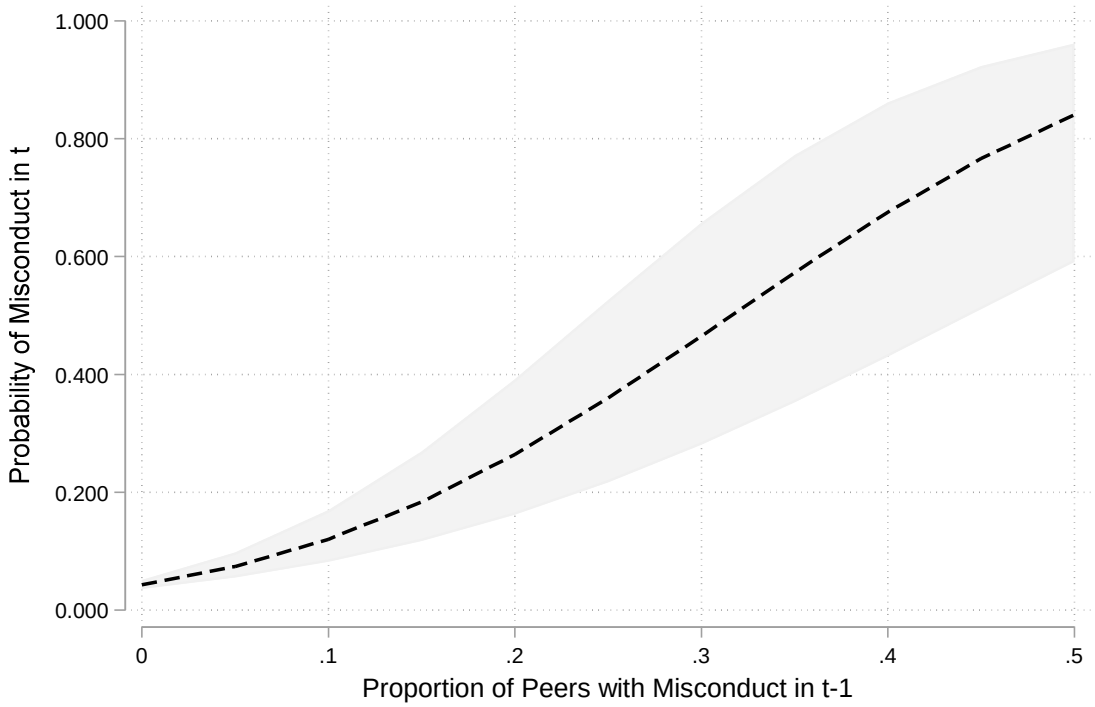


Figure 2. Fitted probability of misconduct at t conditional on the proportion of peers exhibiting events of misconduct in $t-1$. Peer effects are based on estimates of Model 2, IVPROBIT, of Table 2. Estimates are at the mean base levels of covariates. The shaded area represents 95% confidence intervals.

Supplementary Information

Supplementary Methods

Fixed Effects and Random Effects Estimates

Supplementary Table 3 presents results from panel models including both fixed and random effects that do not use instrumental variables. These panel models fit Equation 1 using all quarters in the data, even those in which peers never switch peer groups. While these panel models can be applied to the whole data set, they do not correct for endogeneity.

We find that the panel models show significant but small effects of peer misconduct. But our instrumental variable approach reveals that the panel models greatly underestimate the causal effect of peer misconduct.

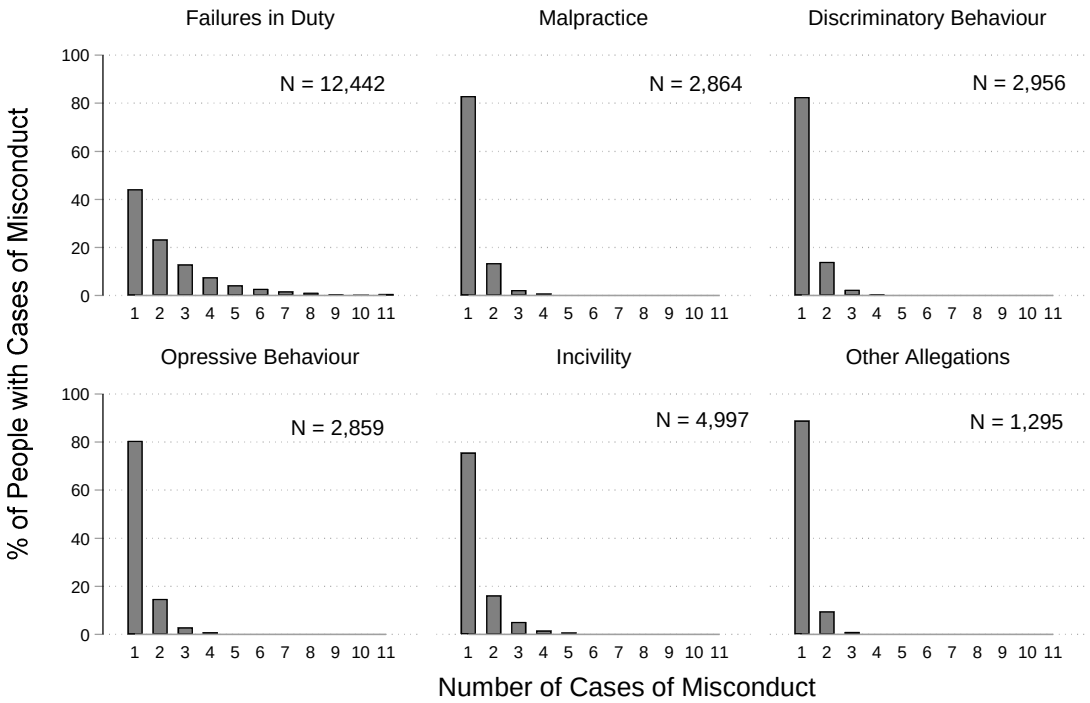
Model 1 of Supplementary Table 3 shows the random effects (RE) estimates of Equation 1. We observe positive and statistically significant peer effects. Model 2 displays fixed effects (FE) estimates that account for any unobserved time invariant characteristic of the individuals. Although FE estimates are smaller in magnitude, they still exhibit the expected positive sign. Models 3 and 4 employ similar estimators but are restricted to the sample of individuals who had at least one incidence of misconduct in the period 2011 to 2014. There is no apparent variation in the size of the peer effects in this subset of the data. These preliminary results indicate that a 10-percentage point increase in the proportion of peers with cases of misconduct in $t-1$ would rise the rate of misconduct in t by between 0.17 (t(359233)=3.48, p=0.001, CI[0.007 - 0.026], Model 2) to 0.66 percentage points (z=14.60, p< 0.001, CI[0.057 - 0.075], Model 1). Although these results suggest that peer misconduct has some small negative spillover effects, part of these effects are potentially spurious because we have not yet accounted for endogeneity in the estimates.

Table 2 in the main text presents the estimates using our instrumental variables approach, which is critical for identifying the causal effect of peer misconduct. We observe that the estimated coefficients of peer effects in Table 2 are much larger to those found in the panel models from Supplementary Table 3. A possible explanation for the large difference in the GMM estimates from Table 2 and the RE and FE estimates from Supplementary Table 3 is measurement errors in the endogenous variable $Peer\ y_{i\,(t-1)}$, which will lead to attenuation bias in the RE and FE estimates[1]. Note that our endogenous variable represents the proportion of peers in $t-1$ with cases of misconduct and so measurement errors could

arise if this proportion does not always capture all peers in $t-1$, probably because peers formally registered under certain line manager are only a subset of the actual peer group.

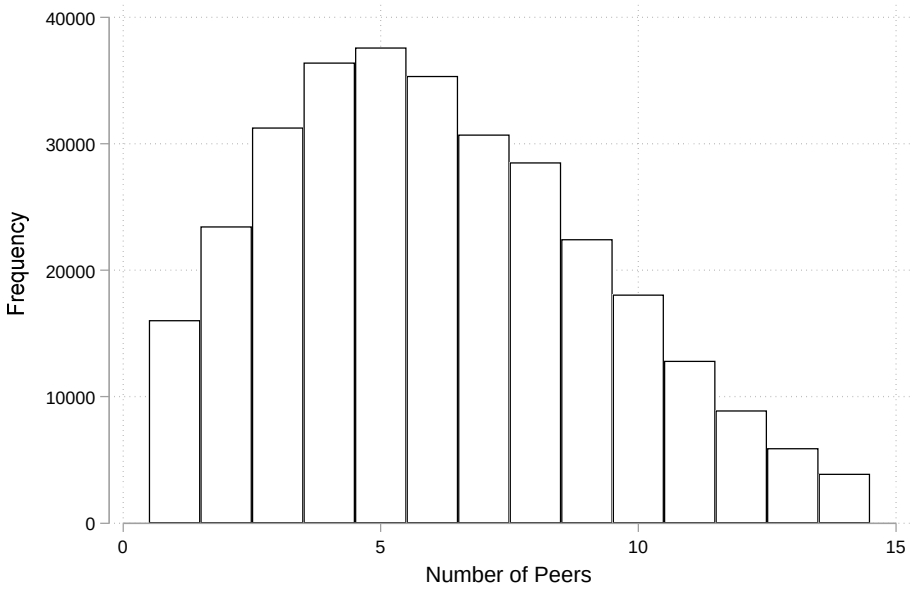
Hence, RE and FE estimates are subject to two sources of bias operating in opposite directions: the upward bias caused by both endogeneity and correlated effects and the downward bias caused by measurement errors. If the endogenous variable is measured with error, our instruments are also subject to measurement error, as they represent the proportion of peers of peers with cases of misconduct. However, to the extent that the measurement errors in our instruments are uncorrelated with the measurement errors in the endogenous variable, our GMM estimator should correct both the endogeneity bias and the attenuation bias. Also, note that in contrast to the endogenous variable that measures the proportion of peers with misconduct of a single individual, our instruments, $\overline{P1}$ and $\overline{P2}$, constitute averages across many individuals and therefore should be subject to smaller measurement errors.

Supplementary Figures

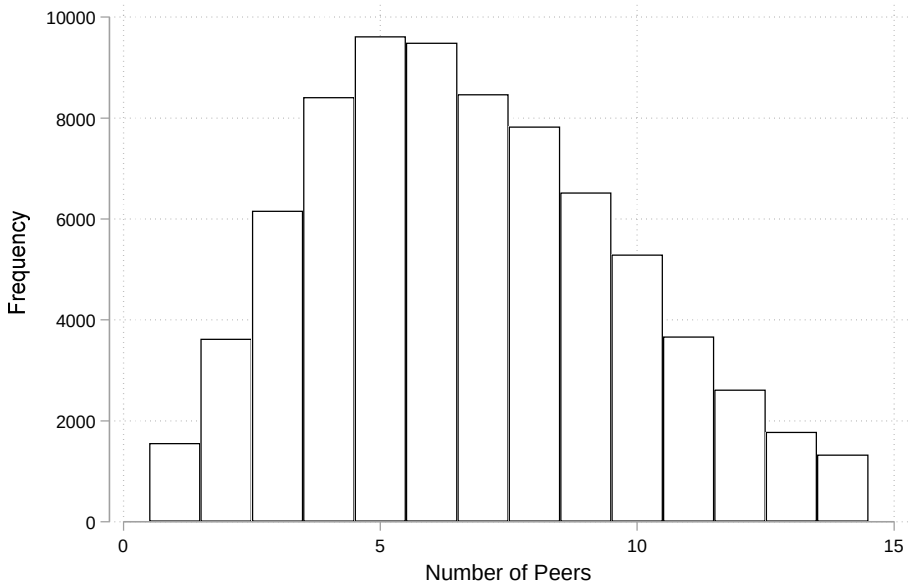


Supplementary Figure 1. The distribution of individuals according to the number and type of misconduct received over the period 2011Q1-2014Q4. Other allegations include traffic allegations. The cohort included 14,915 people. Within each panel, *N* displays the number of individuals with allegations of misconduct. The y-axis shows the percentage of *N* people with records of misconduct. The x-axis shows the number of cases of misconduct received during the period 2011Q1-2014Q4. The first bar refers to the percentage of *N* people with only one case of misconduct; whereas the last bar, with eleven cases of misconduct. Within each panel, bars add to 100%.

a. Whole sample



b. Individuals who experience new peers



Supplementary Figure 2. Distribution of number of peers by sample. The top panel includes all individual × quarter observation in the data (35,777 individuals and 311,652 observations). The target individual is excluded from this count. Thus, the group size is equivalent to the number of peers plus one. The bottom panels restrict the data to those individual × quarter observations that satisfy our criteria for identification (30,047 individuals and 76,423 observations). Outliers below the 5-percentile and above the 95-percentile are excluded.

Supplementary Tables

Supplementary Table 1. *Correlation of Allegations Within Individuals*

	Failures in duty	Malpractice	Discriminatory behaviour	Oppressive behaviour	Incivility
Malpractice	0.171*** (<0.001) [0.156 - 0.187]	1			
Discriminatory behaviour	0.263*** (<0.001) [0.248 - 0.278]	0.083*** (<0.001) [0.067 - 0.099]	1		
Oppressive behaviour	0.271*** (<0.001) [0.256 - 0.286]	0.132*** (<0.001) [0.116 - 0.148]	0.155*** (<0.001) [0.139 - 0.171]	1	
Incivility	0.271*** (<0.001) [0.256 - 0.286]	0.050*** (<0.001) [0.034 - 0.066]	0.214*** (<0.001) [0.199 - 0.230]	0.115*** (<0.001) [0.099 - 0.131]	1
Other	0.049*** (<0.001) [0.033 - 0.065]	0.020* (0.016) [0.004 - 0.036]	0.030*** (<0.001) [0.199 - 0.230]	0.020* (0.013) [0.004 - 0.036]	0.063*** (<0.001) [0.047 - 0.079]

Note. Pearson correlations of allegation types within individuals. Other allegations include traffic allegations. The sample includes 14,915 individuals. 95% CI in brackets and *p*-values in parentheses, Significance levels: *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Supplementary Table 2. *Composition of the Data Used to Estimate Peer Effects*

	Whole sample	Individuals who experience new peers
Gender		
Male	0.65	0.68
Employee type		
Police Constable	0.54	0.61
Police Sergeant	0.12	0.13
Inspector	0.03	0.03
Chief Inspector, Superintendent, Chief Superintendent	0.01	0.01
Special Constabulary	0.00	0.00
Civil Staff	0.30	0.22
Business Group		
TP - Boroughs North	0.07	0.08
TP - Boroughs South	0.10	0.12
TP - Boroughs West	0.08	0.10
TP - Central	0.00	0.00
TP - Criminal Justice & Crime	0.12	0.11
TP - Westminster	0.03	0.03
Specialist Crime and Operations	0.27	0.26
Specialist Operations	0.12	0.11
Other Business Group	0.10	0.05
Length of service (years)	13.45	12.77
Employee Performance Rating		
Exceptional + Competent (above standard)	0.49	0.49
Competent (at required standard)	0.50	0.51
Competent (development required) + Not Yet Competent	0.01	0.01
Events of misconduct		
Incidence of misconduct	0.05	0.06
Incidence of failures in duty	0.04	0.04
Incidence of malpractice	0.01	0.01
Incidence of discriminatory behavior	0.01	0.01
Incidence of oppressive behavior	0.01	0.01
Incidence of incivility	0.01	0.01
Occurrence of Formal disciplinary actions following misconduct	0.00	0.00
Occurrence of Management disciplinary actions following misconduct	0.01	0.01
Occurrence of No disciplinary actions following misconduct	0.04	0.04
Total number of individual × quarter observations	331,023	80,632

Note. The table displays the composition of the whole data (left column) and the subset used to estimate peer effects via instrumental variable regressions (right column). Cells show the proportions for each category of the total individual × quarter observations.

Supplementary Table 3. *The Estimated Likelihood of Misconduct, Peer Effects, Random and Fixed Effects Models*

	Whole sample						Individuals with incidence of misconduct					
	(1)			(2)			(3)			(4)		
VARIABLES	RE	*p*	95% CI	FE	*p*	95% CI	RE	*p*	95% CI	FE	*p*	95% CI
Prop. of peers in $t-1$ with misconduct	0.066***	<0.001	[0.057 - 0.075]	0.017***	0.001	[0.007 - 0.026]	0.063***	<0.001	[0.049 - 0.078]	0.028***	0.001	[0.012 - 0.044]
Gender (reference: Females)												
Male	0.015***	<0.001	[0.014 - 0.017]				0.014***	<0.001	[0.011 - 0.018]			
Employee type (reference: Civil Staff)												
Police Constable	0.032***	<0.001	[0.030 - 0.034]				0.011***	<0.001	[0.007 - 0.015]			
Police Sergeant	0.033***	<0.001	[0.029 - 0.036]				0.012***	<0.001	[0.006 - 0.018]			
Inspector	0.025***	<0.001	[0.020 - 0.030]				0.001	0.809	[-0.009 - 0.011]			
Chief Inspector, Superintendent, Chief Superintendent	0.010***	<0.001	[0.005 - 0.016]				-0.017*	0.025	[-0.032 - -0.002]			
Business Group (reference: TP - Boroughs East)												
TP - Boroughs North	-0.000	0.955	[-0.005 - 0.005]				-0.000	0.896	[-0.007 - 0.006]			
TP - Boroughs South	0.007**	0.003	[0.002 - 0.012]				0.008*	0.010	[0.002 - 0.014]			
TP - Boroughs West	0.001	0.708	[-0.004 - 0.006]				0.003	0.374	[-0.004 - 0.009]			
TP - Central	-0.040***	<0.001	[-0.052 - -0.027]				-0.058*	0.046	[-0.116 - -0.001]			
TP - Criminal Justice & Crime	-0.010***	<0.001	[-0.014 - -0.005]				-0.005	0.116	[-0.011 - 0.001]			
TP - Westminster	-0.001	0.862	[-0.008 - 0.007]				-0.003	0.556	[-0.012 - 0.007]			
Specialist Crime and Operations	-0.029***	<0.001	[-0.032 - -0.025]				-0.020***	<0.001	[-0.025 - -0.015]			
Specialist Operations	-0.045***	<0.001	[-0.048 - -0.041]				-0.032***	<0.001	[-0.039 - -0.026]			
Other Business Group	-0.036***	<0.001	[-0.040 - -0.032]				-0.026***	<0.001	[-0.035 - -0.017]			
Length of service												
Length of service (10 years)	-0.026***	<0.001	[-0.031 - -0.022]	-0.278***	<0.001	[-0.363 - -0.193]	-0.018***	<0.001	[-0.027 - -0.008]	-0.657***	<0.001	[-0.855 - -0.459]
Length of service (10 years)^2	0.004***	<0.001	[0.003 - 0.006]	0.009***	0.001	[0.004 - 0.015]	0.004**	0.004	[0.001 - 0.007]	0.027***	0.001	[0.011 - 0.043]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)												
Exceptional + Competent (above standard)	-0.035***	<0.001	[-0.045 - -0.025]	-0.010	0.138	[-0.023 - 0.003]	-0.035***	<0.001	[-0.053 - -0.018]	-0.019	0.146	[-0.044 - 0.007]
Competent (at required standard)	-0.028***	<0.001	[-0.039 - -0.018]	-0.010	0.150	[-0.023 - 0.003]	-0.030***	0.001	[-0.047 - -0.012]	-0.018	0.160	[-0.042 - 0.007]
Constant	0.078***	<0.001	[0.066 - 0.090]	0.451***	<0.001	[0.322 - 0.580]	0.119***	<0.001	[0.099 - 0.140]	0.944***	<0.001	[0.678 - 1.210]
Observations	331,023			331,022			141,074			141,073		
Number of individuals	35,924			35,923			14,853			14,852		
ICC	0.0235						0					
σ_u	0.0329						0					
Quarter FEs	YES			YES			YES			YES		
Year FEs	YES			YES			YES			YES		

Note. All models estimate the probability of at least one event of misconduct in quarter t conditional on a set of covariates. The variable of interest is the proportion of peers in quarter $t-1$ with reported cases of misconduct. Models 1 to 4 are linear probability panel data models that ignore the endogeneity in the peer misconduct measure. RE estimators in Models 1 and 3 incorporate random individual intercepts that are assumed to be uncorrelated with the explanatory variables. RE estimators use information from both between individual variation and within individual variation in the data. However, FE estimators in Models 2 and 4 use only within individual variation in the data. Thus, time-invariant characteristics in our data, like gender, employee type or business groups, cannot be estimated by FE models. By using only within individual variation, FE estimators allow for correlations between the individual intercepts and the explanatory variables. Alternative specifications applying instrumental variable techniques for the identification of peer effects are presented in Table 2 in the main text. 95% confidence intervals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 4. *The Estimated Likelihood of Misconduct, Peer Effects*

	Individuals experiencing new peers					
VARIABLES	(1) GMM	*p*	95% CI	(2) IV PROBIT	*p*	95% CI
Prop. of peers in $t-1$ with misconduct	0.768***	<0.001	[0.461 - 1.075]	5.426***	<0.001	[4.048 - 6.803]
Gender (reference: Females)						
Male	0.017***	<0.001	[0.013 - 0.020]	0.140***	<0.001	[0.102 - 0.178]
Employee type (reference: Civil Staff)						
Police Constable	0.017***	<0.001	[0.009 - 0.025]	0.210***	<0.001	[0.128 - 0.291]
Police Sergeant	0.022***	<0.001	[0.013 - 0.031]	0.270***	<0.001	[0.178 - 0.361]
Inspector	0.019***	0.001	[0.008 - 0.031]	0.254***	<0.001	[0.145 - 0.364]
Chief Inspector, Superintendent, Chief Superintendent	0.016*	0.011	[0.004 - 0.028]	0.143	0.117	[-0.036 - 0.323]
Business Group (reference: TP - Boroughs East)						
TP - Boroughs North	-0.001	0.852	[-0.010 - 0.008]	-0.005	0.845	[-0.057 - 0.047]
TP - Boroughs South	0.007	0.101	[-0.001 - 0.015]	0.038	0.116	[-0.009 - 0.085]
TP - Boroughs West	0.002	0.656	[-0.007 - 0.010]	0.013	0.603	[-0.037 - 0.064]
TP - Central	-0.033**	0.005	[-0.056 - -0.010]			
TP - Criminal Justice & Crime	0.006	0.164	[-0.003 - 0.015]	0.066*	0.021	[0.010 - 0.121]
TP - Westminster	0.006	0.389	[-0.007 - 0.018]	0.034	0.360	[-0.039 - 0.108]
Specialist Crime and Operations	-0.006	0.253	[-0.016 - 0.004]	-0.053	0.168	[-0.127 - 0.022]
Specialist Operations	-0.013~	0.079	[-0.027 - 0.001]	-0.163**	0.007	[-0.281 - -0.045]
Other Business Group	-0.001	0.938	[-0.015 - 0.014]	-0.204**	0.007	[-0.354 - -0.055]
Length of service						
Length of service (10 years)	-0.013*	0.034	[-0.025 - -0.001]	-0.057	0.240	[-0.152 - 0.038]
Length of service (10 years)^2	0.002	0.126	[-0.001 - 0.005]	-0.003	0.840	[-0.029 - 0.023]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)						
Exceptional + Competent (above standard)	-0.038**	0.001	[-0.062 - -0.015]	-0.285***	<0.001	[-0.428 - -0.142]
Competent (at required standard)	-0.035**	0.003	[-0.059 - -0.012]	-0.256***	<0.001	[-0.392 - -0.120]
Constant	0.038*	0.039	[0.002 - 0.074]	-1.606***	<0.001	[-1.753 - -1.459]
Observations	80,632			80,612		
Number of individuals	30,627			30,617		
LM test statistic for under identification (Kleibergen-Paap)	199.3	<0.001				
F statistic for weak identification (Kleibergen-Paap)	97.75					
Hansen Statistic	0.520	0.471				
Degrees freedom of Hansen Statistic	1					
Wald test of endogeneity, $\chi^2(1)$				29.97	<0.001	
Quarter FEs	YES			YES		
Year FEs	YES			YES		

Note. The table shows full inferential statistics for the estimates presented in Table 2. All models estimate the probability of an event of misconduct in quarter t conditional on a set of covariates. The dependent variable is a dummy equal to one when at least one event of misconduct is reported in quarter t. The independent variable of interest is the proportion of peers in quarter $t-1$ with reported cases of misconduct. Our identification strategy exploits the variation in peer groups experienced by the individuals during the period 2011-2014. We use instrumental variable techniques for the identification of peer effects. Model 1 presents a 2-step GMM linear model and Model 2, an IV PROBIT model. Two instruments are used for identification: the average proportion of peers of peers with incidence of misconduct in $t-2$ and, likewise, the average proportion of peers of peers with incidence of misconduct in $t-3$. The models include dummy controls to account for seasonal variation in the report of misconduct events: Quarter FE and year FE correspond to quarter dummies and year dummies. The first stage results of Model 1 are displayed in Supplementary Table 5. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 5. *Peer Effects on the Likelihood of Misconduct - First Stage GMM*

	Individuals experiencing new peers		
VARIABLES	(1) GMM	*p*	95% CI
Instrument 1	0.048***	<0.001	[0.040 - 0.057]
Instrument 2	0.028***	<0.001	[0.020 - 0.036]
Gender (reference: Females)			
Male	0.003***	<0.001	[0.002 - 0.005]
Employee type (reference: Civil Staff)			
Police Constable	0.022***	<0.001	[0.020 - 0.024]
Police Sergeant	0.020***	<0.001	[0.017 - 0.023]
Inspector	0.016***	<0.001	[0.011 - 0.021]
Chief Inspector, Superintendent, Chief Superintendent	-0.001	0.691	[-0.007 - 0.005]
Business Group (reference: TP - Boroughs East)			
TP - Boroughs North	<0.001	0.833	[-0.003 - 0.004]
TP - Boroughs South	0.002	0.345	[-0.002 - 0.005]
TP - Boroughs West	-0.002	0.214	[-0.006 - 0.001]
TP - Central	-0.018~	0.098	[-0.039 - 0.003]
TP - Criminal Justice & Crime	-0.013***	<0.001	[-0.017 - -0.010]
TP - Westminster	<0.001	0.977	[-0.005 - 0.006]
Specialist Crime and Operations	-0.025***	<0.001	[-0.028 - -0.022]
Specialist Operations	-0.039***	<0.001	[-0.042 - -0.036]
Other Business Group	-0.039***	<0.001	[-0.043 - -0.036]
Length of service			
Length of service (10 years)	-0.022***	<0.001	[-0.026 - -0.017]
Length of service (10 years)^2	0.004***	<0.001	[0.003 - 0.005]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)			
Exceptional + Competent (above standard)	-0.012*	0.010	[-0.022 - -0.003]
Competent (at required standard)	-0.005	0.342	[-0.014 - 0.005]
Constant	0.079***	<0.001	[0.069 - 0.089]
Observations	80,632		
Number of individuals	30,627		
LM test statistic for under identification (Kleibergen-Paap)	199.3	<0.001	
F statistic for weak identification (Kleibergen-Paap)	97.75		
Quarter FEs	YES		
Year FEs	YES		

Note. The regression displays the first stage results of Model 1 in Table 2. The dependent variable is the proportion of peers in quarter $t-1$ with reported cases of misconduct. Two instruments are used for identification: the average proportion of peers of peers with incidence of misconduct in $t-2$ and, likewise, the average proportion of peers of peers with incidence of misconduct in $t-3$. 95% confidence intervals using standard errors clustered by individuals in parentheses. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 6. *Estimated Likelihood of Misconduct, Peer Effects: Falsification Test*

	DV: Prop. of former peers in $t-2$ with cases of misconduct in t										DV: Misconduct in t		
VARIABLES	(1) GMM	*p*	95% CI	(2) GMM	*p*	95% CI	(3) GMM	*p*	95% CI	VARIABLES	(4) GMM	*p*	95% CI
Prop. of peers in $t-1$ with misconduct	0.162	0.235	[-0.105 - 0.428]	0.156	0.308	[-0.144 - 0.456]	0.132	0.393	[-0.171 - 0.436]	Prop. of peers in $t-1$ with misconduct	0.802*	0.020	[0.124 - 1.480]
Gender (reference: Prop. of Females)										Gender [reference: Females]			
Prop. of Males	0.018***	<0.001	[0.011 - 0.024]	0.018***	<0.001	[0.011 - 0.026]	0.018***	<0.001	[0.010 - 0.025]	Male	0.014**	0.001	[0.005 - 0.022]
Employee type (reference: Prop. of Civil Staff)										Employee type [reference: Civil Staff]			
Prop. of Police Constable	0.027***	<0.001	[0.019 - 0.036]	0.029***	<0.001	[0.018 - 0.039]	0.030***	<0.001	[0.019 - 0.040]	Police Constable	0.015~	0.074	[-0.001 - 0.031]
Prop. of Police Sergeant	0.027***	<0.001	[0.018 - 0.036]	0.031***	<0.001	[0.020 - 0.042]	0.034***	<0.001	[0.022 - 0.045]	Police Sergeant	0.026***	<0.001	[0.012 - 0.041]
Prop. of Inspector	0.015**	0.009	[0.004 - 0.026]	0.021**	0.003	[0.007 - 0.034]	0.025***	0.001	[0.011 - 0.038]	Inspector	0.019*	0.037	[0.001 - 0.038]
Prop. of Chief Inspector, Superintendent, Chief Superintendent	0.025**	0.004	[0.008 - 0.042]	0.022*	0.032	[0.002 - 0.042]	0.026*	0.012	[0.006 - 0.046]	Chief Inspector, Superintendent, Chief Superintendent	0.017~	0.090	[-0.003 - 0.036]
Prop. of Special Constabulary	-0.049***	<0.001	[-0.061 - -0.037]	-0.068***	<0.001	[-0.094 - -0.042]	-0.067***	<0.001	[-0.092 - -0.042]	Special Constabulary			
Business Group (reference: Prop. in TP - Boroughs East)										Business Group [reference: TP - Boroughs East]			
Prop. in TP - Boroughs North	-0.004	0.314	[-0.011 - 0.004]	-0.009*	0.048	[-0.018 - -0.000]	-0.009~	0.057	[-0.018 - 0.000]	TP - Boroughs North	-0.001	0.883	[-0.018 - 0.015]
Prop. in TP - Boroughs South	0.006~	0.099	[-0.001 - 0.013]	0.005	0.226	[-0.003 - 0.013]	0.005	0.194	[-0.003 - 0.014]	TP - Boroughs South	0.001	0.911	[-0.014 - 0.016]
Prop. in TP - Boroughs West	-0.006	0.112	[-0.013 - 0.001]	-0.007	0.160	[-0.016 - 0.003]	-0.007	0.128	[-0.016 - 0.002]	TP - Boroughs West	0.001	0.888	[-0.015 - 0.017]
Prop. in TP - Central	-0.011	0.555	[-0.048 - 0.026]	-0.026	0.336	[-0.078 - 0.027]	-0.022	0.430	[-0.075 - 0.032]	TP - Central	-0.029*	0.024	[-0.054 - -0.004]
Prop. in TP - Criminal Justice & Crime	-0.006	0.235	[-0.015 - 0.004]	-0.009~	0.085	[-0.019 - 0.001]	-0.009~	0.084	[-0.019 - 0.001]	TP - Criminal Justice & Crime	0.011	0.259	[-0.008 - 0.029]
Prop. in TP - Westminster	0.007	0.238	[-0.005 - 0.019]	0.007	0.371	[-0.008 - 0.021]	0.007	0.346	[-0.007 - 0.021]	TP - Westminster	0.020	0.117	[-0.005 - 0.045]
Prop. in Specialist Crime and Operations	-0.028***	<0.001	[-0.037 - -0.019]	-0.030***	<0.001	[-0.040 - -0.020]	-0.029***	<0.001	[-0.039 - -0.019]	Specialist Crime and Operations	-0.002	0.847	[-0.022 - 0.018]
Prop. in Specialist Operations	-0.047***	<0.001	[-0.060 - -0.034]	-0.049***	<0.001	[-0.064 - -0.034]	-0.048***	<0.001	[-0.063 - -0.033]	Specialist Operations	-0.005	0.729	[-0.034 - 0.024]
Prop. in Other Business Group	-0.035***	<0.001	[-0.047 - -0.022]	-0.037***	<0.001	[-0.051 - -0.023]	-0.035***	<0.001	[-0.049 - -0.022]	Other Business Group	0.001	0.925	[-0.028 - 0.031]
Length of service										Length of service			
Average Length of service (10 years)	-0.026***	<0.001	[-0.038 - -0.015]	-0.034***	<0.001	[-0.048 - -0.020]	-0.039***	<0.001	[-0.053 - -0.025]	Length of service [10 years]	-0.022*	0.034	[-0.042 - -0.002]
Average Length of service (10 years)^2	0.003~	0.065	[-0.000 - 0.007]	0.006**	0.009	[0.001 - 0.010]	0.007**	0.003	[0.002 - 0.011]	Length of service [10 years]^2	0.004	0.115	[-0.001 - 0.010]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)										Employee Performance Rating in $t-4$ [reference: Competent but development required + Not Yet Competent]			
Prop. of Exceptional + Competent (above standard)				-0.079***	0.001	[-0.125 - -0.034]	-0.080***	<0.001	[-0.125 - -0.035]	Exceptional + Competent [above standard]	-0.048*	0.025	[-0.089 - -0.006]
Prop. of Competent (at required standard)				-0.072**	0.002	[-0.117 - -0.026]	-0.071**	0.002	[-0.116 - -0.025]	Competent [at required standard]	-0.041~	0.050	[-0.083 - 0.000]
Constant	0.057***	<0.001	[0.037 - 0.076]	0.136***	<0.001	[0.088 - 0.184]	0.146***	<0.001	[0.098 - 0.194]	Constant	0.054	0.104	[-0.011 - 0.118]
Observations	27,040			19,796			19,796			Observations	20,374		
Number of individuals	18,506			14,111			14,111			Number of individuals	14,401		
LM test statistic for underidentification (Kleibergen-Paap)	52.51	<0.001		40.70	<0.001		39.53	<0.001		LM test statistic for under identification [Kleibergen-Paap]	35.16	<0.001	
F statistic for weak identification (Kleibergen-Paap)	25.39			19.62			19.09			F statistic for weak identification [Kleibergen-Paap]	16.98		
Hansen Statistic	3.135	0.077		3.486	0.062		3.889	0.049		Hansen Statistic	0.0304	0.862	
Degrees freedom of Hansen Statistic	1			1			1			Degrees freedom of Hansen Statistic	1		
Quarter FEs	NO			NO			YES			Quarter FEs	YES		
Year FEs	NO			NO			YES			Year FEs	YES		

Note. The table shows full inferential statistics for the results presented in Table 3. All models apply instrumental variable techniques for the identification of peer effects. Models 1 to 3 are part of a falsification test that study the behaviour of former peers of an individual i who moves to a different peer group in $t-1$. The outcome constitutes the proportion of these peers who had reports of misconduct at time t. The independent variable of interest is the proportion of peers of i in $t-1$ presenting incidence of misconduct. This variable is instrumented by two measures of conduct of peers of peers of i. By construction, these two measures are expected to have no influence on the outcome variable of these models. Model 4 is presented for comparative purposes and uses the standard outcome variable of the study, misconduct of i at time t. 95% confidence intervals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 7. *Peer Effects on the Likelihood of Misconduct - Line Manager Effects - Exhaustive Geographic Controls*

	Individuals experiencing new peers											
VARIABLES	(1) GMM	*p*	95% CI	(2) IV PROBIT	*p*	95% CI	(3) GMM	*p*	95% CI	(4) IV PROBIT	*p*	95% CI
Prop. of peers in $t-1$ with misconduct	0.731***	<0.001	[0.397 - 1.066]	5.276***	<0.001	[3.708 - 6.844]	0.687***	<0.001	[0.381 - 0.994]	4.985***	<0.001	[3.456 - 6.514]
Gender (reference: Females)												
Male	0.017***	<0.001	[0.013 - 0.020]	0.143***	<0.001	[0.104 - 0.182]	0.016***	<0.001	[0.013 - 0.020]	0.140***	<0.001	[0.103 - 0.177]
Employee type (reference: Civil Staff)												
Police Constable	0.018***	<0.001	[0.010 - 0.026]	0.215***	<0.001	[0.130 - 0.299]	0.019***	<0.001	[0.012 - 0.026]	0.211***	<0.001	[0.134 - 0.289]
Police Sergeant	0.023***	<0.001	[0.014 - 0.032]	0.272***	<0.001	[0.181 - 0.364]	0.021***	<0.001	[0.013 - 0.029]	0.248***	<0.001	[0.165 - 0.331]
Inspector	0.020***	<0.001	[0.009 - 0.031]	0.258***	<0.001	[0.148 - 0.368]	0.017**	0.005	[0.005 - 0.029]	0.217***	<0.001	[0.103 - 0.331]
Chief Inspector, Superintendent, Chief Superintendent	0.016*	0.010	[0.004 - 0.028]	0.141	0.125	[-0.039 - 0.322]	0.011	0.142	[-0.004 - 0.025]	0.087	0.457	[-0.142 - 0.316]
Special Constabulary							-0.012	0.174	[-0.029 - 0.005]	-0.270**	0.006	[-0.464 - -0.076]
Length of service												
Length of service (10 years)	-0.014*	0.026	[-0.026 - -0.002]	-0.062	0.205	[-0.159 - 0.034]	-0.015**	0.006	[-0.026 - -0.004]	-0.069	0.115	[-0.156 - 0.017]
Length of service (10 years)^2	0.002	0.107	[-0.001 - 0.005]	-0.002	0.909	[-0.028 - 0.025]	0.002~	0.085	[-0.000 - 0.005]	-0.004	0.764	[-0.028 - 0.021]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)												
Exceptional + Competent (above standard)	-0.038**	0.001	[-0.062 - -0.015]	-0.288***	<0.001	[-0.433 - -0.143]						
Competent (at required standard)	-0.035**	0.003	[-0.058 - -0.012]	-0.259***	<0.001	[-0.396 - -0.121]						
Line Manager Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)												
Exceptional + Competent (above standard)							-0.039~	0.080	[-0.083 - 0.005]	-0.325*	0.027	[-0.612 - -0.037]
Competent (at required standard)							-0.038~	0.087	[-0.082 - 0.006]	-0.318*	0.030	[-0.605 - -0.031]
Constant	0.033~	0.088	[-0.005 - 0.072]	-1.639***	<0.001	[-1.833 - -1.446]	0.036	0.182	[-0.017 - 0.089]	-1.612***	<0.001	[-1.926 - -1.297]
Observations	80,632			80,612			77,994			77,994		
Number of individuals	30,627			30,617			32,183			32,183		
LM test statistic for under identification (Kleibergen-Paap)	165.4	<0.001					183.1	<0.001				
F statistic for weak identification (Kleibergen-Paap)	80.98						90.24					
Hansen Statistic	0.653	0.419					0.492	0.483				
Degrees freedom of Hansen Statistic	1						1					
Wald test of endogeneity, $\chi^2(1)$				22.70	<0.001					22.29	<0.001	
Geographic FEs	YES			YES			YES			YES		
Quarter FEs	YES			YES			YES			YES		
Year FEs	YES			YES			YES			YES		

Note. The models replicate the regression from Table 2 but including a large number of dummy controls for specific geographic fixed effects (FEs): 32 dummy variables distinguishing 32 Borough Operational Command Units. Geographic FEs also include dummy variables for specific subgroups of Territorial Policing (TP): TP - Central, TP - Westminster, and 6 subgroups that are part of the TP - Criminal Justice & Crime (Met Detention, Met Prosecutions, RTPC - Roads and Transport Policing Com, TP - Capability and Support, TP Crime Recording Investigation Bureau, and TP Crime Recording Investigation Bureau). As in earlier specifications (Tables 2 and 3), our models also incorporate dummy variables for individuals working in the Directorates Specialist Crime and Operations and Specialist Operations. Observe that Models 1 and 2 incorporate controls for the individual performance in quarter $t-4$ and Models 3 and 4, for the performance of the line manager. These performance scores are reported on annual basis in Performance Development Reviews and evaluate competences in operational effectiveness, organizational influence, and resource management. Because the individual performance and his/her line manager performance are mutually determined, these controls are reported in different models. Sample sizes across models differ because of missing performance scores for a small subset of individuals or their line managers. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 8. *Peer Effects on the Likelihood of Misconduct - Line Manager Effects - Exhaustive Geographic Controls - First Stage GMM*

	Individuals experiencing new peers					
	(1)			(2)		
VARIABLES	GMM	*p*	95% CI	GMM	*p*	95% CI
Instrument 1	0.045***	<0.001	[0.037 - 0.054]	0.051***	<0.001	[0.042 - 0.060]
Instrument 2	0.024***	<0.001	[0.016 - 0.032]	0.027***	<0.001	[0.018 - 0.035]
Gender (reference: Females)						
Male	0.003**	0.003	[0.001 - 0.004]	0.002*	0.025	[0.000 - 0.004]
Employee type (reference: Civil Staff)						
Police Constable	0.020***	<0.001	[0.018 - 0.022]	0.019***	<0.001	[0.017 - 0.021]
Police Sergeant	0.016***	<0.001	[0.013 - 0.019]	0.013***	<0.001	[0.010 - 0.017]
Inspector	0.012***	<0.001	[0.007 - 0.017]	0.011***	<0.001	[0.006 - 0.017]
Chief Inspector, Superintendent, Chief Superintendent	-0.004	0.198	[-0.009 - 0.002]	-0.004	0.236	[-0.012 - 0.003]
Special Constabulary				-0.048***	<0.001	[-0.051 - -0.045]
Length of service						
Length of service (10 years)	-0.020***	<0.001	[-0.024 - -0.016]	-0.020***	<0.001	[-0.024 - -0.016]
Length of service (10 years)^2	0.003***	<0.001	[0.002 - 0.005]	0.004***	<0.001	[0.003 - 0.005]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)						
Exceptional + Competent (above standard)	-0.011*	0.017	[-0.021 - -0.002]			
Competent (at required standard)	-0.004	0.374	[-0.014 - 0.005]			
Line Manager Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)						
Exceptional + Competent (above standard)				-0.006	0.468	[-0.021 - 0.010]
Competent (at required standard)				-0.001	0.891	[-0.017 - 0.014]
Constant	0.068***	<0.001	[0.056 - 0.081]	0.071***	<0.001	[0.053 - 0.089]
Observations	80,632			77,994		
Number of individuals	30,627			32,183		
LM test statistic for under identification (Kleibergen-Paap)	165.4	<0.001		183.1	<0.001	
F statistic for weak identification (Kleibergen-Paap)	80.98			90.24		
Geographic FEs	YES			YES		
Quarter FEs	YES			YES		
Year FEs	YES			YES		

Note. The model reports the first stage results of the GMM regressions presented in Models 1 and 2 of Supplementary Table 7. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, ~ $p<0.10$.

Supplementary Table 9. *Peer Effects on the Likelihood of Misconduct - Difference Between Individuals Who Move to a New Peer Group and Individuals Who Have New Incoming Peers to Their Current Peer Group - Exhaustive Geographic Controls*

	Individuals who experience new peers											
	(A) Individuals moving to a different peer group						(B) Individuals with new incoming peers to their current peer group					
	(1)			(2)			(3)			(4)		
VARIABLES	GMM	*p*	95% CI	IV PROBIT	*p*	95% CI	GMM	*p*	95% CI	IV PROBIT	*p*	95% CI
Prop. of peers in $t-1$ with misconduct	0.730*	0.049	[0.005 - 1.456]	4.920**	0.001	[1.913 - 7.928]	0.739***	<0.001	[0.364 - 1.115]	5.493***	<0.001	[3.665 - 7.321]
Gender (reference: Females)												
Male	0.014***	0.001	[0.006 - 0.022]	0.114**	0.006	[0.033 - 0.194]	0.018***	<0.001	[0.013 - 0.022]	0.153***	<0.001	[0.109 - 0.196]
Employee type (reference: Civil Staff)												
Police Constable	0.017*	0.035	[0.001 - 0.033]	0.197*	0.021	[0.030 - 0.364]	0.019***	<0.001	[0.009 - 0.028]	0.219***	<0.001	[0.121 - 0.316]
Police Sergeant	0.027***	<0.001	[0.013 - 0.041]	0.280***	0.001	[0.120 - 0.439]	0.022***	<0.001	[0.011 - 0.033]	0.270***	<0.001	[0.159 - 0.380]
Inspector	0.019*	0.033	[0.002 - 0.037]	0.208*	0.029	[0.021 - 0.395]	0.021**	0.003	[0.007 - 0.035]	0.274***	<0.001	[0.140 - 0.408]
Chief Inspector, Superintendent, Chief Superintendent	0.015	0.156	[-0.006 - 0.035]	0.056	0.719	[-0.247 - 0.359]	0.017*	0.024	[0.002 - 0.032]	0.186~	0.077	[-0.020 - 0.393]
Length of service												
Length of service (10 years)	-0.021*	0.042	[-0.041 - -0.001]	-0.120	0.147	[-0.282 - 0.042]	-0.011	0.147	[-0.025 - 0.004]	-0.036	0.536	[-0.151 - 0.078]
Length of service (10 years)^2	0.004	0.155	[-0.001 - 0.009]	0.012	0.613	[-0.034 - 0.058]	0.002	0.312	[-0.002 - 0.005]	-0.008	0.617	[-0.038 - 0.023]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)												
Exceptional + Competent (above standard)	-0.049*	0.020	[-0.090 - -0.008]	-0.320**	0.010	[-0.564 - -0.077]	-0.034*	0.023	[-0.063 - -0.005]	-0.273**	0.004	[-0.458 - -0.088]
Competent (at required standard)	-0.042*	0.044	[-0.083 - -0.001]	-0.263*	0.023	[-0.490 - -0.036]	-0.032*	0.029	[-0.060 - -0.003]	-0.253**	0.005	[-0.430 - -0.077]
Constant	0.032	0.382	[-0.040 - 0.105]	-1.611***	<0.001	[-2.011 - -1.212]	0.031	0.185	[-0.015 - 0.077]	-1.663***	<0.001	[-1.897 - -1.429]
Observations	20,374			20,364			60,258			60,248		
Number of individuals	14,401			14,395			27,895			27,887		
LM test statistic for under identification (Kleibergen-Paap)	29.68	<0.001					139.2	<0.001				
F statistic for weak identification (Kleibergen-Paap)	14.33						68.51					
Hansen Statistic	0.0336	0.855					1.116	0.291				
Degrees freedom of Hansen Statistic	1						1					
Wald test of endogeneity, $\chi^2(1)$				5.401	0.020					17.76	<0.001	
Geographic FEs	YES			YES			YES			YES		
Quarter FEs	YES			YES			YES			YES		
Year FEs	YES			YES			YES			YES		

Note. All models estimate the probability of an event of misconduct in quarter t conditional on a set of covariates. The dependent variable is a dummy equal to one when at least one event of misconduct is reported in quarter t. The independent variable of interest is the proportion of peers in quarter $t-1$ with reported cases of misconduct. Our identification strategy exploits the variation in peer groups experienced by the individuals during the period 2011-2014. Models 1 and 2 correspond to the sample of individuals who joined a new peer group. Models 3 and 4 correspond to the sample of individuals who had incoming new member to their current peer group. We use instrumental variable techniques for the identification of peer effects. Models 1 and 3 present 2-step GMM linear models and Models 2 and 4 present IV PROBIT models. Two instruments are used for identification: the average proportion of peers of peers with incidence of misconduct in $t-2$ and, likewise, the average proportion of peers of peers with incidence of misconduct in $t-3$. The first stage results of Models 1 and 3 are displayed in Supplementary Table 10. The models include a large number of dummy controls for specific fixed effects (FE). Quarter FE and year FE correspond to quarter dummies and year dummies that account for seasonal variation in the report of misconduct events. Geographic FEs include 32 dummies distinguishing 32 Borough Operational Command Units. Geographic FEs also include dummy variables for specific subgroups of Territorial Policing (TP): TP - Central, TP - Westminster, and 6 subgroups that are part of the TP - Criminal Justice & Crime (Met Detention, Met Prosecutions, RTPC - Roads and Transport Policing Com, TP - Capability and Support, TP Crime Recording Investigation Bureau, and TP Crime Recording Investigation Bureau). As in earlier specifications, we also incorporate dummies for individuals working in the Directorates Specialist Crime and Operations and Specialist Operations. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 10. *Peer Effects on the Likelihood of Misconduct - Difference Between Individuals Who Move to a Different Peer Group and Individuals Who Have New Incoming Peers to Their Current Peer Group - Exhaustive Geographic Controls - First Stage*

	(A) Individuals moving to a different peer group			(B) Individuals with new incoming peers to their current peer group		
VARIABLES	(1) GMM	*p*	95% CI	(2) IV PROBIT	*p*	95% CI
Instrument 1	0.043***	<0.001	[0.023 - 0.062]	0.046***	<0.001	[0.037 - 0.055]
Instrument 2	0.024*	0.015	[0.005 - 0.044]	0.023***	<0.001	[0.015 - 0.032]
Gender (reference: Females)						
Male	0.005**	0.004	[0.002 - 0.009]	0.001	0.156	[-0.001 - 0.003]
Employee type (reference: Civil Staff)						
Police Constable	0.017***	<0.001	[0.012 - 0.022]	0.021***	<0.001	[0.019 - 0.023]
Police Sergeant	0.006~	0.082	[-0.001 - 0.012]	0.020***	<0.001	[0.017 - 0.024]
Inspector	0.001	0.831	[-0.009 - 0.011]	0.016***	<0.001	[0.010 - 0.021]
Chief Inspector, Superintendent, Chief Superintendent	-0.015**	0.005	[-0.025 - -0.004]	0.001	0.853	[-0.006 - 0.007]
Length of service						
Length of service (10 years)	-0.005	0.256	[-0.015 - 0.004]	-0.025***	<0.001	[-0.029 - -0.020]
Length of service (10 years)^2	-0.000	0.992	[-0.003 - 0.003]	0.005***	<0.001	[0.003 - 0.006]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)						
Exceptional + Competent (above standard)	-0.005	0.564	[-0.022 - 0.012]	-0.015**	0.010	[-0.026 - -0.003]
Competent (at required standard)	-0.001	0.875	[-0.018 - 0.016]	-0.006	0.265	[-0.017 - 0.005]
Constant	0.062***	<0.001	[0.035 - 0.089]	0.072***	<0.001	[0.057 - 0.087]
Observations	20,374			60,258		
Number of individuals	14,401			27,895		
LM test statistic for under identification (Kleibergen-Paap)	29.68	<0.001		139.2	<0.001	
F statistic for weak identification (Kleibergen-Paap)	14.33			68.51		
Geographic FEs	YES			YES		
Quarter FEs	YES			YES		
Year FEs	YES			YES		

Note. The model reports the first stage results of the GMM regressions presented in Models 1 and 3 of Supplementary Table 9. Two instruments are used for identification: the average proportion of peers of peers with incidence of misconduct in $t-2$ and, likewise, the average proportion of peers of peers with incidence of misconduct in $t-3$. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 11. *Peer Effects on the Likelihood of Misconduct - Peer Group Size Effects - Exhaustive Geographic Controls*

VARIABLES	(1) GMM	*p*	95% CI	(2) IV PROBIT	*p*	95% CI
Prop. of peers in $t-1$ with misconduct	0.585**	0.008	[0.153 - 1.016]	4.606***	0.001	[2.005 - 7.207]
Prop. of peers in $t-1$ with misconduct # Number of peers in $t-1$	0.027	0.381	[-0.034 - 0.089]	0.128	0.540	[-0.282 - 0.539]
Number of peers in $t-1$	-0.001	0.440	[-0.005 - 0.002]	-0.006	0.625	[-0.030 - 0.018]
Gender (reference: Females)						
Male	0.017***	<0.001	[0.013 - 0.021]	0.141***	<0.001	[0.101 - 0.180]
Employee type (reference: Civil Staff)						
Police Constable						
Police Sergeant	0.019***	<0.001	[0.010 - 0.027]	0.216***	<0.001	[0.127 - 0.305]
Inspector	0.024***	<0.001	[0.015 - 0.034]	0.281***	<0.001	[0.183 - 0.378]
Chief Inspector, Superintendent, Chief Superintendent	0.022***	<0.001	[0.011 - 0.034]	0.273***	<0.001	[0.159 - 0.387]
Chief Inspector, Superintendent, Chief Superintendent	0.015*	0.023	[0.002 - 0.027]	0.139	0.142	[-0.046 - 0.324]
Length of service						
Length of service (10 years)	-0.014*	0.045	[-0.027 - -0.000]	-0.065	0.228	[-0.170 - 0.040]
Length of service (10 years)^2	0.003	0.122	[-0.001 - 0.006]	-<0.001	0.984	[-0.028 - 0.027]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)						
Exceptional + Competent (above standard)	-0.035**	0.004	[-0.059 - -0.011]	-0.262***	0.001	[-0.411 - -0.113]
Competent (at required standard)	-0.032**	0.008	[-0.055 - -0.008]	-0.232**	0.001	[-0.373 - -0.091]
Constant	0.039~	0.058	[-0.001 - 0.080]	-1.605***	<0.001	[-1.856 - -1.353]
Observations	77,427			77,408		
Number of individuals	30,163			30,153		
LM test statistic for under identification (Kleibergen-Paap)	157.9	<0.001				
F statistic for weak identification (Kleibergen-Paap)	38.58					
Hansen Statistic	1.005	0.605				
Degrees freedom of Hansen Statistic	2					
Wald test of endogeneity, $\chi^2(1)$				22.99	<0.001	
Geographic FEs	YES			YES		
Quarter FEs	YES			YES		
Year FEs	YES			YES		

Note. All models estimate the probability of an event of misconduct in quarter t conditional on a set of covariates. The dependent variable is a dummy equal to one when at least one event of misconduct is reported in quarter t. The independent variable of interest is the proportion of peers in quarter $t-1$ with reported cases of misconduct. Our identification strategy exploits the variation in peer groups experienced by the individuals during the period 2011-2014. The models are similar to those in Supplementary Table 7, Models 1 and 2. But we added an interaction between the endogenous variable (Prop. of peers in $t-1$ with misconduct) and the number of peers. We excluded observations above the 95-percentile of number of peers (more than 15 peers). We use instrumental variable techniques for the identification of peer effects. Model 1 presents 2-step GMM linear model and Model 2 presents an IV PROBIT model. Four instruments are used for identification: the average proportion of peers of peers with incidence of misconduct in $t-2$, the average proportion of peers of peers with incidence of misconduct in $t-3$, and the interaction of each of these variables with the number of peers in quarter $t-1$. The first stage results of Model 1 are displayed in Supplementary Table 12. The models include a large number of dummy controls for specific fixed effects (FE). Quarter FE and year FE correspond to quarter dummies and year dummies that account for seasonal variation in the report of misconduct events. Geographic FEs include 32 dummies distinguishing 32 Borough Operational Command Units. Geographic FEs also include dummy variables for specific subgroups of Territorial Policing (TP): TP - Central, TP - Westminster, and 6 subgroups that are part of the TP - Criminal Justice & Crime (Met Detention, Met Prosecutions, RTPC - Roads and Transport Policing Com, TP - Capability and Support, TP Crime Recording Investigation Bureau, and TP Crime Recording Investigation Bureau). As in earlier specifications, we also incorporate dummies for individuals working in the Directorates Specialist Crime and Operations and Specialist Operations. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, ~ $p<0.10$.

Supplementary Table 12. *Peer Effects on the Likelihood of Misconduct - Peer Group Size Effects - Exhaustive Geographic Controls - First Stage*

	DV= Prop. of peers in $t-1$ with misconduct			DV=Prop. of peers in $t-1$ with misconduct # Number of peers in $t-1$		
VARIABLES	(1) GMM	*p*	95% CI	(2) GMM	*p*	95% CI
Instrument 1	0.060***	<0.001	[0.037 - 0.083]	-0.031	0.553	[-0.132 - 0.071]
Instrument 2	0.040***	0.001	[0.017 - 0.063]	-0.233***	<0.001	[-0.332 - -0.134]
Instrument 1 # Number of peers in $t-1$	-0.002~	0.094	[-0.005 - 0.000]	0.043***	<0.001	[0.025 - 0.061]
Instrument 2 # Number of peers in $t-1$	-0.003~	0.075	[-0.005 - 0.000]	0.056***	<0.001	[0.038 - 0.074]
Number of peers in $t-1$	0.001***	<0.001	[0.001 - 0.001]	0.053***	<0.001	[0.051 - 0.055]
Gender (reference: Females)						
Male	0.003**	0.005	[0.001 - 0.004]	0.020**	0.001	[0.008 - 0.031]
Employee type (reference: Civil Staff)						
Police Constable	0.020***	<0.001	[0.018 - 0.022]	0.128***	<0.001	[0.114 - 0.143]
Police Sergeant	0.017***	<0.001	[0.014 - 0.020]	0.105***	<0.001	[0.085 - 0.124]
Inspector	0.013***	<0.001	[0.008 - 0.018]	0.069***	<0.001	[0.042 - 0.095]
Chief Inspector, Superintendent, Chief Superintendent	-0.002	0.475	[-0.008 - 0.004]	0.034*	0.018	[0.006 - 0.063]
Length of service						
Length of service (10 years)	-0.020***	<0.001	[-0.024 - -0.015]	-0.178***	<0.001	[-0.208 - -0.148]
Length of service (10 years)^2	0.003***	<0.001	[0.002 - 0.005]	0.034***	<0.001	[0.026 - 0.042]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)						
Exceptional + Competent (above standard)	-0.013**	0.010	[-0.022 - -0.003]	-0.065*	0.025	[-0.122 - -0.008]
Competent (at required standard)	-0.006	0.222	[-0.016 - 0.004]	-0.028	0.334	[-0.085 - 0.029]
Constant	0.062***	<0.001	[0.049 - 0.076]	0.087*	0.040	[0.004 - 0.170]
Observations	77,427			77,427		
Number of people	30,163			30,163		
LM test statistic for under identification (Kleibergen-Paap)	157.9	<0.001		157.9	<0.001	
F statistic for weak identification (Kleibergen-Paap)	38.58			38.58		
Geographic FEs	YES			YES		
Quarter FEs	YES			YES		
Year FEs	YES			YES		

Note. The model reports the first stage results of the GMM regressions presented in Model 1 of Supplementary Table 11. Four instruments are used for identification: the average proportion of peers of peers with incidence of misconduct in $t-2$, the average proportion of peers of peers with incidence of misconduct in $t-3$, and the interaction of each of these variables with the number of peers in quarter $t-1$. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary Table 13. *Likelihood of Switching Line Manager - Whole Data - Exhaustive Geographic Controls*

VARIABLES	(1) RE	*p*	95% CI	(2) RE	*p*	95% CI	(3) RE	*p*	95% CI
Complaint in previous semester =1	0.046***	<0.001	[0.038 - 0.053]						
Gender (reference: Females)									
Male							-0.018***	<0.001	[-0.023 - -0.013]
Employee type (reference: Civil Staff)									
Police Constable							0.086***	<0.001	[0.080 - 0.092]
Police Sergeant							0.121***	<0.001	[0.112 - 0.131]
Inspector							0.132***	<0.001	[0.118 - 0.146]
Chief Inspector, Superintendent, Chief Superintendent							0.077***	<0.001	[0.056 - 0.099]
Length of service									
Length of service (10 years)							-0.065***	<0.001	[-0.078 - -0.053]
Length of service (10 years)^2							0.009***	<0.001	[0.005 - 0.012]
Employee Performance Rating in $t-4$ (reference: Competent but development required + Not Yet Competent)									
Exceptional + Competent (above standard)				-0.036**	0.004	[-0.062 - -0.011]			
Competent (at required standard)				-0.022~	0.079	[-0.047 - 0.003]			
Constant	0.357***	<0.001	[0.355 - 0.360]	0.391***	<0.001	[0.366 - 0.415]	0.480***	<0.001	[0.454 - 0.505]
Observations	195,363			195,363			195,363		
Number of individuals	35,923			35,923			35,923		
ICC	0.0615			0.0667			0.0372		
σ_u	0.118			0.124			0.0906		
Year FEs	NO			NO			YES		
Geographic FEs	NO			NO			YES		

Note. The model reports Random Effects (RE) regression estimates. The dependent variable is a dummy equal to one when an individual is assigned to a different line manager during semester t. Data observations are semi-annual since line managers changes are registered every six months. Model 1 shows the effects of a misconduct event during the preceding semester on the likelihood of a subsequent change in line manager. Model 2 shows the effects of past performance scores on the likelihood of a subsequent change in line manager. These performance scores are reported on annual basis in Performance Development Reviews and evaluate competences in operational effectiveness, organizational influence, and resource management. Because the individual performance and his/her incidents of misconducts are correlated, these controls are reported in different models. Model 3 shows the effects of demographic controls. 95% confidence intervals using standard errors clustered by individuals in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10.

Supplementary References

1 Cameron, A. C. & Trivedi, P. K. *Microeconometrics : methods and applications*. (2005).