‘Who You Gonna Call?’ 911 Call Takers and Police Discretion

Austin V. Smith*

August 14, 2025

Abstract

Police officers make high-stakes decisions under uncertainty, often acting on information from 911 call takers. Using quasi-random assignment of calls to call-takers in Dallas, I show that variation in call takers’ risk assessments significantly affects arrests. Calls upgraded to priority increase arrests by 53% relative to the mean, primarily for low-level offenses. These effects are partly driven by officers using call information to inform evaluations of suspect culpability, with stronger responses among inexperienced officers or in racially discordant neighborhoods. The results highlight the critical role of call takers and officer information processing in shaping enforcement.

JEL Codes: D91, J15, K42

* *Northeastern University*, email: austinvsmith.econ@gmail.com. I am deeply grateful to my committee members, Dan Herbst, Evan Taylor, Ashley Langer, and Juan Pantano, for their invaluable guidance, support, and mentorship. I would also like to thank Jesse Bruhn, Spencer Cooper, Lance Gui, Ami Ichikawa, Tom Kirchmaier, Jonathan Moreno-Medina, Matthew B. Ross, CarlyWill Sloan and seminar participants at the University of Arizona, Texas Economics of Crime Workshop, and WEC Jr. for their helpful feedback. Additionally, I thank Jeremy Price and numerous members of the Dallas PD Public Information Office for helping me with institutional details. All remaining errors are my own.

1 Introduction

Police officers make decisions under uncertainty, time pressure, and the potential for danger to themselves or others. These decisions, such as whether to make an arrest, are high-stakes, as they impose substantial economic and social costs on arrested individuals and their communities. While individual officer discretion is a determinant of law enforcement outcomes (Weisburst, 2024), police often act on information they have not gathered themselves. In the US, half of all police contacts with the public are initiated by civilians, typically through 911 (Tapp and Davis, 2024). Before an officer is dispatched to a 911 call, a call taker speaks with the civilian caller and determines whether the call should be prioritized for police response. Priority reflects the severity of the incident and the risk to civilians. Call takers have substantial discretion in making these classifications, yet we know little about the role their discretion plays in high-stakes police decisions.

In this paper, I provide the first empirical evidence that variation in call takers' risk assessment preferences meaningfully affects police enforcement decisions. Specifically, I estimate the effect of priority classifications on arrest outcomes. Because priority status is more likely to be given to incidents with greater unobserved potential for arrest, I employ an examiner design identification strategy that exploits the quasi-random assignment of 911 calls to call takers in Dallas, Texas (Doyle, 2007; Dobbie et al., 2018; Arnold et al., 2018). In my setting, calls are assigned to call takers through an automated system that routes calls to the next available taker according to a predetermined order. I instrument for priority classification using *Call Taker Score*, a measure of each call taker's intrinsic propensity to assign priority.

I first document substantial variation in call takers' preferences. Replacing a call taker in the 10th percentile of the Call Taker Score distribution with one in the 90th percentile increases the likelihood of a call receiving priority status by 18pp—a 38% change relative to the average call priority rate. Call takers who upgrade calls more frequently tend to be relatively inexperienced; however, a concentrated right tail of experienced call takers consistently upgrades calls at higher rates than most peers.

The instrument allows me to estimate differences in officer behavior driven solely by variation in priority status arising from call takers' preferences. Using a sample composed primarily of 911 calls for low-level offenses, I find that priority signals lead officers to make more arrests at calls that are otherwise observationally similar to non-priority incidents. A priority designation results in 1.5 additional arrests per 100 calls, a 53% increase relative to the mean arrest rate. These results are stable across several alternative specifications addressing threats to the IV exclusion restriction, including endogenous sample selection and multiple dimensions of call taker decisions. The effect is driven by non-index arrests, suggesting that call takers primarily influence enforcement for low-level crimes.

I then show that these effects are driven, at least in part, by officers using priority signals to inform their arrest decisions. I provide two pieces of evidence supporting this information-processing mechanism. First, misdemeanor arrests are 16.2pp less likely to result in conviction when made at a marginally upgraded priority call, suggesting that officers incorporate these signals into assessments of suspect culpability and are induced to make lower-quality arrests. Second, officers rely more on call taker information in situations where their prior beliefs over suspect guilt are likely weaker. Specifically, I examine two important response contexts where this is likely to occur: when officers interact with same-race suspects and as they accumulate on-the-job experience. I implement a difference-in-differences design exploiting within-officer variation in the timing and location of dispatches to estimate how the arrest gap between priority and non-priority calls changes as officers gain experience and work in demographically similar neighborhoods.

I find that arrests respond *less* to priority signals when officers are dispatched to same-race neighborhoods. A 10pp increase in the share of neighborhood residents who share an officer's race reduces the priority effect by 0.0014pp—a 5% reduction relative to the mean. This attenuation is driven by Black and White officers, whose priority arrest gaps shrink in Black (White) neighborhoods relative to their non-Black (non-White) counterparts. I also show that the effect of priority on arrests declines as officers gain experience. By 25 years of service, the effect of marginal priority classification on arrest decisions becomes negligible. These findings suggest that officers are more likely to incorporate noisy external signals when their own priors are weaker.

Taken together, the results highlight the role of call takers in shaping police responses and the importance of officer information processing in enforcement decisions. They suggest two broad avenues for policy reform. First, given the large effects of priority signals and the substantial variation in call takers' propensity to assign them, these findings point to a need for standardized training protocols that reduce discretion in priority classification. Second, the fact that officers respond differently to the same information depending on their prior knowledge suggests benefits to recruiting officers more familiar with the communities they police. In particular, given both the racial heterogeneity in responses to priority signals and existing evidence that minority civilians are more likely to face adverse policing outcomes disproportionate to their offending behavior (e.g. Goncalves and Mello, 2021; Hoekstra and Sloan, 2022), these findings support the view that increasing minority officer representation may reduce disparities by increasing officers' familiarity with the civilians they encounter.

This paper contributes to three literatures. First, it adds to research on the determinants of police discretion. A broad literature identifies factors shaping enforcement decisions—including union protections, peer influence, training, and officer preferences (e.g. Mas, 2006; Cunningham et al., 2021; Holz et al., 2023; Weisburst, 2024; Rivera, 2025). I build on this work by showing that external information generated by call takers substantially impacts arrests. My findings suggest there may be large, unrealized social returns to policy changes targeting this under-analyzed but impactful segment of law enforcement. In particular, I complement Dube et al. (2025), who shows that officers may rely on readily available but noisy information when making arrests or using force. By providing direct evidence that police decisions are shaped by call taker priority signals—an upstream, institutionally generated cue—I highlight a channel through which behaviorally informed interventions may influence enforcement.

Second, my findings contribute to research on the effects of workforce diversity on social outcomes. Prior work shows that minority practitioners in education, medicine, and law enforcement generate better outcomes for minority individuals than their White counterparts (Ba et al., 2021b; Gershenson et al., 2022; Frakes and Gruber, 2022). I provide evidence on a key mechanism that may explain these results: in my context, officers rely less on external signals when policing racially concordant neighborhoods, suggesting that familiarity or race-specific information improves decision-making. This pattern aligns with evidence from medicine showing that White doctors often have lower-quality information about Black patients (Hoffman et al., 2016). In policing, my findings reinforce the view that increasing minority representation can reduce racial disparities by identifying how such reforms improve officer decision-making (Gudgeon et al., 2023; Rivera, 2025).

Finally, by providing evidence on the causal effects of priority classifications, I clarify findings from descriptive work in criminology and sociology on the role of 911 call takers in law enforcement. This literature emphasizes call takers as gatekeepers to the criminal justice system, shaping which incidents receive a police response and influencing efforts to divert resources from low-level offenses (Lum et al., 2020; Goodier and Lum, 2022). More recent work highlights their role as risk appraisers: in addition to triaging calls, call takers may influence officer expectations by framing incident severity (Gillooly, 2020). I extend this literature by providing the first causal evidence on how call taker risk assessments shape high-stakes officer decisions. My findings build on Gillooly (2022), who uses a similar design to show that priority upgrades cause officers to perceive incidents as more urgent, which she attributes to anchoring bias. While her design captures changes in officer beliefs, mine measures changes in the high-stakes officer decisions that directly impact civilians. Moreover, I document that officers' reliance on this information depends critically on the social context in which it is presented.

The rest of the paper proceeds as follows. Section 2 discusses the institutional details of the Dallas Police Department; Section 3 describes the data; Section 4 details the empirical model; Section 5 presents the results; Section 6 discusses mechanisms; Section 7 concludes.

2 Institutional Background

The setting for my analysis is the Dallas Police Department (DPD), where 911 calls play a central role: 41% of arrests originate from one. Calls are assigned to available call takers through an automated rotation system that alerts the next unoccupied call taker via headset. Unlike cities with multiple call centers, Dallas operates a single centralized facility, ensuring all call takers on a shift receive the same call stream. Shifts are determined once or twice per year through a seniority-based bidding process, similar to officer shift assignments (Ba et al., 2021a). Call takers work five days per week on one of three shifts—morning/mid-afternoon, afternoon/evening, or overnight.

Calls are occasionally transferred within the center, most often when the caller prefers to speak a language other than English. Given that nearly 40% of Dallas residents are Hispanic, Spanish is the most common alternative. If a bilingual call taker is unavailable, a third-party translation service is used. Content-based transfers—where a call is passed because of its nature—are rare and generally discouraged, occurring only when a supervisor intervenes after a call taker loses control of an incident.

Upon answering a call, the call taker assigns one of 85 incident classifications, each covering related event types (e.g., burglaries, shootings) and mapped to a priority level, or queue order, from 1 (highest) to 4 (lowest). Queue order determines response speed.[1] Most classifications are hard-coded to a specific queue order. However, seven incident types allow a choice between two queue levels: residential, business, and vehicle burglaries; thefts; criminal mischief; missing persons; and "Other," a broad categorization for low-level offenses, including trespassing, public intoxication, and outstanding warrants. Within these incident types, call takers can upgrade priority based on risk of violence, presence of weapons, or number of victims. My empirical strategy focuses on these seven incident types. I define a priority call as one where the higher of the two possible queue levels is chosen for the classified call type. Such upgrades increase dispatch speed and signal greater severity, meaning two otherwise similar incidents may receive different responses solely because of discretionary classification.

After classification, the call is routed to a dispatcher, who uses the Computer-Aided Dispatch (CAD) system to assign officers based on location, incident type, and priority. The

[1] Call takers may also add free-text comments for officer reference; these are not observed in the data.

nearest available unit is typically sent, though when several are nearby, one may volunteer.[2] Given the stakes of 911 calls and the unpredictability of future demand, officers have little scope to systematically sort to calls. Consistent with this, Weisburst (2024) shows that, conditional on location-by-time fixed effects, officers in Dallas are quasi-randomly assigned to calls during the time period of this study.

3 Data

This paper uses administrative data from the Dallas Police Department (DPD) and the Dallas County District Attorney's Office, obtained through public records requests. I combine the universe of civilian calls-for-service to DPD between 2015 and 2019 with arrest records, use-of-force reports, criminal charges, police personnel files, and case disposition data from Dallas County courts.

The calls-for-service data come from the Computer Aided Dispatch (CAD) system and include all incidents responded to by DPD officers, whether initiated via 911, 311, or self-initiated activity. Following Weisburst (2024), I restrict the sample to incidents most likely originating from 911 calls. I exclude "Mark-Outs" and low-level incidents with instantaneous dispatch—both indicating self-initiated activity. I also drop calls with unusually long response times and those with classification codes signaling police-initiated events (e.g., bait car activation, officer assistance). The resulting 911 call sample contains the time, location, classification, and queue order of each call, along with badge numbers of responding officers and names of call takers. Call disposition codes—recorded at shift end—indicate incident outcomes, with common values such as "Report" (offense report filed), "No Police Action," and "No Complainant" (complainant absent on arrival).

I link 911 calls to arrest and use-of-force reports via a call identifier, coding a call as involving an arrest or use of force if any associated report shows one occurred. For each arrest, I classify offense severity using two schemes: index offenses and felony/misdemeanors. Index offenses include the eight FBI-tracked serious crimes: murder, rape, robbery, aggravated assault, burglary, theft, arson, and car theft; all others are non-index. About one-fourth of arrests lack charge data, typically reflecting non-criminal matters (e.g., mental health detentions, alcohol detox) or low-level civil infractions that result in citation rather than detention. I label these as "unclassified." Felony/misdemeanor status comes from the charge data, which also determine the appropriate Dallas County court system. Arrests for outstanding war-

[2] Officers can view active calls and details on in-car terminals but do not see which call taker handled a call. According to DPD personnel, officers and call takers are not personally familiar and officers interact almost exclusively with dispatchers.

rants are treated as non-index, but their felony/misdemeanor status is unobserved. When multiple charges exist, I classify by the most serious charge.

To link arrests to court outcomes, I perform a fuzzy match between the arrestee's name and court case records using first and last names and the offense date.[3] I define a defendant as guilty if convicted by judge or jury or if they plead guilty (including plea deals). At the call level, a guilty outcome is recorded if any arrestee from that call is found guilty in court. Finally, I merge the 911 data with 2010 Census and American Community Survey block group demographics, using geocoded call coordinates to assign the origin block group's characteristics. The raw sample contains 2,411,148 calls handled by 1,444 unique call takers.

I apply two more restrictions to construct the analysis sample. First, I limit to calls handled by full-time call takers and responded to by patrol officers. Trainees and officers temporarily reassigned to the call center are excluded. Using DPD staffing records, I identify full-time call takers as the top-ranked individuals by annual call volume, up to the number reported as full-time each year. For example, if 70 full-time call takers are reported in 2014, I retain the 70 call takers with the highest call counts that year. This procedure is repeated annually. These restrictions yield 1,907,202 911 calls handled by 164 call takers. Second, I restrict to the seven incident classifications with variation in priority assignment, as described in Section 2. This final restriction yields a baseline sample of 597,973 calls across all 164 full-time call takers.

3.1 Summary Statistics

Table 1 presents summary statistics for each call in the analysis sample, broken out by call type. These calls represent roughly 1/3 of all 911 calls in the sample period. The "Other" category is the second most frequent call code in the full 911 data, accounting for 20% of all calls. It is also by far the most frequent incident type in the analysis sample, comprising over half of all calls. Calls result in arrest 2.9% of the time, and 38% are upgraded to priority status. Nearly half of "Other" calls are given priority status, and 4.1% result in arrest—both higher than the other call types, which have priority rates between 10% and 37% and arrest rates below 1%.

Nearly half of arrests carry misdemeanor charges, and 12% (not listed) involve felonies. The remainder either lack formal charges (e.g., detentions for mental health crises) or are for outstanding warrants, which do not have charge information. Only 21.6% of arrests result in conviction in Dallas courts. To give a clearer sense of the types of offense present in my

[3] I match using the first two letters of the first name and the first three letters of the last name. Matches are retained if the first and last names have a Levenshtein distance of two or less and the offense date is within 10 days of the arrest date. Matching is implemented using the `stringdist` package in R.

sample, Supplementary Appendix Table A1 lists the 20 most frequent charges for in-sample arrests. Nearly one-third are for prior warrants, which are unlikely to link to court data because offense and arrest dates may not align, and the warrant may be for an outside agency. A substantial share of arrests are for non-index crimes—such as disorderly conduct, drug offenses, trespassing, and public intoxication. In total two-thirds of all arrests in the sample are for non-index offenses.

Supplementary Appendix Table A2 provides summary statistics for the 164 full-time call takers. Call takers are less white than patrol officers: only 21% are white, compared to nearly half of patrol officers. On average, each answers over 3,500 in-sample 911 calls. Priority assignment rates vary widely—from 21% to 66%—partly due to turnover, as not all call takers appear in the data for the full period. The average call taker is in the sample for 2.1 years.[4] Only 3.5% of calls are transferred, consistent with the low transfer rates reported by DPD call takers I interviewed. There are no notable outliers; the maximum transfer rate is 5%. For the analysis, I assign the call taker as the person who first answered the call.

Arrest rates differ sharply between priority and non-priority calls (Figure 1). Only 1.28% of non-priority calls lead to arrest, compared with 5.42% of priority calls—a pattern observed across all call types. This reflects both differences in the alleged crimes and in the police response. On average, arrests from priority calls are of lower quality: conviction rates are 9 percentage points *lower* for these arrests. Supplementary Appendix Figure A1 shows that this negative gap is driven mainly by "Other" and Missing Persons calls, the only incident types where conviction rates drop for priority arrests relative to arrests made at non-priority calls.

Overall, the descriptive evidence indicates that priority and non-priority calls yield different outcomes. However, without plausibly exogenous variation in priority assignment, it is impossible to separate the causal effects of priority status from selection into different types of incidents, which motivates the examiner design strategy described in the next section.

4 Empirical Design

My empirical strategy exploits the quasi-random assignment of call takers to estimate the causal effect of priority designations on enforcement decisions. I begin with the following baseline linear specification:

[4] While I cannot observe total employment duration for all call takers, I can identify those who are newly full-time during the sample period, who account for 45% of call takers. This aligns with internal DPD documents indicating high turnover among call takers at the time.

$$Arrest_c = \beta_0 + \beta_1 Priority_c + \beta_2 \mathbf{X_c} + \epsilon_c, \tag{1}$$

where $Arrest_c$ is the arrest outcome for call c; $Priority_c \in \{0, 1\}$ is the call's priority, which equals 1 when the higher-priority call code is chosen for the classified incident type; and $\mathbf{X_c}$ includes pre-determined location features, such as police division and block group demographics. The error term ϵ_c captures unobserved determinants of arrest, including suspect behavior. Because call takers assign priority based on information not fully captured in the data, $Priority_c$ is likely correlated with ϵ_c, rendering OLS estimates of β_1 biased upward.

To address this, I implement an examiner design using call takers' inherent propensities to upgrade calls, which I denote as the *Call Taker Score*, as an instrument for individual call priority. Call Taker Score is constructed using a leave-out mean to avoid mechanical correlation that would arise from using an observation's priority to construct its own instrument value (Arnold et al., 2018). For call c, I first residualize the priority indicator:

$$P_c^* = Priority_c - \alpha \mathbf{X_c^p} \tag{2}$$

$$= Z_{cj} + \epsilon_c, \tag{3}$$

where $\mathbf{X_c^p}$ includes day-of-week-by-hour, month-by-year, and Hispanic call taker fixed effects. These controls isolate variation in priority arising from call taker differences rather than systematic assignment features, such as shift schedules, seasonal crime trends, or potential translation requirements for Spanish-speaking callers. The residualized priority P_c^* is then aggregated to compute Call Taker Score using a leave-out mean:

$$Z_{cj} = \left(\frac{1}{n_j - 1} \right) \left(\sum_{k=0}^{n_j} P_{kj}^* - P_{cj}^* \right), \tag{4}$$

where n_j is the number of calls answered by call taker j. Figure 2 shows substantial variation in Call Taker Score, with a 90th percentile call taker 18 percentage points more likely to assign priority than a 10th percentile call taker—47% of the mean priority rate. Figures A2 and A3 further show that less white and less experienced call takers are more likely to upgrade calls, though there is a concentrated tail of experienced call takers with high priority propensities.

Under three assumptions, IV estimation of equation 1 using Call Taker Score as an instrument for Priority will identify a Local Average Treatment Effect (LATE). In particular, the LATE identified is a variance-weighted average of the effects of a call being given priority status just because a call taker with a slightly higher propensity to give a priority designation was assigned (Imbens and Angrist, 1994).

The first assumption is that the instrument must affect priority assignment. The first-stage relationship is:

$$Priority_c = \gamma_1 Z_{cj} + \gamma_2 \mathbf{X_c} + u_c. \tag{5}$$

Figure 2 plots the first stage relationship using a nonparametric local linear estimator. The OLS estimates of the first stage are presented in column 1 of Table A3. The OLS estimate of γ_1 is 0.9983. The first stage F-statistic is 13,382, which is well past the rule of thumb threshold.[5] This suggests that the relevance assumption holds.

The second assumption is that Call Taker Score is correlated with a call's arrest outcome only through the priority level assigned to it, conditional on the time and call taker ethnicity controls. In my setting, this assumption relies on quasi-random assignment of call takers to calls. As a test of this assumption, I evaluate whether Call Taker Score is correlated with features of the call's location, which are the only observables in the data that are determined before the call taker answers the phone. Specifically, I regress Call Taker Score on the call-level location observables, along with the fixed effects necessary for identification:

$$CallTakerScore_c = \psi' \mathbf{X_c} + \epsilon_c. \tag{6}$$

I report the estimates, along with estimates from a benchmark regression that uses the endogenous Priority indicator as the outcome, in Table 2. The p-value for an F-test of joint significance of the location variables is 0.123, suggesting that call location does not determine the assigned call taker. Two of the covariates for the Call Taker Score regression are statistically significant at the 5% level (Southeast and Southwest Division), while minority proportion of the call's Census Block Group is also statistically significant at the 10% level. However, the point estimates are small, not exceeding 1.5% of standard deviation of Call Taker Score, and are not indicative of systematic sorting of call takers to calls. Location variables explain little of the variation in the instrument; the incremental R^2 for the relevant controls is 0.00004. In contrast, location is highly correlated with a call's realized priority. Over half (6/10) of the location coefficients are statistically significant at the 1% level, and an F-test of joint significance soundly rejects joint nullity (p-value ≈ 0). Given the tight links between crime and location (Weisburd and Eck, 2004), the lack of systematic association between call takers and location features suggests it is unlikely that call takers select

[5] I use the robust Kleibergen-Paap F-statistic, which is equivalent to Olea and Pflueger (2013)'s effective F-statistic in the just-identified case (Andrews et al., 2019).

on unobserved features of the call. I interpret this evidence as support for the exclusion restriction.

It is still possible for the exclusion restriction to be violated if call takers have a direct effect on arrests. Because officers do not observe the call taker and call takers do not directly communicate with officers, this is unlikely. One might still worry that call taker notes justifying priority choices could influence officer behavior. Column 2 of Table A3 presents the reduced form regression of Call Taker Score on arrests, capturing the effect of receiving a call from a call taker more likely to upgrade calls, including any effects through channels other than priority. These estimates are statistically indistinguishable from the IV estimates in the next section.

Other features of the setting could threaten the exclusion restriction. First, because I restrict the sample to call types with call taker discretion, there could be endogenous selection. Supplementary Appendix Table A5 shows that Call Taker Score is uncorrelated with selection into the sample, mitigating this concern. Second, call takers may influence other call dimensions, such as incident type. Section 5 addresses this by instrumenting for all possible incident types, yielding similar results. Finally, call takers may transfer calls to others; my assignment of the initial call taker reduces this concern, and Section 5 shows that explicitly controlling for transfer propensity does not alter results.

The third assumption required for a LATE interpretation is that the effect of Call Taker Score on priority setting is monotonic across calls. That is, there can be no calls for which a call taker with a low score would assign priority but a high score call taker would not. I perform a partial test of the monotonicity assumption by calculating the first stage relationship between Call Taker Score and priority within subgroups of the analysis sample. In Supplementary Appendix Table A4, I show that these first stage relationships are positive and statistically significant across all tested subsamples of the data.

5 Results

Columns 1 and 2 of Table 3 present the baseline OLS and IV estimates for Equation (1), respectively. Using two-stage least squares with the Call Taker Score instrument, the estimated coefficient on $Priority$ is 0.015, or 1.5 additional arrests per 100 calls, roughly a 53% increase relative to the mean arrest rate. The OLS estimate for β_1 is 0.039, more than twice the IV estimate. This causal effect implies that officers are more likely to make an arrest solely because a call is assigned a higher priority. Marginal decisions by call takers thus have large downstream effects on enforcement outcomes.

Despite evidence from Section 4 suggesting quasi-random assignment of call takers, other violations of the exclusion restriction could influence the baseline IV estimates. In columns 3–5 of Table 3, I implement alternative specifications to assess the robustness of the results. First, I consider potential endogenous selection of call takers into the incident types in the sample. Column 3 controls for the call taker's propensity to classify calls as each incident type in the sample. I construct this variable using the full call dataset as the leave-out mean in-sample incident rate, following Herbst (2023). The estimated effect of priority is similar to the baseline. Second, to address the multidimensional nature of call-taker decisions—classifying both crime type and priority—I control in column 4 for call takers' propensities to classify calls into each incident type, also calculated as leave-out means (Bhuller et al., 2020). The estimate is slightly smaller at 0.013, but the baseline estimate remains well within the 95% confidence interval. Finally, column 5 controls for the assigned call taker's propensity to transfer calls to other call takers, again using a leave-out mean; the estimate remains similar. Taken together, these results support a clear causal interpretation of the baseline IV estimate.

Which arrests are responsive to upgraded priority? To answer this, I separately estimate the effect of priority on index and non-index arrests. I also examine "unclassified" arrests, which include non-criminal apprehensions or citations for low-level ordinance violations. Results are reported in Supplementary Appendix Table A6. Priority has statistically significant effects only on non-index and unclassified arrests, with nearly two-thirds of the total effect arising from non-index offenses and the remaining third from non-criminal behaviors. The effect on index arrests is smaller in magnitude and statistically indistinguishable from zero. These findings suggest that discretionary arrests are most responsive to exogenous changes in upstream information. Non-index arrests primarily involve victimless offenses such as disorderly conduct, drug possession, and public intoxication. Unclassified arrests also likely involve substantial discretion, as officers may, for example, detain individuals in crisis when perceiving heightened risk.

An important secondary outcome is use of force. Supplementary Appendix Table A7 reports the effect of priority on officer use of force. The estimate is positive, but less than 10% of the mean and not statistically significant. The estimate is quite imprecise, as I cannot rule out large positive effects exceeding 100% of the mean. These results suggest that priority designations may primarily affect officers' perception of whether a crime occurred, rather than their assessment of personal risk. However, given the rarity of use of force, the estimate has limited power and cannot definitively rule out meaningful effects.

Overall, these results indicate that call taker discretion has substantial implications for police enforcement. For calls where call takers may disagree on the appropriate designation, assigning a priority classification significantly increases the likelihood of an arrest, especially for low-severity offenses where officer discretion carries more weight. While policy changes targeting inefficiencies in policing have focused overwhelmingly on law enforcement officers, these findings suggest that reforms in 911 call taking—an under-examined component of the policing process—may also have meaningful impact.

6 Mechanism: Information Processing

With the results of the previous section in mind, an open question is whether call takers' effects are driven by changes to officer perceptions, as opposed to more mechanical reductions in response time. In Supplementary Appendix B, I show that faster response times likely explain part of the effect. In this section, I explore whether any of the effect can be attributed to officers' evaluations of suspects. Since information provided by call takers is a critical input to police decisions, understanding how officers use it provides insight into their decision-making process and has implications for policies aimed at reducing aggressive policing.

To test for information effects, I first examine the effect of priority on arrest quality, proxied by court convictions. If officers use priority to inform their evaluation of a suspect's guilt, arrests at marginally upgraded calls should be less successful in court than arrests at non-upgraded calls. I restrict the sample to calls that result in arrests with charges (excluding warrants and non-criminal arrests) and estimate the baseline specification using conviction as the outcome. Because Dallas has separate courts for felony and misdemeanor offenses, I estimate specifications separately for each. Results are presented in Table A8. The smaller sample reduces first-stage strength (Kleibergen-Paap F statistics of 155 and 53 for misdemeanor and felony arrests, respectively), though the instrument remains relevant. Point estimates are negative for both felonies and misdemeanors, suggesting that arrests at priority calls are less likely to result in convictions; however, only the misdemeanor effect is statistically significant. For misdemeanors, the estimated effect is -16.2 percentage points, over a 50% reduction relative to the mean conviction rate. I interpret this as evidence that priority effects on arrests are partially driven by the information supplied to officers.

Officer sensitivity to priority likely depends on the strength of their priors regarding suspect guilt. Priority should have less influence when officers have stronger priors. I consider two cases in which priors plausibly vary: interactions with suspects who are the same race as the officer and officer experience. Officers may have tighter priors when interacting with same-race suspects due to shared socio-cultural knowledge and more effective communication with witnesses. Similarly, greater tenure provides more experience to form priors, reducing the influence of any single piece of information.

I first consider calls dispatched to neighborhoods that are racially similar to the officer.

Using officer-call level data and following (Hoekstra and Sloan, 2022), I estimate a difference-in-difference specification:

$$\begin{aligned} Arrest_{ic} = {} & \alpha_0 + \alpha_1(ProportionSameRace_{ic}) + \alpha_2 Priority_c \\ & + \alpha_3(Priority * ProportionSameRace)_{ic} \\ & + Officer_i + \alpha_4 X_c + u_{ic}, \end{aligned} \tag{7}$$

where $ProportionSameRace_{ic}$ is the proportion of civilians in the call's Census Block Group who share officer i's race. The coefficient α_3 measures how the difference between priority and non-priority arrest rates changes as neighborhoods increasingly resemble the officer's race. Officer fixed effects, $Officer_i$, control for the possibility that officers more responsive to priority calls are concentrated in high-crime areas. I instrument for $Priority_c$ and $(PriorityProportionSameRace)_{ic}$ using $CallTakerScore_c$ and $(CallTakerScore * ProportionSameRace)_{ic}$ to address the fact that priority calls are more likely to result in arrest for reasons observed by the officer but unobserved by the econometrician.

Figure 3 illustrates the intuition: in neighborhoods with few same-race residents, the gap in arrest rates between likely priority and non-priority calls is largest; as neighborhoods become more racially similar to the officer, the gap shrinks. In column 1 of Table 4, I present results for IV estimation of equation 7. These officer-level regressions are weighted by the inverse number of responding officers at the call. The α_3 estimate is -0.0143, significant at the 5% level, implying that a 10pp increase in neighborhood racial similarity reduces the effect of a priority signal on arrests by 0.0014pp. Marginal priority changes thus have less influence in racially similar neighborhoods.

I further examine which officers are driving this phenomenon using a triple-differences specification. For an officer of generic "Race", these regressions take the following form:

$$\begin{aligned} Arrest_{ic} = {} & a_0 + a_1(Proportion\text{“}Race\text{”}_{ic}) + a_2 Priority_c + Officer_i \\ & + a_3 OtherInteractions_{ic} + a_4(Priority * \text{“}Race\text{”}Officer * Proportion\text{“}Race\text{”})_{ic} \\ & + a_5 X_c + e_{ic}, \end{aligned} \tag{8}$$

where $OtherInteractions_{ic}$ includes all lower-order interactions of $Priority_c$, officer race, and neighborhood racial composition. The coefficient a_4 measures how the priority arrest gap changes as a neighborhood becomes more racially aligned with the officer, relative to non-"Race" officers. Supplementary Appendix Figure A4 shows that white and Black officers become relatively less responsive to priority in same-race neighborhoods, whereas Hispanic officers exhibit minimal change..

Columns 2–4 of Table 4 confirm these patterns. For white officers, a 10pp increase in neighborhood whiteness reduces the priority arrest gap by 0.37pp (13% of the mean arrest rate). For Black officers, a 10pp increase in neighborhood Black share reduces the gap by 0.51pp (17% of the mean). These estimates are significant at the 5% and 10% levels, respectively. For Hispanic officers, the triple interaction estimate is 0.0163, implying slightly higher responsiveness in Hispanic neighborhoods, though the effect is imprecisely estimated and I am unable to rule out an economically meaningful negative estimate.

I next examine officer experience by replacing $ProportionSameRace_{ic}$ in equation 7 with years of experience. Table 5 reports the results. The interaction term is -0.0008, significant at the 5% level, indicating that each additional year of experience reduces the effect of priority by 0.08pp. These effect sizes imply that rookie officers are 2pp more likely to arrest at marginal priority calls, and this effect shrinks as the officer gains experience, disappearing entirely by the time an officer has been working for 25 years.

To ensure my results are not driven by officers sorting to assignments at more granular spatial or temporal levels, I reproduce the difference-in-difference estimates from Tables 4 and 5, replacing the location and time fixed effects from the baseline models with high-dimensional call beat-by-time fixed effects. Compared to divisions, of which there are only 7 in Dallas, there are over 230 unique police beats, allowing this specification to control for hyper-local variation in call severity. Specifically, I control for beat-by-shift-by-day of week-by-year fixed effects, so that the interaction terms are estimated relative of the average arrest rate within a particular day of the week, 8-hour shift, year, and patrol beat cell. The results, presented in Supplementary Appendix Table A9, are similar to the baseline specification.

These results provide strong evidence that officers' information processing is crucial in explaining how call takers' priority decisions affect arrests. Moreover, officers rely more on priority information in contexts where their priors are weaker, highlighting the role of information processing in generating disparate and inconsistent police decisions across situations that should otherwise be judged similarly.

7 Conclusion

This paper provides the first empirical evidence that 911 call takers' risk assessments affect police officers' enforcement decisions. Leveraging the quasi-random assignment of call takers to 911 calls in Dallas, Texas, I show that police are more likely to make arrests at calls upgraded to priority solely due to the underlying call taker's risk preferences. This effect is strongest for low-level arrests, where officer discretion is most salient. I further demonstrate that this effect arises, at least in part, from officers' reliance on priority as a key source of information when deciding whether to arrest. Misdemeanor arrests at marginally upgraded calls are less likely to result in convictions, and officers are more responsive to priority upgrades when they are inexperienced or when the call occurs in a neighborhood that does not match their own racial or ethnic identity. These findings highlight that information from 911—and the discretion exercised by call takers—is a critical yet underappreciated input into policing decisions.

One simple and potentially cost-effective tool for mitigating the unintended effects of call taker discretion could be to provide officers with the identity of the call taker responsible for the assignment. In many large police agencies, call takers and dispatchers occupy separate roles, and officers communicate almost exclusively with the dispatchers (Transform911, 2022). Providing officers with information about who handled the call could allow them to adjust for systematic differences in call taker classifications after observing enough calls from each individual. In jurisdictions where 911 call center workers both take calls and dispatch officers (Gillooly, 2023), examining officer responses to call information could offer insight into the potential effectiveness of such a policy, since officers would directly communicate with the people who make call classifications.

These results also suggest directions for future research on the 911 process. For instance, some U.S. jurisdictions have recently implemented text-to-911 systems, which remove many of the emotional cues that call takers use to classify calls. Additionally, training requirements for 911 call takers vary widely across states. Given the substantial influence that this paper documents, understanding how such changes to call-taking processes affect police outcomes represents an important avenue for future work.

References

Andrews, I., Stock, J., and Sun, L. (2019). Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*, 1.

Arnold, D., Dobbie, W., and Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, 133:1885–1932.

Ba, B., Bayer, P., Rim, N., Rivera, R., and Sidibé, M. (2021a). Police officer assignment and neighborhood crime. *Working Paper*.

Ba, B. A., Knox, D., Mummolo, J., and Rivera, R. (2021b). The role of officer race and gender in police-civilian interactions in chicago. *Science*, 371:696–702.

Bhuller, M., Dahl, G. B., Løken, K. V., and Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, 128:1269–1324.

Cunningham, J., Feir, D., and Gillezeau, R. (2021). Collective bargaining rights, policing, and civilian deaths. *Working Paper*.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108:201–240.

Doyle, J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review*, 97:1583–1610.

Dube, O., MacArthur, S. J., and Shah, A. K. (2025). A cognitive view of policing. *The Quarterly Journal of Economics*, 140:745–791.

Frakes, M. and Gruber, J. (2022). Racial concordance and the quality of medical care: Evidence from the military. *NBER Working Paper 30767*.

Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C. A., and Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, 14:300–342.

Gillooly, J. W. (2020). How 911 callers and call-takers impact police encounters with the public: The case of the henry louis gates jr. arrest. *Criminology and Public Policy*, 19:787–804.

Gillooly, J. W. (2022). “lights and sirens”: Variation in 911 call-taker risk appraisal and its effects on police officer perceptions at the scene. *Journal of Policy Analysis and Management*, 41:762–786.

Gillooly, J. W. (2023). Collaborative gatekeeping: Consensus-seeking practices among emergency call-takers. *Policing (Oxford)*, 17.

Goncalves, F. and Mello, S. (2021). A few bad apples? racial bias in policing. *American Economic Review*, 111:1406–1441.

Goodier, M. and Lum, C. (2022). First point of contact: Can procedural justice be applied by emergency calltakers? *Policing: A Journal of Policy and Practice*.

Gudgeon, M., Jordan, A., and Kim, T. (2023). Do teams perform differently under black and hispanic leaders? evidence from the chicago police department. *Working Paper*.

Herbst, D. (2023). The impact of income-driven repayment on student borrower outcomes. *American Economic Journal: Applied Economics*, 15:1–25.

Hoekstra, M. and Sloan, C. W. (2022). Does race matter for police use of force? evidence from 911 calls. *American Economic Review*, 112:827–860.

Hoffman, K. M., Trawalter, S., Axt, J. R., and Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America*, 113:4296–4301.

Holz, J. E., Rivera, R. G., and Ba, B. A. (2023). Peer effects in police use of force. *American Economic Journal: Economic Policy*, 15:256–291.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62:467.

Lum, C., Koper, C. S., Stoltz, M., Goodier, M., Johnson, W., Prince, H., and Wu, X. (2020). Constrained gatekeepers of the criminal justice footprint: A systematic social observation study of 9-1-1 calltakers and dispatchers. *Justice Quarterly*, 37:1176–1198.

Mas, A. (2006). Pay, reference points, and police performance. *Quarterly Journal of Economics*, 121:783–821.

Olea, J. L. M. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics*, 31:358–369.

Rivera, R. (2025). Do peers matter in the police academy? *American Economic Journal: Applied Economics*, 17:127–164.

Tapp, S. and Davis, E. (2024). Contacts between police and the public, 2022. Technical report, U.S. Department of Justice, Bureau of Justice Statistics.

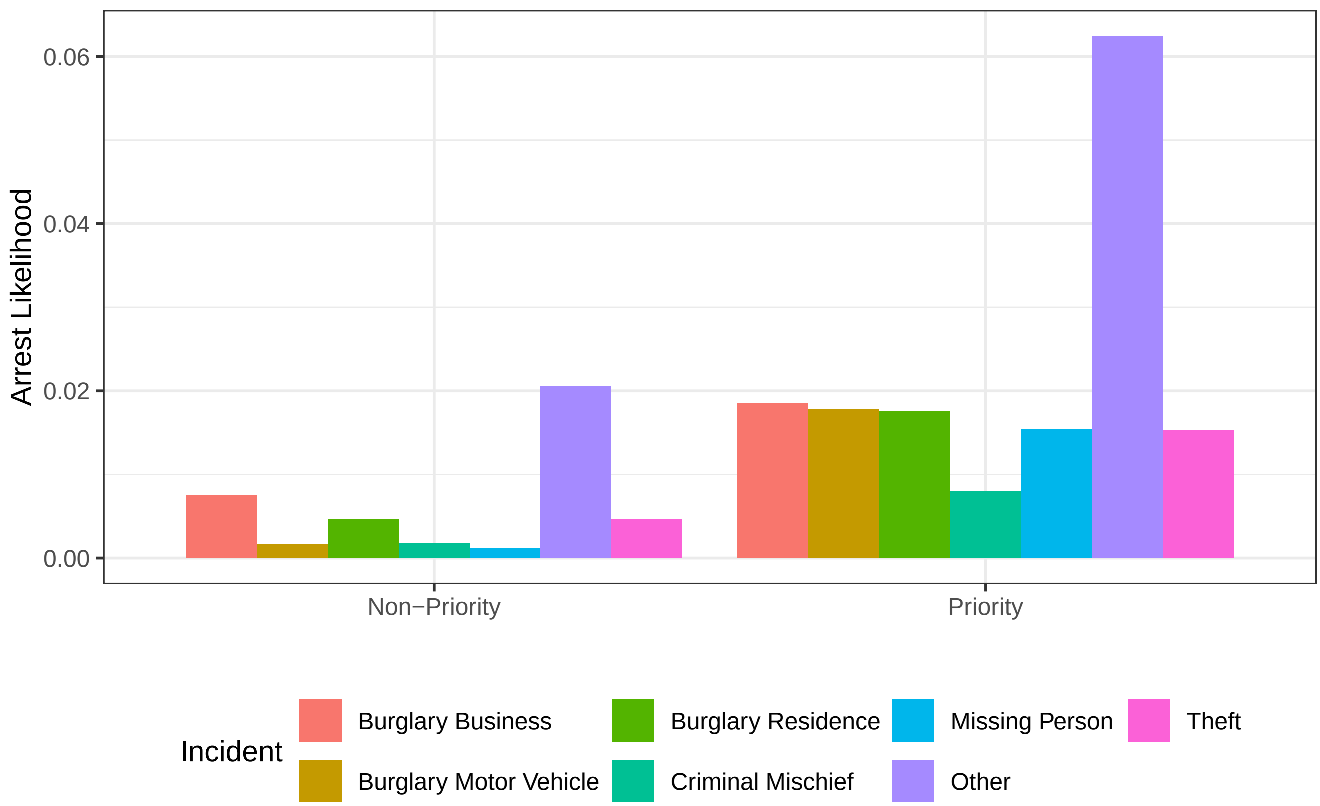
Transform911 (2022). Transforming 911 report.

Weisburd, D. and Eck, J. E. (2004). What can police do to reduce crime, disorder, and fear? *Annals of the American Academy of Political and Social Science*, 593:42–65.

Weisburst, E. K. (2024). Whose help is on the way? *Journal of Human Resources*, 59:1122–1149.

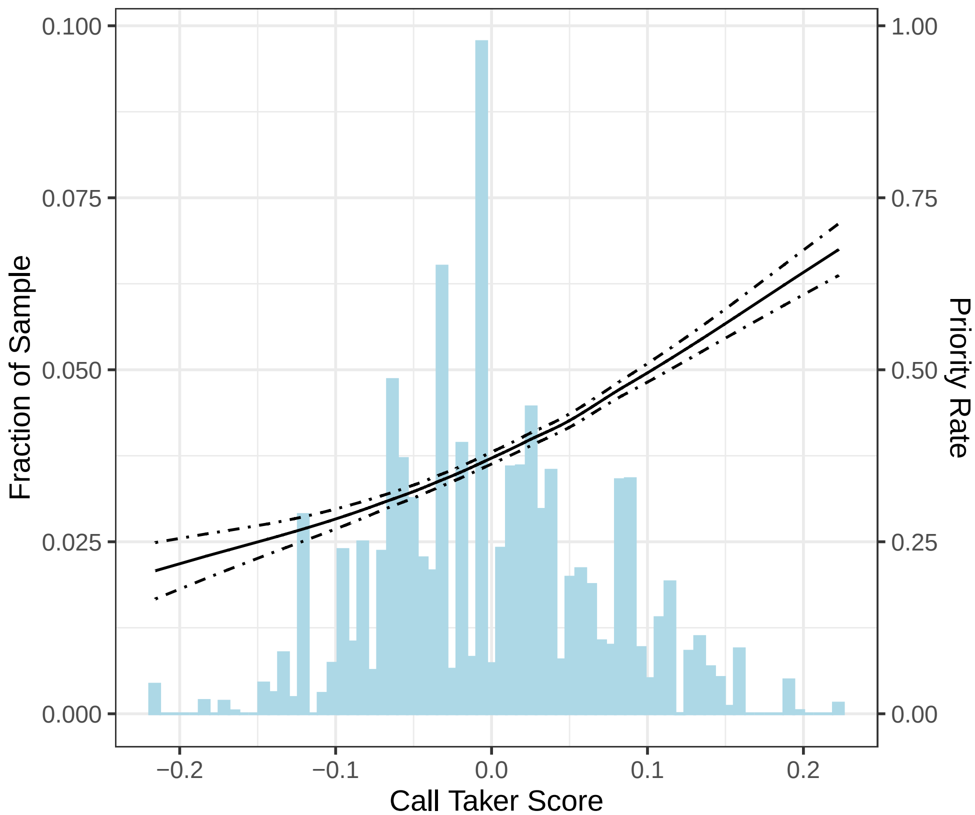
8 Tables and Figures

Figure 1: Arrest Rate by Priority



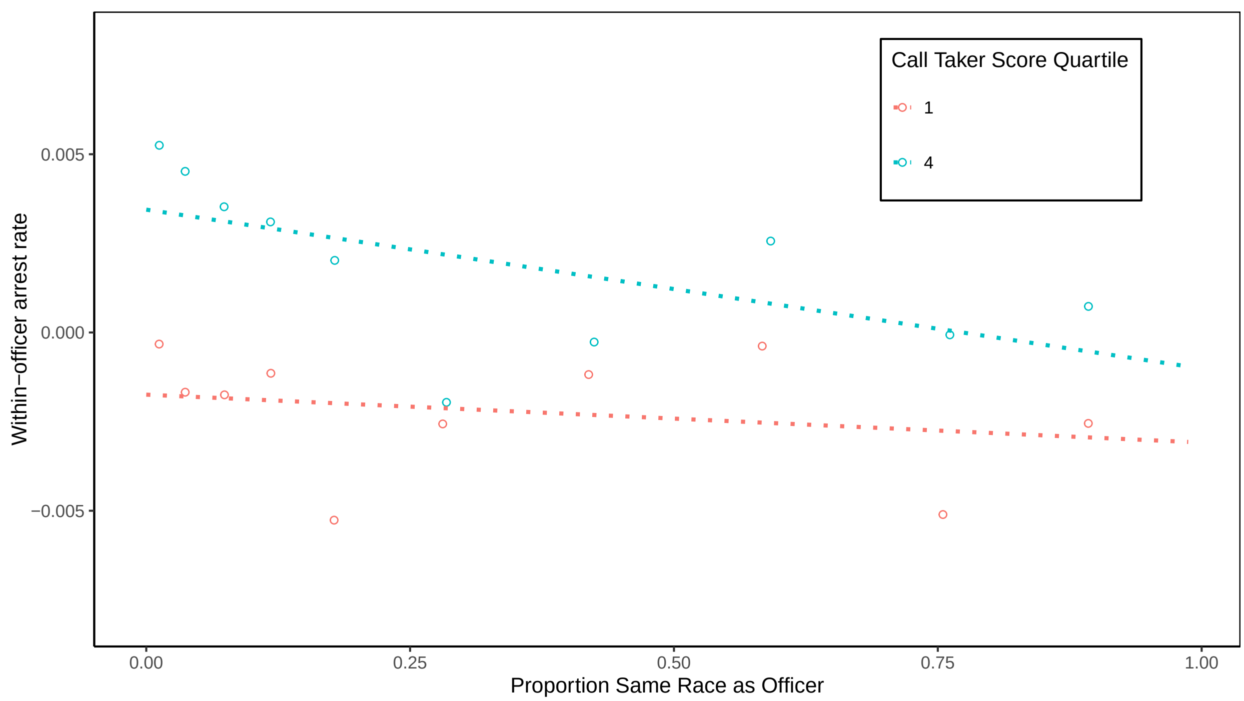
Notes: This figure depicts arrest rate for the 7 incident types included in the analysis sample, separated by Priority classification.

Figure 2: Call Taker Score Variation



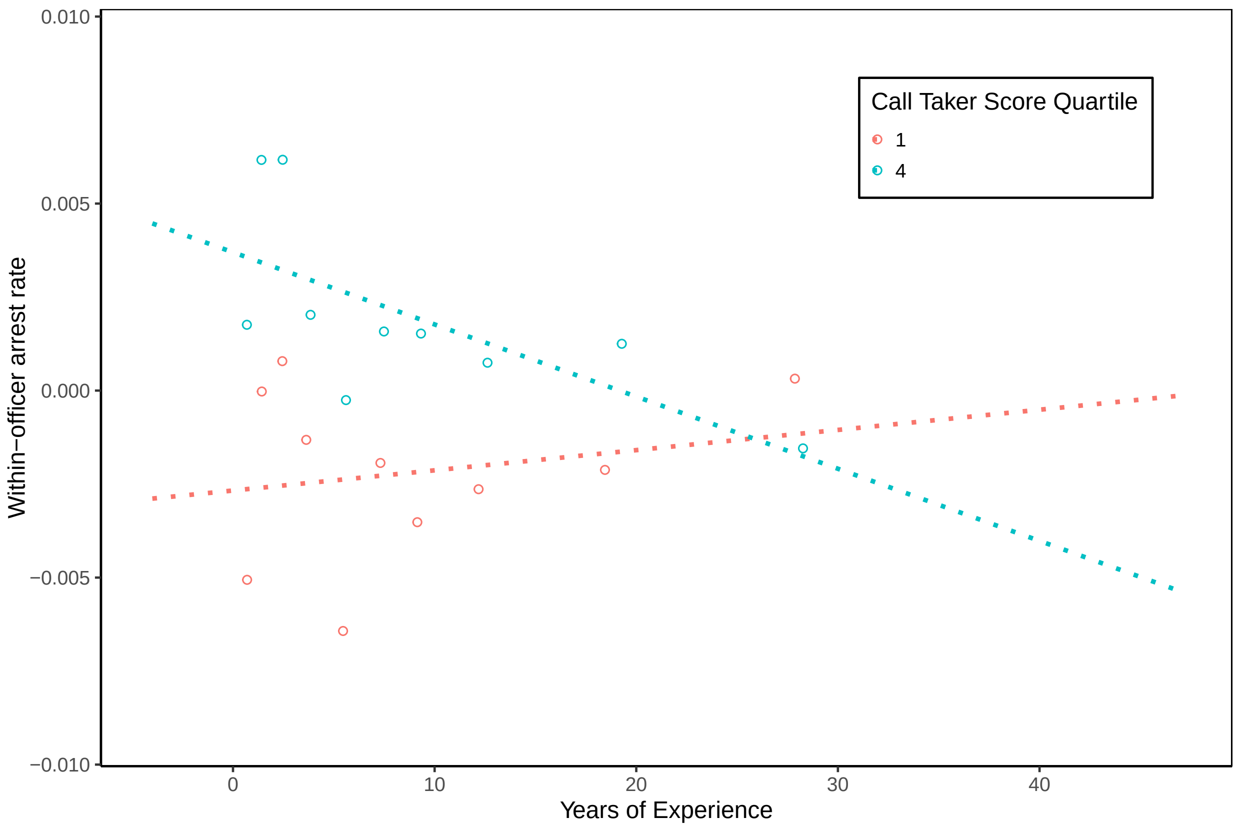
Notes: This figure reports the distribution of Call Taker Score and the first-stage relationship between Call Taker Score and call priority. Call Taker Score is the leave-out mean priority rate, calculated using data from other calls answered by the call taker following the procedure described in Section 4. The solid black line, plotted against the right axis, denotes predicted means from a local linear regression of Priority signal on Call Taker Score. The histogram, plotted against the left axis, provides the distribution of Call Taker Score measures across all 911 calls in the analysis sample.

Figure 3: Arrest Rates by Neighborhood Racial Composition



Notes: The y-axis measures the likelihood of an arrest relative to an officer's average arrest rate in the sample. The x-axis is the proportion of the Census Block Group that is the same race as the officer. Observations are grouped so that each point includes an equal number of calls. Only calls in the top (blue) and bottom (red) quartiles of Call Taker Score are included in the figure. The fitted lines are linear fits across each of the plotted Call Taker Score quartiles.

Figure 4: Arrest Rates by Experience



Notes: The y-axis measures the likelihood of an arrest relative to an officer's average arrest rate in the sample. The x-axis is the officer's years of experience in the Dallas Police Department. Observations are grouped so that each point includes an equal number of calls. Only calls in the top (blue) and bottom (red) quartiles of Call Taker Score are included in the figure. The fitted lines are linear fits across each of the plotted call taker score quartiles.

Table 1: Summary Statistics

Incident Type	Total Calls	% of All Calls	Priority Rate	Pr(Arrest)	Misdemeanor Charges|Arrest	Conviction Rate
Burglary Business	12558	0.007	0.107	0.009	0.193	0.496
Burglary Motor Vehicle	56384	0.030	0.105	0.003	0.544	0.526
Burglary Residence	38385	0.020	0.170	0.007	0.271	0.456
Criminal Mischief	35129	0.018	0.185	0.003	0.390	0.264
Missing Person	21239	0.011	0.368	0.006	0.044	0.043
Other	391991	0.206	0.482	0.041	0.476	0.205
Theft	42287	0.022	0.276	0.008	0.413	0.371
Analysis Sample	597973	0.314	0.383	0.029	0.467	0.216

Notes: This table presents summary statistics for the analysis sample, separated by call type. The column Prop. of All Calls measures the proportion of all 911 calls dispatched in Dallas that comprise of the listed call type.

Table 2: Covariate Balance

Dependent Variables:	Priority	Call Taker Score
Model:	(1)	(2)
Variables		
Proportion Minority	-0.0655***	-0.0015*
	(0.0057)	(0.0008)
Proportion No Degree	0.0014	0.0011
	(0.0062)	(0.0019)
Proportion Unemployed	0.0178	-0.0006
	(0.0121)	(0.0016)
Log(Income per Capita)	0.0028	0.0001
	(0.0023)	(0.0003)
Division - North Central	-0.0222***	0.0008
	(0.0037)	(0.0006)
Division - Northeast	-0.0015	0.0004
	(0.0030)	(0.0004)
Division - Northwest	-0.0122***	0.0000
	(0.0031)	(0.0005)
Division - South Central	0.0383***	0.0008
	(0.0034)	(0.0005)
Division - Southeast	0.0251***	0.0011**
	(0.0032)	(0.0005)
Division - Southwest	0.0145***	0.0010**
	(0.0031)	(0.0005)
Fit statistics		
Observations	597,973	597,973
Incremental R^2	0.00144	0.00004
F Stat	73.795	1.5259
p-value	0.00	0.12291

Notes: This table reports results from a test of covariate balance. Column 1 uses the endogeneous priority variable as the outcome and Column 2 uses the Call Taker Score as the outcome. Incremental R^2 reports the R^2 added to the regression for just the variables with reported estimates. Proportion Minority, Proportion No Degree, Proportion Unemployed, and Log(Income Per Capita) are each calculated at the call Census Block Group level. Each regression includes month-by-year and day of week-by-hour fixed effects, and an indicator for whether the call taker is Hispanic. Standard errors are clustered at the call taker level. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 3: Effect of Priority on Arrests

	(1)	(2)	(3)	(4)	(5)
Priority	0.0394***	0.0152***	0.0154***	0.0130***	0.0151***
	(0.0009)	(0.0044)	(0.0042)	(0.0044)	(0.0044)
Estimation Type	OLS	IV	IV	IV	IV
In-sample Propensity	No	No	Yes	No	No
Incident Type Propensities	No	No	No	Yes	No
Transfer Propensity	No	No	No	No	Yes
Observations	597,973	597,973	597,973	597,973	597,973
R^2	0.01834	0.01345	0.01316	0.01257	0.01342
Y mean	0.02862	0.02862	0.02862	0.02862	0.02862
KP F-stat		13382.8083	13378.8372	10492.2772	13293.9095

Notes: This table contains results for the estimation of β_1 under various specifications of equation 1. Column 1 estimates the equation using OLS. Columns 2-5 use 2SLS with Call Taker Score as the instrument for Priority. Column 3 adds a control for the handling call taker's propensity to classify calls into the sample, calculated as leave-out mean in-sample incident rate. Colum 4 controls for the assigned call taker's propensity to assign each of the 7 different incident types in the sample, also calculated using leave-out means. Column 5 controls for the assigned call taker's propensity to transfer calls, calculated using a leave-out mean. Each regression includes month-by-year, day of week-by-hour, and Division fixed effects, an indicator for whether the call taker is Hispanic, and Census Block Group controls for proportion minority, proportion no high school degree, proportion unemployed, and log per capita income. KP F-stat denotes the Kleibergen-Paap Robust F-Statistic from the first-stage regression. Standard errors are clustered at the call taker level. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 4: IV Results by Race

Dependent Variable:	Arrest			
Model:	Diff-in-Diff	Triple-Diff	Triple-Diff	Triple-Diff
“Race” =		White	Black	Hispanic
	(1)	(2)	(3)	(4)
Variables				
Priority	0.0178***	0.0095	0.0100	0.0118**
	(0.0049)	(0.0060)	(0.0063)	(0.0059)
Proportion Same Race	0.0044			
	(0.0028)			
Priority * Proportion Same Race	-0.0143**			
	(0.0072)			
Priority * Proportion ”Race” * ”Race” Officer		-0.0370**	-0.0511***	0.0163
		(0.0154)	(0.0167)	(0.0148)
Priority * Proportion“Race”		0.0185	0.0115	0.0011
		(0.0138)	(0.0140)	(0.0110)
Priority * “Race” Officer		0.0080	0.0180**	-0.0033
		(0.0056)	(0.0078)	(0.0074)
Proportion “Race” * “Race” Officer		0.0126**	0.0164**	-0.0062
		(0.0063)	(0.0065)	(0.0056)
Proportion “Race”		-0.0071	-0.0017	0.0001
		(0.0055)	(0.0058)	(0.0044)
Fit statistics				
Observations	1,170,500	1,170,500	1,170,500	1,170,500
R^2	0.02184	0.02179	0.02171	0.02198
Y mean	0.02862	0.02862	0.02862	0.02862

Notes: The table reports results for regressions using the specifications in equations 7 and 8. Regressions are performed at the officer-by-call level. Each observation within a call is weighted by the inverse number of responding officers. Column 1 reports the difference-in-difference specification from equation 7. Proportion Same Race is given by the proportion of civilians who are the same race as the focal officer in the Census Block Group in which the call was made. In columns 2-4, I report results from estimations of equation 8. The race of the officer and block group measured by “Race” is denoted in the row immediately above the column numbers. All interaction terms that use Priority are instrumented using the interaction of Call Taker Score and the other terms in the interaction. Additional interaction terms from the regressions are omitted. Each regression includes month-by-year, day of week-by-hour, officer, and Division fixed effects; an indicator for whether the call taker is Hispanic; Census Block Group controls for proportion no high school degree, proportion unemployed, and log per capita income; and officer experience. Standard errors are clustered at the call taker level. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 5: IV Results by Experience

Dependent Variable:	Arrest
Model:	(1)
Variables	
Priority	0.0203***
	(0.0057)
Years of Experience	0.0002
	(0.0005)
Priority * Years of Experience	-0.0008***
	(0.0003)
Fit statistics	
Observations	1,170,500
R^2	0.02204
Y mean	0.02862

Notes: This table reports results for a specification similar to equation 7 that uses Years of Experience as the interaction term instead of Proportion Same Race. Priority * Years of Experience is instrumented using Call Taker Score * Priority. Each regression includes month-by-year, day of week-by-hour, officer, and Division fixed effects; an indicator for whether the call taker is Hispanic; and Census Block Group controls for proportion no high school degree, proportion unemployed, proportion minority, and log per capita income Standard errors are clustered at the call taker level. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.