Mislearning from Selected Crime Data: Selection Neglect and Statistical Discrimination in Policing

Juan B. González Santiago De Martini Ervyn Norza Santiago M. Perez-Vincent *

January 26, 2026

(Most Recent Version Here)

PRELIMINARY AND INCOMPLETE DRAFT, PLEASE DO NOT DISTRIBUTE

Abstract

There is growing evidence that people often form beliefs from selected samples as if these were representative, a cognitive bias known as selection neglect. We study selection neglect in policing, a domain where officers routinely form crime predictions from selected crime data: only a subset of crime events becomes observable, and observability endogenously varies across locations and groups, with minorities being over-selected into crime data. If police officers form beliefs by taking selected data at face value rather than adjusting for the selection process, biased data can become biased crime predictions. We conduct a lab-in-the-field experiment with senior police officers in Colombia, eliciting crime predictions directly while exogenously varying the selection of crime data. We find that officers substantially neglect data selection, leading to large and economically meaningful prediction errors. A series of treatments show that neglect reflects representational failures rather than lack of information or computational capacity, so that providing information or unbiased advice does not eliminate neglect. Neglect maps selection bias into prediction bias: when selection is asymmetric across groups, neglect generates systematically biased crime beliefs; when data selection is endogenous to past predictions, neglect produces persistent and self-reinforcing belief distortions. Our results provide a micro-foundation for statistical discrimination in crime beliefs, demonstrating how biased predictions can arise mechanically from selected data even in the absence of preferences or stereotypes.

*González: Department of Economics, University of Southern California, juanbgon@usc.edu. De Martini: Department of Economics, University of Southern California, sd59576@usc.edu. Norza: National Police of Colombia, eh.norza10@uniandes.edu.co. Perez-Vincent: Inter American Development Bank, santiagoper@iadb.org. We are grateful to the Center for Effective Global Action (CEGA) for generously funding this study and to the Police Postgraduate School (ESPOL) in Bogotá for facilitating the experiments. We thank Felipe Gonçalves, Jonathan Libgober, Alejandro Martínez-Marquina, Paulina Oliva, Emanuel Vespa, and the participants of LAX at UCLA and BEEC at Caltech and UCSB for thoughtful feedback on the study.

I Introduction

Consider how people form beliefs about the political views of others. The opinions you observe are not a random sample of society, but a selected subset shaped by who your friends are, which accounts you follow on social media, and which media you consume. As your beliefs evolve, they can in turn influence this selection by changing who you engage with or what information you are exposed to, creating echo chambers. Similar selection processes arise whenever data are generated endogenously by past predictions rather than randomly sampled from the world. A growing experimental literature shows that people often fail to account for such selection when forming beliefs, treating observed data as if it were representative of the underlying environment (Brundage et al., 2024; DellaVigna & Kaplan, 2007; Enke, 2020). This cognitive failure, known as selection neglect, maps selection bias directly into belief bias. Despite its potential importance for inference, however, we still know little about how selection neglect operates among experts, what mechanisms sustain it, or how it affects data-driven institutions.

Policing provides a particularly revealing setting to study selection neglect because it increasingly relies on crime data to make crime predictions and guide enforcement (Brayne, 2020). At the same time, observed crime data is not a representative sample of the underlying crime. Observability differs systematically across groups and locations, and is often endogenous to past predictions and enforcement decisions (Chen et al., 2023; Gonçalves, Mello, & Weisburst, 2025). These features imply that the data used to infer crime patterns are both asymmetrically and endogenously generated. Importantly, learning from selected data is not a challenge unique to human decision-makers: a growing literature documents that predictive algorithms trained on crime data reproduce its selection bias (Arnold et al., 2021, 2025). Policing is thus a case of a more general inference problem: when predictions are formed from data that are asymmetrically selected across groups and selection is neglected, belief distortions can arise even in the absence of animus. In this sense, selection neglect provides a micro-foundation for statistical discrimination in policing. For these reasons, policing is not only a relevant policy domain, but also a canonical setting to study how practitioners form beliefs from endogenously selected data.

Despite the centrality of selected crime data in policing, direct evidence on how officers form beliefs from such data remains scarce. How police officers interpret crime data when forming beliefs is not directly observed in administrative records, which capture enforcement actions rather than the predictions that inform them. Identifying selection neglect from observational data would require exogenous variation in the selection of crime data that is unrelated to officers' behavior and to underlying crime, so that changes in observed decisions could be attributed to inference rather than to incentives, constraints, or strategic considerations. In practice, such variation is difficult to obtain. Selection into crime data is often shaped by policing itself through patrol allocation, stops, and arrests, and even when driven by citizen reporting, reporting behavior responds to police presence and trust (Ang et al., 2025). Moreover, changes in observability affect criminal behavior (Gonçalves, Jacome, & Weisburst, 2025), so shifts in selection rarely leave crime unchanged. As a result, although empirical research has documented selection bias in crime data and theory has linked it to statistical discrimination, direct evidence on whether and how police officers neglect data selection when forming crime beliefs remains elusive.

In this paper, we address this challenge using a lab-in-the-field experiment with senior officers of the National Police of Colombia. This setting is particularly well suited to study selection neglect in policing. Captains in the National Police are responsible for forming crime predictions based on *reported* crime data, this is, crime data generated through citizen reporting, to allocate police resources across space and time. Selection through citizen reporting has several advantages for studying belief formation: it is transparent to officers, plausibly exogenous to their short-run decisions, and reporting bias is well documented through victimization surveys (DANE, 2023). Officers are therefore used to working with selected data, and many are aware that reported crime is an imperfect measure of underlying crime, allowing us to study whether data selection is nevertheless neglected at the belief-formation stage.

Our design mirrors the core inference problem officers face in the field. Crime is selectively observed, reported crime is the primary data source, and reporting rates vary across locations. In each round of the Crime Prediction Task, officers observe the number of crime reports in two abstract neighborhoods and are incentivized to predict which neighborhood experienced more underlying crime (unobserved to participants) based on this selected data. Crucially, officers know the data-generating process and observe the reporting rate: the probability that a crime is reported in each neighborhood. As a result, higher reported crime does not mechanically imply higher true crime, and correct inference requires adjusting for observability. By exogenously varying reporting rates within-subjects while holding underlying crime constant, the design allows us to identify whether officers account for selection or instead take reported crime at face value.[1] The design isolates selection neglect in crime inference while abstracting from informational constraints —officers know the reporting rates of each area— and taste-based motives.

We find robust evidence that police officers neglect selection when making predictions from crime data. The experimental design generates two types of decisions: cases where the area with more reports is also the area with more crime, so the naïve and the selection-adjusted predictions are identical; and cases where adjusting for selection is necessary to identify the area with more crime. Because selection is exogenously varied, comparing prediction accuracy across these two cases isolates selection neglect from other inference errors. This comparison shows large and consequential mistakes. Officers make the correct prediction in 77% of decisions where adjusting for selection does not affect the optimal choice, but in only 41% of decisions where accounting for selection is required. These errors are not trivial: selection neglect leads officers to incur prediction errors equivalent to roughly 40% of the average crime gap between areas. The pattern is widespread across officers and is not driven by poor numeracy or confusion about the task. Moreover, because officers are systematically more likely to predict higher crime in areas where crime is more observable, these results show that selection neglect maps selection bias in crime data into biased crime predictions.

Failing to adjust for selection in crime data is consistent with two broad mechanisms. One possibility is that selection never enters the officer's mental model of the inference problem. If reported crime is treated as a complete measure of underlying crime, adjustment will never be attempted. We test this extensive-margin mechanism directly by allowing officers to choose which information to view before making a prediction. We find that 36% of officers choose not to view reporting rates when given the option, indicating that selection does not enter their inference problem at all. This behavior is not driven by confusion or inability. Officers who ignore reporting rates perform similarly to others in decisions where adjustment is not required, and neglect persists even among those who demonstrate full understanding of the task. Instead, extensive-margin neglect varies systematically with experience. Officers whose most

[1] For example, if neighborhood A receives more reports than neighborhood B but has a substantially higher reporting rate, accounting for selection reverses the correct prediction.

recent assignment involved analyzing reported crime data to make predictions and allocate patrols are significantly more likely to view reporting rates than officers in administrative roles. Together with the finding that experimentally increasing the salience of reporting rates does not reduce neglect, this pattern points to top-down mental models that exclude data selection, rather than bottom-up inattention, as a central driver of extensive-margin selection neglect.

Even when selection enters officers' inference problem, many fail to adjust correctly for it. This indicates that extensive-margin neglect cannot fully account for the observed errors. Conditional on recognizing that reported crime is selectively observed, two mechanisms could explain persistent neglect at this intensive margin. First, officers may hold incorrect mental models of how selection distorts reported crime data and therefore fail to apply the appropriate adjustment. Second, officers may understand the adjustment in principle but find it costly to implement, reverting to simpler heuristics when computation is effortful. The evidence suggests that incorrect mental models play a central role. Officers with better knowledge of reporting bias and lower confidence in the representativeness of reported crime data are substantially more likely to adjust correctly for selection, while those who view reported crime as largely reliable exhibit large accuracy losses even when selection is salient. At the same time, there is evidence that computation is costly: adjusting for selection increases response times, and many officers adjust only intermittently across selection-sensitive decisions despite performing well when adjustment is not required. Therefore, our evidence is consistent with both incorrect mental models and computational frictions playing a role in neglect.

Algorithmic advice provides a sharp diagnostic test of which of these two mechanisms drive the intensive margin of neglect. We randomly assign officers to receive algorithmic advice that either adjusts for selection or takes reported crime at face value, and inform officers of which algorithm is assigned to them. If the primary barrier were computation, recommendations that correct for selection should substantially reduce errors by eliminating the need to perform the adjustment. Instead, we find that neglecting officers frequently override correcting algorithmic recommendations precisely when those recommendations contradict their inference rule. In contrast, these officers are more likely to follow algorithms that validate their neglecting heuristic. This pattern indicates that selection neglect is a fundamentally representational problem. Officers do not ignore selection because it is unavailable or insufficiently salient, but because it does not enter their inference problem in the appropriate way. As a result, interventions based on information provision or unbiased advice may fail to change beliefs and can be overridden when they conflict with existing mental models.

Having established that selection neglect defines a mapping from selection bias in observed data to bias in crime predictions, we next examine how this inference error interacts with the structure of the selection environment. Crime data selection commonly exhibits two features: it is asymmetric across groups, with some groups overrepresented in observed crime data, and it is endogenously generated by enforcement decisions guided by that same data. If selection neglect remains stable across these environments, it can mechanically translate these features of data selection into belief distortions. To test this implication directly, we exogenously vary the selection environment while holding underlying crime constant. When selection is asymmetric across groups, neglect maps this asymmetry into systematically biased crime predictions: officers become significantly more likely to predict higher crime in the group where crime is more observable, even though true crime is identical across these abstract groups. When selection is endogenous to past predictions, this mapping generates dynamic feedback: initial prediction errors affect future observability, which in turn reinforces subsequent beliefs. As a result, officers exhibit persistent, self-reinforcing cycles of bias in crime predictions. Taken together, these findings show that selection neglect does not merely generate isolated inference errors, but maps the selection environment into systematic and persistent belief distortions. Our results therefore provide a micro-foundation for statistical discrimination in crime beliefs: biased predictions can arise mechanically from selected data, even in the absence of preferences or stereotypes.

This paper contributes to the literature on selection neglect by documenting its prevalence and mechanisms among expert practitioners. A growing experimental literature shows that individuals often fail to account for data selection in laboratory settings (Ali et al., 2021; Araujo et al., 2021; Barron et al., 2024; Enke, 2020; Esponda & Vespa, 2018; Jin et al., 2021). However, it remains unclear how this bias operates in professional contexts where decision-makers routinely work with selected data, with prior work finding inconsistent results (Koehler & Mercer, 2009; Malmendier & Shanthikumar, 2007). Using a lab-in-the-field experiment with senior police officers, we show that selection neglect persists in such a setting and generates large and economically meaningful prediction errors. Studying practitioners further allows us to leverage meaningful heterogeneity in experience with selected data and knowledge of the selected data generating process. This heterogeneity reveals that selection neglect primarily reflects representational failures rather than inattention or lack of effort, a distinction that is difficult to establish with standard experimental samples.

A second contribution of the paper is to provide direct belief-level evidence linking selected crime data to statistical discrimination in policing. A defining feature of policing is that not all crimes are observed, and crime data is not a representative sample of the underlying crime. A broad consensus shows that minorities are patrolled, stopped, searched, and arrested more often (Aggarwal et al., 2025; Ba et al., 2021; Chen et al., 2023; Feigenberg & Miller, 2025; Gonçalves & Mello, 2021; National Academies of Sciences, Engineering, and Medicine, 2023; Pierson et al., 2020; Vomfell & Stewart, 2021). Due to these disparities in crime observability across groups, crime data suffers from selection bias, with minorities overrepresented in crime data (Gonçalves, Mello, & Weisburst, 2025). While theoretical work has emphasized that inferences drawn from selected crime data can generate group-level disparities even absent animus (Hübert & Little, 2023), direct evidence on this link was missing. We show that selection neglect mechanically maps selection bias into biased crime predictions, even when groups are abstract, true crime is held constant, and no preferences or stereotypes are present. When crime data are asymmetrically selected across groups, officers systematically form asymmetric crime beliefs. These results establish a micro-foundation for statistical discrimination in crime beliefs, demonstrating that biased predictions can arise purely from a cognitive inference error operating on selected data, consistent with prior findings from blind algorithms (Arnold et al., 2025). This paper thus complements a growing body of work in policing that emphasizes the role of cognitive and emotional mechanisms in officer decision-making (Dube et al., 2024; Ferrazares, 2025; Holz et al., 2023; Owens et al., 2018).

Finally, our results highlight the representational nature of selection neglect. Officers don't neglect selection in crime data because this information is not available or because of computational issues, but because they fail to correctly represent the inference problem. As a result, redirecting officers attention or providing unbiased advice doesn't eliminate neglect. In this sense, our paper contributes to a growing literature showing the limits of algorithmic advice in other high-stakes settings, because experts like judges and physicians override algorithms when they conflict with their own inference rule (Angelova et al., 2025; Mullainathan & Obermeyer, 2022). More broadly, our findings contribute to an expanding literature on how expert decision makers form beliefs from data (Arnold et al., 2018; Bhuller & Sigstad, 2024; Feigenberg & Miller, 2025; Prendergast, 2021).

The paper proceeds as follows. Section II introduces the setting of policing in Colombia and our sample of police officers. Section III describes the experimental design and the empirical strategy. Section IV presents the results on the extent, relevance, and robustness of selection neglect in crime predictions. Section V explores the mechanisms driving neglect, using algorithmic advice to tell them apart. Section VI documents the implications of this inference error across selection environments. Section VII discusses related literature and concludes.

II Setting

A Data-Driven Policing in Colombia

Policing in Colombia is explicitly data-driven. Station commanders and operational officers routinely rely on administrative crime data to form beliefs about where and when crime is most likely to occur, and to respond accordingly through patrol allocation and other operational decisions. While objectives may differ across officers and contexts, accurate beliefs about the spatial and temporal distribution of crime are a necessary input to any policing strategy. The primary source of information the National Police of Colombia uses for these crime predictions is reported crime data. Internal monitoring systems aggregate citizen crime reports across locations and time, and officers are trained to interpret variation in reported crime as informative about underlying criminal activity. In practice, reported crime counts are the dominant metric used to summarize crime patterns within an officer's area of responsibility.

A central challenge, however, is that reported crime data is not a representative sample of crime. Crimes enter administrative records only if they are reported by citizens, and the probability that a crime is reported varies substantially across places, crime types, and populations. According to national victimization surveys, only around 31% of crimes in Colombia are ever reported to the police (DANE, 2023). Reporting rates are high for some offenses, such as homicide, but extremely low for others, such as extortion; they also differ markedly across urban and rural areas, and across the income distribution [2]. As a result, two locations with identical underlying crime levels can generate very different volumes of reported crime solely because crimes are observed with different probabilities.

This reporting-based selection creates a selection problem for belief formation. If officers treat reported crime counts as a representative measure of underlying crime, variation in reporting behavior will be mechanically confounded with variation in criminal activity. A rational decision-maker would therefore need to adjust reported crime data for the probability that crimes are reported in order to form accurate

[2] National victimization surveys uncover systematic variation in reporting rates. The ECSC survey by DANE (2023) documents that while 43% of crimes are reported in Bogotá, only 22% are in rural areas. Within Bogotá, citizens with higher socioeconomic status (strata V and VI) are twice as likely to report a crime than lower status ones (strata I and II). In terms of types of crime, while less than 25% of violence and extortion cases are reported, around half of vehicle thefts are reported, and most homicides are.

beliefs about true crime levels. One might argue that officers fail to adjust reported crime for selection simply because they lack sufficiently granular information about reporting behavior. While officers may not observe reporting rates at fine geographic levels, coarse information about selection is available in victimization surveys, and yet this knowledge is rarely operationalized in practice. Our experimental design addresses this concern directly by making reporting rates fully observable, allowing us to test whether officers adjust for selection when informational constraints are removed.

Importantly for our purposes, the selection problem of crime data is both transparent and plausibly exogenous to individual officers' short-run decisions. In the Colombian context, selection into crime data is driven primarily by citizen reporting behavior rather than by discretionary police actions such as stops or arrests [3]. While policing may affect reporting over longer horizons through trust and visibility, officers do not directly control whether specific crimes are reported at the moment the data are generated. These features make Colombia an unusually well-suited setting to study belief formation from selected data. Officers routinely rely on reported crime data to form beliefs; the selection process that generates this data is simple, salient, and understandable; and selection is not mechanically tied to officers' own enforcement actions in the short run. This combination allows us to isolate whether expert decision-makers correctly adjust for known data selection while abstracting from preferences, stereotypes, and institutional constraints that confound inference in observational settings.

B Sample

Our participants are Captains of the National Police of Colombia, the rank that typically serves as station commander and is directly responsible for interpreting crime data and forming crime predictions that guide patrol allocation. We focus on captains because they are the officers who routinely work with reported crime data in practice and whose beliefs about crime patterns are operationally relevant.

Logistics. — In September 2025, 202 captains of the National Police of Colombia were attending the Police Postgraduate School (ESPOL) in Bogotá, completing the mandatory training required for promotion to Major (NATO OF-4). These officers constitute the universe of captains eligible for promotion in that cohort and were instructed by superior officers to attend our experimental session as part of their training activities. Upon arrival, officers were invited to participate in a "Policing Science Study" and were offered an incentive of USD 5 to 20. Eighteen officers (8.9%) declined to participate before receiving any information about the experiment. Of the 184 officers who began the study, all but two completed it, leaving a final sample of 182 participants and no evidence of endogenous attrition.

Final sample. — The final sample consists of 182 captains with substantial professional experience. Officers have an average of 14 years of service in the National Police of Colombia (minimum 12 years), and all hold at least an undergraduate degree, a requirement for promotion to this rank; 15% also hold a master's degree. Importantly for our purposes, participants have extensive experience making operational decisions: 25% were serving as station commanders immediately prior to the training, and an additional 40% held operative positions involving the allocation of police resources. Because our sample consists of captains eligible for promotion, it is, if anything, positively selected in terms of ability and experience. In terms of demographics, 83% of participants are men, and ages range from 32 to 45 years (mean 36).

[3] See Gonçalves, Mello, and Weisburst (2025) for how officer behavior determines selection into crime data in the US context.

Overall, the sample comprises experienced practitioners who routinely interpret crime data in the field, making them a particularly appropriate population to study belief formation from selected data.

III Design

Our experimental design isolates the belief-formation problem that police officers face when interpreting crime data, abstracting from enforcement choices, institutional constraints, and incentives. In practice, policing decisions depend on a variety of considerations, including objectives, resources, and strategic interactions. Rather than attempting to replicate this complexity, our design focuses on a more fundamental input to those decisions: officers' beliefs about where crime is more likely to occur. We therefore study crime prediction as a cognitive task in its own right. In the experiment, officers are repeatedly asked to predict which of two locations experienced more crime based on available data. These predictions are not treated as enforcement decisions and do not carry operational consequences. Instead, they are interpreted as discrete elicited beliefs about relative crime levels, analogous to the implicit predictions officers form when reading crime reports in the field. This abstraction is deliberate. By separating belief formation from action, we are able to cleanly identify whether officers correctly account for data selection when forming predictions. Throughout the paper, we therefore interpret experimental choices as measures of crime prediction accuracy rather than as proxies for real-world policing behavior.

A Crime Prediction Task

We now describe the Crime Prediction Task, the core belief-elicitation environment in the experiment. Participants observe fictitious data for two cities $g \in \{A, B\}$, each consisting of a continuum of neighborhoods. In each neighborhood gn, a number of crimes c_{gn} occurs, which is unobservable to the participant.[4] In every round, officers observe a pair of neighborhoods (A_n, B_n) and are incentivized to predict which neighborhood experienced more crime. One decision is randomly selected for payment, and participants earn \$5 if their prediction is correct. Using binary predictions allows us to cleanly identify whether officers adjust reported data for selection or instead rely on reports at face value, without imposing an unfamiliar belief-elicitation method.

To form predictions about the latent crime level c_{gn}, participants observe crime data. Specifically, they see the number of reported crimes r_{gn} and the reporting rate s_{gn} of each neighborhood, defined as the probability that a crime is reported. Because crime observability is determined by reporting, reporting rates govern the selection of crime data. Two neighborhoods with identical crime levels can therefore generate different numbers of reports solely because of differences in reporting rates. In the first part of the experiment, participants also observe demographic characteristics $\mathbf{X}_{gn}$—average age, unemployment rate, and socio-economic status (SES).[5] These demographic variables serve two purposes. First, they act as decoys that reduce the salience of the reporting rate. Second, if officers rely on demographics in a setting where crime data are fully informative, their choices provide a measure of the extent to which non-crime information enters routine crime predictions.

[4]Cities serve as abstract labels for groups that could differ in their crime distributions (e.g., socio-economic status or other characteristics). By using abstract cities rather than socially salient groups, the design rules out taste-based discrimination and isolates inference from selected data.

[5]In Colombia, residential areas are classified into socio-economic strata (*estrato*) ranging from 1 (low) to 6 (high). This classification is used to set utility prices and target public services and is a familiar proxy for income for police officers.

Data-generating process. — Crime in each neighborhood is independently drawn from a city-specific distribution, $c_{gn} \sim \mathcal{N}(\mu_g, \sigma^2)$, with no prior information provided about the distributional parameters. Reporting rates s_{gn} are drawn independently from a uniform distribution $\mathcal{U}[0.1, 0.9]$ and rounded to the nearest multiple of five to ease computation. In principle, reported crimes follow $r_{gn} \sim \text{Bin}(c_{gn}, s_{gn})$, with $E[r_{gn}] = c_{gn}s_{gn}$. Reported crime is therefore a selected signal of underlying crime, and the optimal prediction adjusts for selection by scaling reports by the reporting rate, $\hat{c}_{gn} = r_{gn}/s_{gn}$. To isolate inference errors from statistical noise, we fix reported crimes to their expected value, setting $r_{gn} = c_{gn}s_{gn}$. Under this design, a decision-maker who correctly adjusts for selection always makes the correct prediction. Any deviation from the optimal rule therefore reflects cognitive error rather than sampling variation. Demographic variables $\mathbf{X}_{gn}$ are drawn independently from uniform distributions and are, by construction, uninformative about crime.

Information provided to participants. — Participants are informed that each neighborhood has an uncertain level of crime c_{gn} and that reported crime depends on reporting behavior. They are explicitly told that the reporting rate s_{gn} represents the probability that a crime is reported, so that in expectation $E[r_{gn}] = c_{gn}s_{gn}$.[6] To ensure comprehension, participants answer questions that require applying this logic (e.g., inferring the number of crimes from reported crimes and reporting rates). By making the selection process fully observable, the design rules out informational constraints about the data-generating process and allows us to test whether officers adjust for selection when the relevant information is available.

B Experiment Structure

The experiment proceeds in stages that progressively build on this core prediction task. We begin by measuring whether officers adjust for data selection when forming crime predictions in a baseline environment where the reporting process is fully observable and sampling noise is eliminated. We then vary the information environment to diagnose the mechanisms behind selection neglect, distinguishing failures of attention from failures of computation. To rule out simple explanations based on numeracy or misunderstanding of the task, we separately measure participants' basic arithmetic skills and comprehension of the data-generating process. Next, we introduce systematic differences in data selection across abstract groups to study whether selection neglect mechanically maps biased data into biased crime beliefs, even when true crime is identical across groups. We further examine how these distortions evolve when data selection becomes endogenous to past predictions, capturing a key feature of real-world data environments. Finally, we test whether a commonly proposed intervention —algorithmic recommendations that correct for selection— can mitigate inference errors, or whether selection neglect extends to the rejection of corrective advice. Throughout, we relate experimental behavior to officers' self-reported beliefs about the reliability of reported crime data in the field, linking laboratory inference errors to practitioners' real-world mental models.

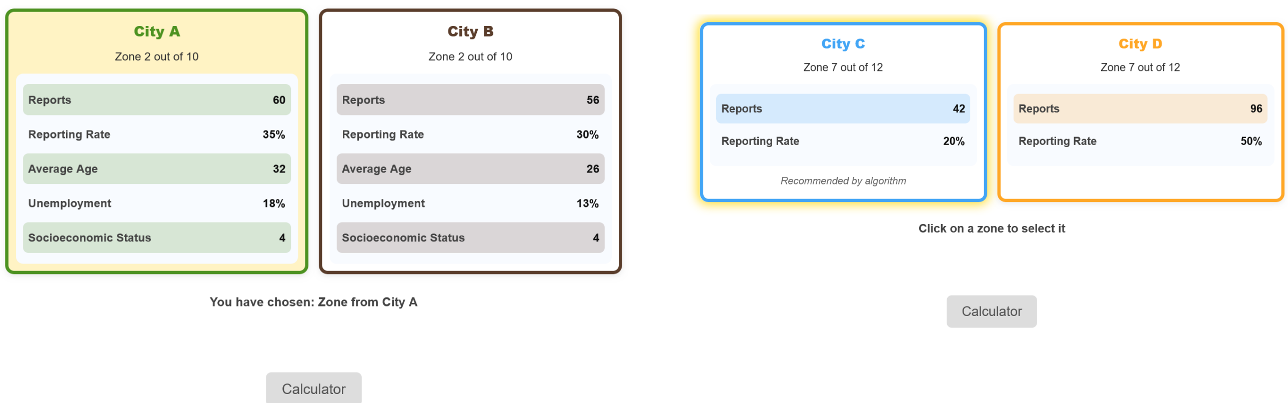
Introduction — After listening to the instructions, participants complete a brief training and answer comprehension questions designed to verify their understanding of the data-generating process, including the relationship between reported crime, reporting rates, and underlying crime. Incorrect answers are corrected before participants proceed. Participants also complete a small number of incentivized arithmetic questions involving the basic operations required to adjust reported crime for selection. These steps allow

[6] Instructions include examples such as: "Imagine that 200 crimes occurred in a neighborhood. If the reporting rate is 50%, we expect 100 crimes to be reported."

us to assess whether observed prediction errors reflect misunderstanding or limited numeracy, rather than inference failures, and are used only for diagnostic purposes in the analysis.

Figure 1: Crime Prediction Task Screens

(a) Part I - Low Selection Salience **(b)** Part II - High Salience and Algorithm



Notes: Panel (a) shows the decision screen for Part I, in an example where the participant has guessed the neighborhood from City A to have more crime. Note that in Part I, reports and reporting rates are accompanied by demographic decoys, and that they are always presented as the first data sources. This makes it more difficult to to neglect selection. Panel (b) shows the decision screen for Part II, in an example where the participant has not yet made a prediction but sees the recommendation of the algorithm as a golden halo around the zone and a small span in the bottom. Algorithm predictions, as detailed below, are only provided for the last six decisions of Part II and for a subset of participants. Without the recommendation, the screen is identical but without the halo and the text at bottom. Note the on-screen calculator available throughout the experiment.

Part I — The purpose of Part I is to elicit crime predictions in an environment where data selection is present but not salient. Officers repeatedly predict which of two neighborhoods experienced more crime while observing reported crime data alongside additional information that does not affect the optimal prediction. This setting mirrors real-world contexts in which reporting bias is known but embedded among multiple data sources rather than highlighted.

Part I consists of ten crime predictions across neighborhoods of cities A and B. In the first decision, officers observe only the number of reported crimes in each neighborhood, providing a baseline check that participants understand the task and interpret higher reports as indicative of higher crime. In the subsequent eight decisions, officers observe reported crimes, reporting rates, and demographic characteristics. The demographic variables —average age, unemployment, and socio-economic status— are intentionally uninformative about crime and serve to reduce the salience of the reporting rate. This allows us to observe whether officers incorporate data selection into their predictions when it is not made focal. In the final decision of Part I, officers choose which additional variable —out of the reporting rate and the demographic variables— to observe alongside reported crime before making their prediction. Selecting the reporting rate reveals that data selection enters the officer's mental model of the inference problem, while selecting any other variable indicates that selection is not treated as relevant for prediction. This choice provides a direct measure of whether officers consider selection when interpreting crime data in a low-salience environment.

Participants have 1 minute to make each prediction in Part I. To ease computations, officers can use a calculator on screen. We randomly select one prediction of Part I for payment, and pay $5 if officers correctly guessed the neighborhood with the higher number of crimes.

Part II — Part II builds on the core prediction task by increasing the salience of data selection and studying its implications in more structured environments. To avoid learning or carryover from Part I, all decisions in Part II involve neighborhoods drawn from two new cities, C and D. Officers make twelve crime predictions, each comparing a pair of neighborhoods from these cities. Throughout Part II, participants observe only the number of reported crimes and the reporting rate for each neighborhood; no demographic information is shown. By removing decoy information, this environment makes the selection information focal. As in Part I, participants have one minute to make each prediction, may use an on-screen calculator, and one decision is randomly selected for payment.

We first study whether systematic differences in data selection across groups translate into asymmetric crime predictions. Participants are randomly assigned to one of two selection environments. In the *Symmetric Selection* condition, reporting rates for cities C and D are independently drawn from the same distribution, as in Part I. In the *Asymmetric Selection* condition, one city has systematically higher reporting rates than the other: $s_{Cn} \sim \mathcal{U}[0.1, 0.6]$ and $s_{Dn} \sim \mathcal{U}[0.4, 0.9]$, or vice versa, generating a 30 percentage point reporting gap in expectation. True crime is drawn symmetrically across cities, so any persistent differences in predictions reflect the interaction between selection bias and officers' inference.

We then examine how selection neglect interacts with endogenous data generation. For the last six decisions of Part II, participants are further randomized into an *Exogenous Selection* or an *Endogenous Selection* condition. In the exogenous condition, reporting rates are drawn independently of past predictions. In the endogenous condition, officers' previous predictions affect future observability: if an officer predicts higher crime in city g in decision $n-1$, the reporting rate in that city increases by 15 percentage points in decision n, $s'_{gn} = s_{gn} + 0.15$. This captures a key feature of real-world data environments, where crime beliefs and enforcement influence which events are subsequently observed (Ang et al., 2025; Chen et al., 2023; Gonçalves, Jacome, & Weisburst, 2025).

Finally, we test whether algorithmic recommendations mitigate or amplify inference errors. For the last six decisions of Part II, participants are randomly assigned to one of three conditions: no algorithm (*Control*), a *Correcting Algorithm*, or a *Neglecting Algorithm*. Both algorithms use reported crime data to recommend which neighborhood likely experienced more crime, with some noise. The correcting algorithm adjusts for reporting rates, while the neglecting algorithm relies only on raw reports. Participants are informed of the type of algorithm they receive and may choose to follow or override its recommendation. This design allows us to test whether algorithmic correction compensates for selection neglect, or whether neglect extends to the rejection of corrective advice.

End Questions — After completing the experiment, participants answer a short, unincentivized survey. We ask officers to assess their own performance and whether they would like to receive feedback on how to improve their crime predictions. Crucially, officers report their beliefs about the reliability of reported crime data, including whether crime reports capture spatial and temporal crime patterns, whether they are the best available data, whether citizens know how to report crimes, and whether reported data is biased. These measures allow us to relate experimental inference errors to officers' real-world beliefs about crime data. Participants also report their years of experience in the National Police of Colombia and the nature of their most recent position (administrative, operative, or surveillance). Officers in surveillance positions are those acting as station commanders and directly assigning patrols. Finally, participants provide basic demographic information.

Table 1: Summary of Experimental Design

Part I	Ten crime predictions First prediction: only see number of reports, sanity check Eight predictions: see number of reports, reporting rate, demographics Tenth decision: see number of reports and 1 data of their choice
Part II	Twelve crime predictions Higher salience of selection: only see number of reports, reporting rate All predictions: systematic selection treatment Last six predictions: endogenous selection treatment and algorithm interventions
End	Performance guess, trust in reported data, experience, and demographics

Why an Experiment

Identifying selection neglect in policing using administrative data is difficult because the object of interest—how officers form beliefs from crime data—is not directly observed. Doing so would require exogenous variation in the selection of crime data that is unrelated to officer behavior and to true crime, so that changes in decisions could be cleanly attributed to inference rather than to incentives or constraints. In practice, such variation is unlikely to exist. Selection into crime data is often determined by policing itself through patrols, stops, and arrests, and even when driven by citizen reporting, reporting behavior responds to police presence and trust (Ang et al., 2025). Moreover, changes in observability typically affect criminal behavior, so shocks to selection do not leave crime unchanged (Gonçalves, Jacome, & Weisburst, 2025). As a result, observational data cannot disentangle how officers interpret selected crime data from the behavioral and institutional processes that generate it. These challenges motivate an experimental approach. By exogenously varying data selection, holding true crime constant, and eliciting crime predictions directly, the experiment isolates belief formation from enforcement behavior and allows us to identify selection neglect in crime prediction.

C Empirical Strategy

Our analysis follows directly from the experimental design. For each officer and decision, we observe which neighborhood gn they predict to have higher crime. For every decision, there is a unique accurate prediction: choosing the neighborhood with the higher number of crimes c_{gn}. Although officers do not observe c_{gn}, the optimal prediction rule —ranking neighborhoods by adjusted reports $\max_g r_{gn}/s_{gn}$— is always correct because we eliminate sampling noise from the reporting process. The optimal prediction therefore does not depend on priors or assumptions about the underlying crime distributions. Throughout the analysis, the unit of observation is the officer–decision pair. Because predictions are binary, we summarize decision-relevant information using gaps between neighborhoods, $\Delta(x_n) \equiv x_{An} - x_{Bn}$.

Our empirical strategy focuses on whether officers adjust for or neglect data selection when forming crime predictions. In this setting, neglecting selection corresponds to ranking neighborhoods by reported crimes r_{gn} rather than by selection-adjusted reports r_{gn}/s_{gn}. Because reporting rates are independently drawn across neighborhoods, ignoring selection does not always change the optimal ranking. In *selection-neutral decisions* the neighborhood with more reports is also the neighborhood with more crimes. In these cases, even an officer who ignores reporting rates makes the correct prediction. We therefore use selection-neutral decisions as a benchmark that captures inference errors unrelated to selection neglect. In contrast, in *selection-sensitive decisions* adjusting for selection reverses the ranking of neighborhoods. In these cases, neglecting selection mechanically leads to an incorrect prediction. Selection-neutral and selection-sensitive decisions are otherwise comparable in structure; they differ only in whether accounting for selection affects the optimal choice.[7] The difference in prediction accuracy between selection-neutral and selection-sensitive decisions therefore identifies prediction errors attributable to selection neglect.

While accuracy provides a natural measure of inference quality, it abstracts from the magnitude of officers' mistakes. If officers were only neglecting selection in cases where the underlying crime gap is minimal, a large accuracy loss would not necessarily imply large prediction errors. To capture the relevance of selection neglect, we define *neglected crimes* as the difference in crimes between the optimal and the actual guess. This measure quantifies how much crime is mispredicted due to selection neglect, not merely whether a prediction is correct.

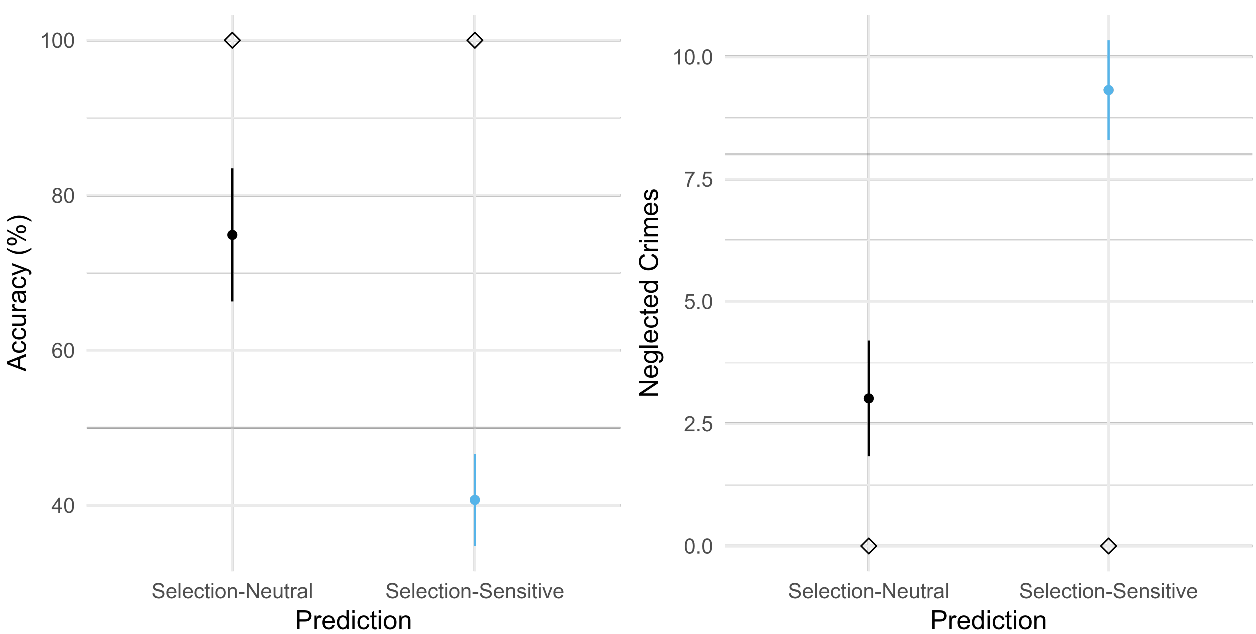
[7]Formally, we denote selection-neutral decisions those that satisfy $\Delta(r_n)\Delta(r_n/s_n) > 0$, while selection-sensitive decisions satisfy $\Delta(r_n)\Delta(r_n/s_n) < 0$. This distinction is not apparent in the data participants observe unless they do the calculations to adjust for selection. For example, if $r_A = 30$ and $r_B = 60$, then with reporting rates $s_A = 21\%$ and $s_B = 40\%$ both raw and adjusted reports imply higher crime in B, making this a selection-neutral decision. With $s_A = 19\%$ and $s_B = 40\%$, raw reports still favor B, but adjusted reports imply higher crime in A, making this a selection-sensitive decision.

IV Selection Neglect in Crime Predictions

We begin by quantifying how much selection neglect reduces the accuracy of crime predictions. Panel (a) of Figure 2 shows that in decisions where neglecting selection does not change the optimal prediction (what we denote *selection-neutral* decisions), officers correctly predict crime in 77% of cases. In contrast, accuracy falls to 41% when the correct prediction requires adjusting for selection (*selection-sensitive* decisions). Table A.1 confirms this 36 percentage point drop in accuracy is significant ($p < 0.001$). Furthermore, these average effects do not reflect a small number of extreme decision-makers: prediction errors due to selection are widespread across officers, a pattern we document in the next section.

Figure 2: Accuracy Loss Due to Selection Neglect

(a) Selection and Accuracy **(b)** Selection and Neglected Crimes



Notes: Panel (a) shows the average prediction accuracy, by prediction type. Panel (b) shows the average number of neglected crimes —the difference in crimes between the optimal prediction and the actual one, by prediction type. Neglected crimes can be interpreted as the average crime gap multiplied by the average inaccuracy rate. Both panels divide the data by whether adjusting for selection was needed to make the accurate prediction. Selection-neutral predictions, where adjusting for selection doesn't change the optimal prediction, are depicted in black and in the left column; selection-sensitive predictions, where adjusting for selection does change the optimal prediction, are depicted in blue and in the right column. Error bars in both panels represent the 95% confidence interval of the mean, with standard errors clustered at the officer level. The horizontal gray lines represent the expected accuracy and neglected crimes of a random decision-maker.

We next quantify the consequences of selection neglect in crime predictions. We examine how many crimes officers miss relative to the optimal prediction, this is, the crime gap between neighborhoods in inaccurate predictions. *Neglected crimes* translate the accuracy loss documented above into a metric directly relevant for policing: how much crime goes unaccounted for because of inference error. While this mapping might seem mechanical, it crucially depends on whether officers just neglect selection in close calls, this is, decisions with a small crime gap between neighborhoods. If this was the case, a significant loss in accuracy would not translate into a relevant increase in neglected crimes. The evidence shows the opposite: selection neglect produces materially worse crime predictions.

Panel (b) of Figure 2 finds that, while officers only neglect 3 crimes on average in predictions where adjusting for selection is not needed, this number rises to 9.4 crimes in selection-sensitive decisions. On average, officers miss 6.4 additional crimes ($p < 0.001$) due to selection neglect, representing roughly 40% the average crime gap. Figure A.1 provides further insight on why neglected crimes increase due to selection: while selection-neutral predictions become easier as the underlying crime gap increases, accuracy remains flat in selection-sensitive predictions. As a result, the accuracy loss due to selection neglect is highest in precisely the situations where prediction errors carry the highest cost in terms of neglected crimes.

Beyond reducing accuracy, selection neglect also systematically distorts crime beliefs. When reporting rates differ, officers who neglect selection may interpret higher observability as higher crime. As a result, differences in data selection can generate systematic asymmetries in crime predictions even when true crime is the same. We test this implication by examining how officers' predictions respond to differences in reporting rates, conditional on the true crime gap. Table A.3 shows that officers are significantly more likely to predict higher crime in neighborhoods with higher reporting rates: a 1 percentage point increase in the reporting-rate gap increases the probability of predicting higher crime in that neighborhood by 0.3 percentage points. This pattern holds even when true crime is held constant, persists across parts of the experiment, and is robust to including officer fixed effects. These results show that selection neglect mechanically maps selection bias in crime data into biased crime predictions.

Neglecting the selection of crime data is not explained by confusion about the task or by limited math skills. Table A.4 finds that selection neglect is robust among officers who answered all nine instruction comprehension questions correctly and those who correctly answered all math questions. Similarly, ignoring selection causes an increase in neglected crimes for both groups. Although there is some evidence that officers who lack the necessary math skills show more neglect, officers who have these skills still fail to adjust for selection. In short, selection neglect arises despite understanding the task and having the basic arithmetic needed to solve it.

The evidence so far shows that selection neglect is both large and costly. We now ask why officers neglect selection: does selection fail to enter the prediction process in the first place, or does the adjustment break down once it comes to mind?

V Mechanisms

A Framework

We now delineate a simple conceptual framework to build intuition about the possible mechanisms of selection neglect. To correctly adjust for selection in our setting, a decision maker must (i) recognize data selection as a relevant feature of the inference problem and (ii) understand the selection process to correctly incorporate it into predictions. We refer to failures at these two margins as extensive and intensive selection neglect. These margins are conceptually distinct and empirically separable in our design, though they need not operate sequentially.

At the extensive margin, selection neglect arises when data selection does not enter the decision maker's mental model of the inference problem. In our setting, crime prediction depends on two features: the number of reported crimes and the process through which these reports are generated, captured by reporting rates. If selection is not represented as a relevant feature, officers may rely on reported crime counts alone, effectively neglecting selection. Prior work highlights two channels through which this can occur. First, salience may draw attention to some features of the problem and not others (Bordalo et al., 2023). Second, mental models determine which variables are treated as relevant for inference, so a model that does not incorporate selection will mechanically generate neglect (Jehiel, 2018; Schwartzstein, 2014).

At the intensive margin, selection neglect arises when selection is represented as relevant but is not correctly incorporated into predictions. Even conditional on recognizing that reporting rates matter, a decision maker must understand the selection process well enough to apply the appropriate adjustment. In our experiment, this adjustment rule is intentionally simple and follows directly from the data-generating process: dividing reported crimes by the reporting rate yields the correct prediction.[8] Prior work shows that individuals may fail to undo selection even when they are aware of it, either because they misunderstand how selection distorts the data or because the adjustment is cognitively demanding (Esponda & Vespa, 2018; Farina & Herman, 2025; Jin et al., 2021).

Our experimental design allows us to test failures at both margins of selection neglect. To study the extensive margin, we vary the salience of data selection and directly elicit whether officers treat reporting rates as relevant information for crime prediction. To study the intensive margin, we examine whether officers who recognize selection correctly apply the adjustment, and how this varies with their understanding of the data-generating process and with computational frictions. Finally, we use algorithmic recommendations to further distinguish these channels: because following a correcting algorithm eliminates computational costs, such recommendations should reduce neglect if failures are primarily computational, but will be overridden if selection is not incorporated into officers' mental models. Together, these design elements allow us to separate failures of representation from failures of implementation in crime prediction.

B Evidence on Two Margins of Selection Neglect

We begin by examining failures at the extensive margin: whether data selection enters officers' mental models of the crime prediction problem at all. In the final decision of Part I, officers choose which additional variable to observe alongside reported crime. Selecting the reporting rate indicates that selection

[8] Because reported crimes are fixed at their expected value, this adjustment is not only optimal but accurate.

is treated as relevant for prediction. Figure A.4 finds that 36% of officers choose not to view the reporting rate, even though the data-generating process is transparent and the reporting rate is directly informative about underlying crime. This pattern is not explained by confusion or lack of basic numeracy, but varies systematically with officers' professional experience. Officers whose most recent assignments involved surveillance and patrol planning (tasks that rely heavily on interpreting selected crime data) are 20 percentage points ($p = 0.035$) more likely to incorporate reporting rates than administrative officers. These results show that a substantial share of officers exclude data selection from their prediction models, and that this exclusion is linked to differences in experience with selected data.

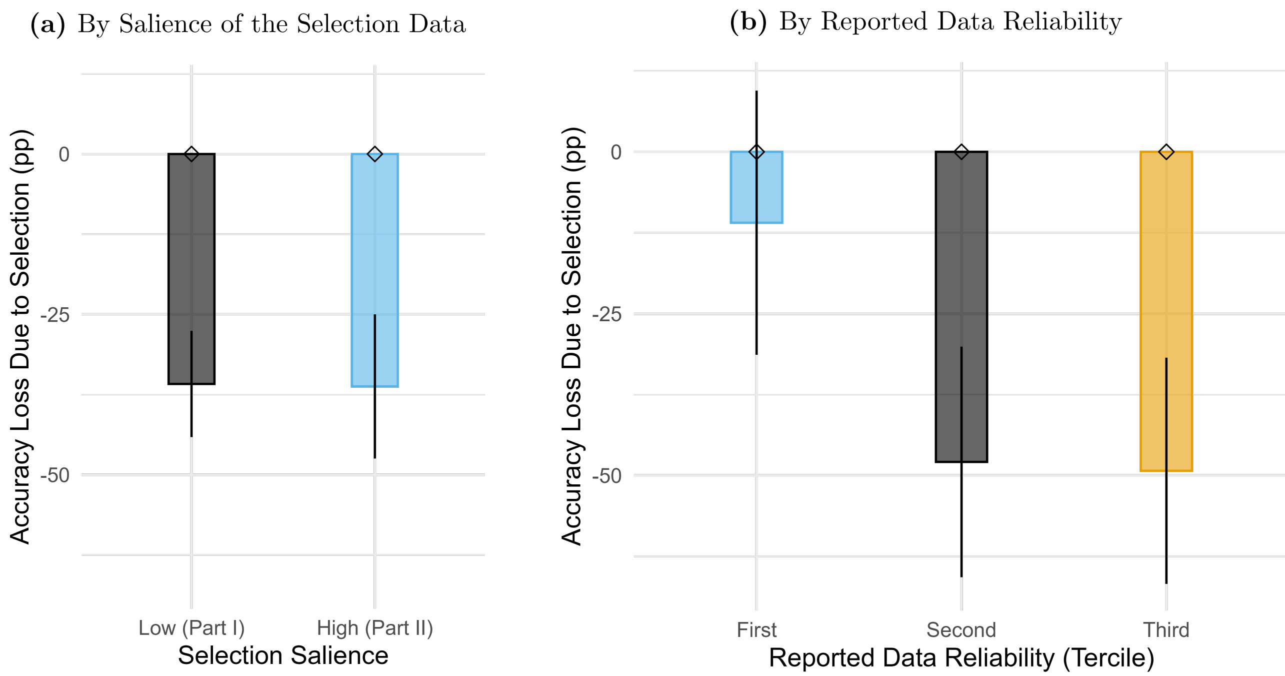
We next examine whether extensive-margin neglect reflects bottom-up inattention to selection information or top-down mental models that do not incorporate selection. If neglect were primarily driven by inattention, making selection more salient should reduce errors. To test this, we compare predictions across environments with low and high selection salience. In Part I, reporting rates and crime reports are shown alongside multiple uninformative variables; in Part II, these decoys are removed, making selection visually unavoidable. Despite this increase in salience, Figure 3 shows that prediction accuracy in selection-sensitive decisions does not improve. Selection neglect persists even when officers are forced to attend to reporting rates. Together with the heterogeneity in information choice, this evidence suggests that extensive-margin neglect is not primarily a salience or reminder problem. Selection fails to enter predictions not because it is unseen, but because it is not represented as a relevant feature of the inference problem.

Failures at the extensive margin, however, cannot fully account for the magnitude of selection neglect we observe. A.1 documents that even among officers who choose to view reporting rates, and in environments where reporting rates are made salient and unavoidable, prediction accuracy remains substantially lower in selection-sensitive decisions. This indicates that selection neglect persists even when selection enters the inference problem, pointing to failures at the intensive margin. Conditional on recognizing that reporting rates matter, there are two broad reasons why officers may still fail to adjust correctly. First, they may hold an inaccurate understanding of how selection distorts reported crime data and how to undo it. Second, even with the correct mental model, applying the adjustment may be cognitively costly, leading officers to rely on simpler heuristics when computation is effortful.

We first examine whether failures at the intensive margin reflect differences in officers' understanding of how selection distorts reported crime data. Even conditional on recognizing that reporting rates matter, officers may hold inaccurate mental models of the selection process and therefore fail to apply the correct adjustment. To assess this channel, we study heterogeneity in selection neglect by officers' beliefs about the reliability of reported crime data. At the end of the experiment, officers report their agreement with statements capturing whether crime reports accurately reflect spatial and temporal crime patterns, whether citizens know how and where to report crimes, and whether reported data provide a biased picture of underlying crime. We aggregate these responses into an index of *reported data reliability*, where higher values reflect stronger beliefs that reported crime data faithfully represent true crime [9]. Panel (b) of Figure 3 shows a stark gradient: officers who are most aware of biases in reported data exhibit little to no selection neglect ($p = 0.298$), while those who view reported crime as largely reliable display large accuracy losses when adjustment is required — 48.8 percentage points ($p < 0.001$). This pattern suggests

[9] We follow Kling et al. (2007) and aggregate z-scores of each component of the index.

Figure 3: Heterogeneity in Selection Neglect



Notes: Panel (a) shows the accuracy loss due to selection —the estimated coefficient of selection-sensitivity on accuracy, disaggregated by selection salience. The left bar plots the effect for Part I, with low selection salience, and the right bar plots it for Part II, with higher selection salience. Error bars show the 95% confidence interval of the coefficient. Panel (b) shows the accuracy loss due to selection, disaggregated by reported data reliability. This panel only uses prediction from Part II. We plot the coefficient of selection-sensitivity estimated separately by terciles of reported data reliability. Error bars show the 95% confidence interval of the coefficient. Diamonds show the optimal benchmark in both panels.

that, beyond simply attending to selection, correctly adjusting for it requires a deeper understanding of how reporting behavior shapes observed crime data.

A second reason for intensive-margin failures is that, even with the correct mental model, implementing the selection adjustment may be cognitively costly. Several pieces of evidence are consistent with this interpretation. First, making accurate predictions in selection-sensitive decisions entails a significant time cost: A.9 finds that response times are substantially higher when adjustment is required, but no such cost appears in selection-neutral decisions. This pattern suggests that errors are not driven by general inattention or low effort, but by the additional cognitive demands of selection-adjustment. Second, Figure A.2 shows many officers adjust only intermittently. While a majority exhibit lower accuracy in selection-sensitive decisions, they do not consistently make errors across such decisions. Importantly, these same officers maintain high accuracy in selection-neutral predictions, ruling out explanations based on random responding or disengagement. Instead, the evidence is consistent with selective failures to implement the adjustment when its cognitive cost is high.

A final piece of evidence comes from the use of the on-screen calculator. Calculator use can only improve predictions for officers who both recognize the relevance of selection and understand how to correct for it; it does not supply the adjustment rule itself. Consistent with this logic, Table A.1 finds officers who use the calculator perform similarly to non-users in selection-neutral decisions ($p = 0.983$)

but are 36 percentage points more accurate in selection-sensitive ones ($p < 0.001$). Among calculator users, the accuracy loss due to selection neglect is sharply reduced and no longer statistically detectable ($p = 0.438$). Together, these results highlight the interaction between the extensive and intensive margins. Computational frictions matter, but only conditional on selection entering the mental model and being understood. When both representation and computation are in place, selection neglect largely disappears; when either margin fails, errors persist.

Selection neglect in policing is not a single error. Whether officers adjust depends first on whether data selection enters their mental model of crime data (extensive margin) and then on whether this mental model leads to the correct adjustment (intensive margin). This dual structure explains why some officers never adjust, why understanding of the selection process is linked to less neglect, why many adjust only intermittently, why adjustment depends on cognitive load, and why selection neglect disappears when both understanding and computation are in place.

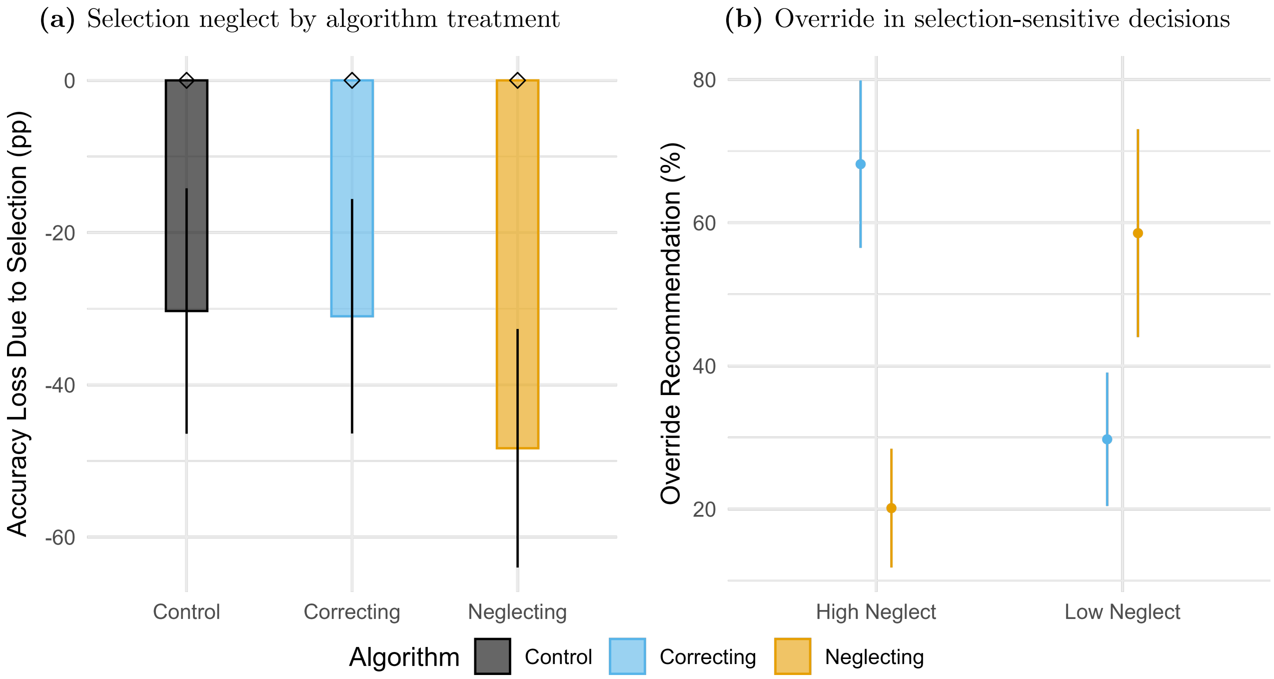
C Selection Neglect and Advice Override

We next use algorithmic recommendations as a diagnostic test of the mechanism underlying selection neglect. If neglect reflected primarily computational frictions, providing a recommendation that already adjusts for selection should substantially improve predictions by removing the need to perform the adjustment. In contrast, if neglect reflects an inaccurate mental model of how reported crime data should be interpreted, officers may reject or override corrective recommendations when they conflict with their (potentially wrong) inference rule. Comparing adherence to algorithms that do and do not adjust for selection allows us to distinguish between these explanations.

To implement this test, officers are randomly assigned to receive either no algorithmic recommendation (*control*), a *correcting* algorithm that adjusts reported crime for reporting rates, or a *neglecting* algorithm that relies on raw reports. In both algorithm conditions, recommendations include some noise, so that even the correcting algorithm's recommendations are highly informative but not mechanically optimal in every decision. In all cases, officers retain full discretion to follow or override the recommendation, and are informed of whether the algorithm accounts for data selection. If officers' inference model does not incorporate selection, they may override the *correcting* algorithm when its recommendation contradicts their neglecting rule, while following the *neglecting* algorithm because it validates that rule. In contrast, if selection neglect reflects primarily computational frictions, officers should follow the *correcting* algorithm, using it as a substitute for performing the adjustment, while overriding the *neglecting* one.

Panel (a) of Figure 4 shows that algorithmic recommendations do not reduce selection neglect. As expected, the neglecting algorithm increases neglect, but the correcting one doesn't change performance with respect to the control group. To understand why a correcting algorithm fails to improve performance, Panel (b) examines officers' adherence to algorithmic advice. We classify officers based on their accuracy in selection-sensitive decisions during the first six rounds of Part II, before algorithms are introduced. Officers with high levels of selection neglect override the *correcting* algorithm in 69% of selection-sensitive decisions, but override the *neglecting* algorithm only 18% of the time. Officers with low levels of neglect display the opposite pattern, frequently overriding the neglecting algorithm while following the correcting one. Thus, selection neglect extends to ignoring algorithmic recommendations: officers selectively adhere to algorithmic recommendations when they align with their inference rule and override them when they

Figure 4: Selection neglect and adherence to algorithm recommendations



Notes: Panel (a) shows the accuracy loss due to selection —the estimated coefficient of selection-sensitivity on accuracy, disaggregated by algorithm intervention. Error bars show the 95% confidence interval of the coefficient, with standard errors clustered at the officer level. Panel (b) shows the share of selection-sensitive decisions in which officers override the algorithm recommendation, dividing the sample in half according to their level of neglect in the first six decisions of Part II. The left column shows officers with *High Neglect* —below-median accuracy (< 33%) in selection-sensitive decisions, while the right column shows officers with *Low Neglect* —above-median accuracy. Both panels only use the last six decisions of Part II. Colors show the algorithm treatments: *Control* in black, *Correcting* in blue, and *Neglecting* in orange.

conflict with it.

Panel (a) of Figure 4 finds that algorithms do *not* reduce selection neglect. Both algorithms improve accuracy in selection-neutral decisions, but neither significantly increases accuracy in decisions where adjusting for selection is critical. Why does an algorithm that literally points to the correct choice not improve performance? Panel (b) suggests the answer: selection neglect extends to ignoring algorithmic recommendations. We divide the sample in half, above and below the median accuracy (33%) for selection-sensitive cases among the first six decisions of Part II (before the algorithms are introduced). Officers with high levels of selection neglect override the algorithm that corrects their heuristic in 68% of selection-sensitive decisions, yet override the neglecting algorithm only 20% of the time. Officers with low neglect show the opposite pattern: overriding the neglecting algorithm while following the correcting one. In other words, officers selectively trust algorithms when the recommendation *validates* their inference rule and override them when the recommendation *contradicts* it.

These findings reinforce the two-margin structure of selection neglect. Algorithmic recommendations eliminate the need to perform the selection adjustment, but they do not supply the missing inference rule. Officers who neglect selection override corrective advice precisely when it conflicts with their mental model, while following advice that validates it, paralleling findings from judicial settings in which judges override risk-prediction algorithms and frequently make lower-quality decisions as a result (Angelova et al., 2025). Selection neglect therefore extends beyond independent inference to the interpretation of external information: advice is filtered through existing inference rules rather than substituting for them. This pattern is consistent with neglect arising from representational failures rather than from computational difficulty alone.

VI Implications of Neglect Across Selection Environments

We have shown that selection neglect defines a stable mapping from the structure of data selection to crime predictions. If this mapping is stable, then features that commonly characterize real-world crime data should translate mechanically into biased crime beliefs. In particular, crime data are rarely selected at random: reporting rates differ systematically across locations and groups, and observability is often endogenous to past predictions and enforcement. Because these selection structures are central to policing and more familiar to officers than purely random variation, it is an empirical question whether the extent of selection neglect remains stable across these environments. In this section, we test this directly by exogenously varying the selection process and examining how different selection structures affect crime predictions.

A Asymmetric Selection Across Groups

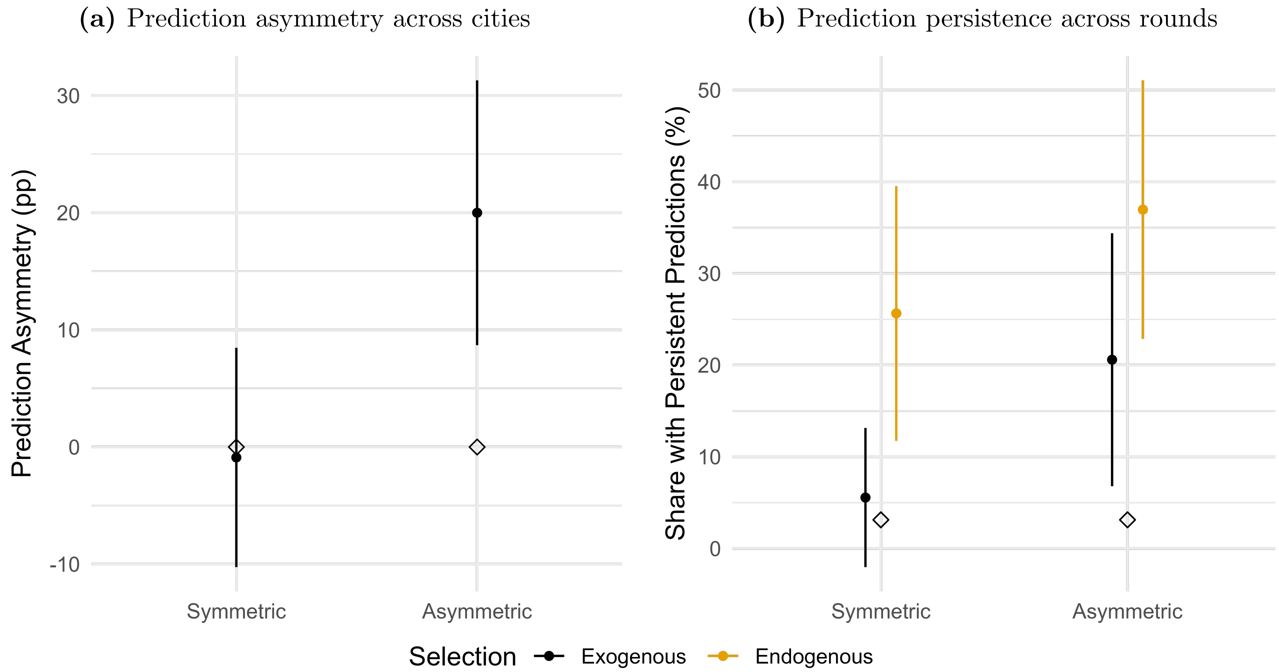
So far, we have exploited random variation in selection rates within groups to show that selection neglect maps differences in observability into prediction errors. In practice, however, selection into crime data is rarely random: reporting and observability are asymmetric across locations and groups. In the United States, Black and Hispanic individuals are overrepresented in crime data because they are more likely to be patrolled, stopped, and arrested (Ba et al., 2021; Gonçalves, Mello, & Weisburst, 2025). In Colombia, crime reporting rates are asymmetric across areas and income groups (DANE, 2023). These features imply persistent asymmetries in which crimes enter administrative data. If the selection-neglect mapping identified earlier is stable in such environments, asymmetric selection across groups should translate mechanically into asymmetric crime predictions, even when underlying crime was similar across groups.

We test this implication by experimentally varying the selection structure across abstract groups. In the experiment, we represent groups as two abstract *cities*, each consisting of a set of neighborhoods. The use of abstract groups rules out preferences, stereotypes, or priors about group-level crime. In the baseline environment, selection is symmetric across cities: reporting rates for neighborhoods in both cities are independently drawn from the same distribution, so the expected reporting-rate gap is zero. In contrast, in the *Asymmetric Selection* treatment implemented in Part II, reporting rates for the two cities are drawn from different distributions, generating an expected reporting-rate gap of 30 percentage points. Importantly, underlying crime distributions remain identical across cities. Under this design, systematic prediction asymmetries between cities can only be attributed to selection neglect.

Table A.7 documents that officers still neglect selection when it is asymmetric across groups. Because neglect remains stable in this environment, Figure 5 shows that asymmetric selection translates into asymmetric crime predictions. Panel (a) indicates that officers in the Systematic Selection Bias treatment are 20 percentage points ($p = 0.006$) more likely to predict higher crime in the over-selected city than in the under-selected one, despite identical true crime levels. No such asymmetry appears in the uniform-selection environment. Panel (b) shows that these asymmetric predictions persist across repeated decisions: 20.6% of officers in the systematic-selection condition predict the same city to have more crime in all six decisions, compared to 5.5% in the uniform-selection condition, a rate consistent with chance. Table A.8 confirms that these differences are statistically significant.

Overall, these results show that when selection is asymmetric across groups, as it is in crime data, selection neglect maps it into persistent asymmetries in crime predictions, even in the absence of group-specific preferences, incentives, or differences in true crime.

Figure 5: Asymmetry and persistence in crime predictions across selection environments



Notes: Panel (a) shows the *prediction asymmetry* across cities: how much more likely the average participant is to predict more crime in the city with higher average reporting rate, in percentage points. The left column shows officers in the *Uniform Selection Bias* treatment, where reporting rates for both cities are equal in expectation. The right column shows officers in the *Systematic Selection Bias* treatment, where reporting rates are 30 percentage points higher in expectation for one city. The figure only includes the first six predictions of Part II. Error bars show the 95% confidence interval of the means, with standard errors clustered by officer. The diamonds show the optimal benchmark: because both cities have the same number of crimes in expectation, the optimal prediction is perfectly symmetric. The panel documents that when data selection is systematically biased towards one group, officers become significantly more likely to predict higher crime within that group. Panel (b) shows the share of officers who make *persistent predictions*, this is, consistently predict the same city to have higher crime across the last six decisions of Part II. This share is disaggregated across selection environments. The left column shows officers in the *Uniform Selection Bias* treatment, while the right column shows officers in the *Systematic Selection Bias* treatment. Colors represent the endogeneity treatments. In black, we show officers in the *Exogenous Selection* group, where crime predictions don't affect reporting rates, and in orange the *Endogenous Selection* treatment, where predicting higher crime in one city increases reporting rates by 15 percentage points in that same city for the next decision. Error bars show the 95% confidence interval of the means, with standard errors clustered by officer. The diamonds show the optimal benchmark: because crimes are equally distributed across cities, the chance that it's optimal to guess the same city in 6 consecutive decisions is $2 \times 0.5^6 = 3.1\%$.

B Endogenous Selection

We next examine how selection neglect operates when data selection is endogenous to past predictions. Thus far, we have shown that officers neglect data selection when observability is independent of their crime predictions. In many policing settings, however, predictions also influence which events become observable in the future. For instance, crime predictions guide patrol allocation, which affects crime observability, and enforcement decisions shape trust in the police and thus citizens' willingness to report crimes (Ang et al., 2025; Chen et al., 2023). When predictions affect selection, neglecting selection can generate self-reinforcing cycles of bias: dynamic feedback loops in which predictions shape data selection and neglected selection feeds back into subsequent predictions. In earlier sections, we established one side of this cycle: when selection is exogenous, neglect mechanically maps it into biased crime predictions. In this section, we close the loop by experimentally varying whether predictions affect selection, testing whether endogenous selection generates cycles of persistent predictions.

To test this implication of neglect, we extend the crime prediction task to an environment in which data selection is endogenous to past predictions. In the *Endogenous Selection* treatment, officers' predictions affect future observability: predicting higher crime in a given city increases its reporting rate in the subsequent decision by 15 percentage points. In the control *Exogenous Selection* condition, implemented throughout the experiment, reporting rates are drawn independently of past predictions. Importantly, the underlying crime distribution remains unchanged. If selection neglect is stable across these environments, endogenous selection can transform prediction errors into persistent and self-reinforcing belief cycles relative to the exogenous benchmark.

The results are consistent with endogenous selection generating self-reinforcing cycles of bias. First, endogenous selection increases short-run prediction persistence. When selection is exogenous, officers are as likely to predict higher crime in the same city in consecutive decisions as to switch cities. In contrast, under endogenous selection, officers become 35.6 percentage points ($p = 0.027$) more likely to predict the same city rather than switch. This increase reflects a mechanical feedback: once a city is predicted to have higher crime, higher subsequent observability reinforces that belief under selection neglect. Second, Figure 5 finds that this short-run persistence translates into longer-run persistent predictions. Under endogenous selection, officers are 21 percentage points ($p = 0.065$) more likely to repeatedly predict the same city across all six decisions, generating persistent crime predictions even though underlying crime are identical across cities. This pattern is strongest when selection is also asymmetric across groups, when 37% of officers persistently predict higher crime in the same city, indicating that asymmetric selection interacts with feedback in data generation to sustain cycles of bias in beliefs over time.

Taken together, these findings show that selection neglect defines a mapping from the selection environment to crime predictions and that this mapping is stable across selection environments common in crime data. When selection is asymmetric across groups, neglect translates that asymmetry into systematically biased crime predictions. When predictions affect data selection, neglect generates self-reinforcing cycles of bias in beliefs. Importantly, we find no evidence that the extent of selection neglect itself varies across these environments; instead, different selection environments determine the implications of this persistent inference error. These findings highlight why selection neglect is particularly consequential in policing contexts and motivate a broader discussion of its implications for data-driven decision-making.

VII Discussion

This paper studies how decision-makers form beliefs from selectively generated data, and documents how selection neglect distorts crime predictions even when the data-generating process is transparent. Using a lab-in-the-field experiment with police officers, we show that selection neglect is prevalent among experienced practitioners and leads to large and economically meaningful prediction errors. We identify two distinct margins underlying this bias: many officers fail to represent data selection as a relevant feature of the inference problem, while others recognize selection but fail to correctly implement the adjustment. Importantly, this inference error is stable across data environments and mechanically maps features of crime data, such as asymmetric and endogenous selection, into systematically biased and persistent crime beliefs. Together, these findings highlight how a cognitive bias in belief formation can have first-order implications in data-driven systems such as policing.

Our findings build on a growing experimental literature documenting selection neglect in lab experiments (Ali et al., 2021; Araujo et al., 2021; Barron et al., 2024; Enke, 2020; Esponda & Vespa, 2018; Farina & Herman, 2025; Jin et al., 2021). Studying experienced practitioners allows us to extend this work in two important ways. First, we show that selection neglect persists in a professional setting where participants routinely work with selected data and have strong incentives to form accurate beliefs. Second, the practitioner context provides meaningful heterogeneity in experience and domain-specific beliefs, which we leverage to shed light on the mechanisms underlying neglect. Differences in officers' understanding of the reliability of crime data and their experience with selected information predict whether selection enters their mental model and whether it is correctly incorporated into predictions. This heterogeneity allows us to distinguish representational failures from computational frictions in a way that is difficult to achieve with standard experimental samples. Consistent with this interpretation, prior work in the financial sector finds that selection neglect declines with trading volume and experience (Koehler & Mercer, 2009; Malmendier & Shanthikumar, 2007). More broadly, our findings complement a growing literature documenting systematic biases in belief updating among criminal justice practitioners, including judges and bail officers, highlighting the role of cognitive processes in high-stakes institutional settings (Arnold et al., 2018; Bhuller & Sigstad, 2024).

A defining feature of policing is that not all crimes are observed, and crime data is not a representative sample of the underlying crime. Who the police interact with, which crimes are observed or reported, and which incidents enter administrative records all depend on non-random selection processes. In the United States, there is broad consensus that minorities are patrolled, stopped, searched, and arrested more often (Aggarwal et al., 2025; Ba et al., 2021; Chen et al., 2023; Feigenberg & Miller, 2025; Gonçalves & Mello, 2021; National Academies of Sciences, Engineering, and Medicine, 2023). These disparities in enforcement —and thus observability— generate selection bias in crime data, making minorities overrepresented in administrative records (Gonçalves, Mello, & Weisburst, 2025). This has motivated recent theories of statistical discrimination in policing, which emphasize how inferences drawn from selected data can generate group-level disparities even absent animus (Che et al., 2024; Hübert & Little, 2023). What has been missing, however, is direct evidence on whether neglect generates such discrimination at the belief-formation stage. We fill this gap by showing that selection neglect mechanically maps selection into biased crime predictions, even when groups are abstract and carry no social meaning, and in the absence of preferences or stereotypes. In this sense, our findings establish that statistical discrimination in crime beliefs can emerge purely from a cognitive inference error operating on selected data. This contribution complements a growing body of work in policing that emphasizes the role of cognitive and emotional mechanisms in officer decision-making (Dube et al., 2024; Ferrazares, 2025; Holz et al., 2023; Owens et al., 2018), situating selection neglect as a belief-based channel within this broader behavioral perspective on policing.

By identifying selection neglect at the belief-formation stage, our results provide a micro-foundation for statistical discrimination in criminal justice systems. Policing occupies a unique position in this respect because it sits at the origin of the criminal justice data pipeline. Beliefs formed at this stage shape which events and individuals become observable, which cases enter administrative records, and therefore the data that downstream actors, such as judges and bail officers, use to form their own beliefs. When beliefs are formed from selected data and selection is neglected, distortions arise even in the absence of preferences or stereotypes. If similar inference processes operate at later stages of the system, belief distortions generated at policing can be mechanically propagated through successive rounds of selective observation and inference. In this way, selection neglect at the point where criminal justice data are first produced can trigger a chain of belief-level statistical discrimination that extends beyond policing itself.

The dynamics we document are not specific to policing, but arise more generally in environments where data are selectively observed and endogenously generated. In any domain where search of new information is prediction-based (Callaway et al., 2021; Reutskaja et al., 2011), current predictions influence which information become observable in the future, creating feedback between beliefs and data. When inference errors are stable, such environments give rise to cycles of bias in which distorted beliefs shape data generation and selected data reinforce those beliefs. This logic applies not only to human decision-makers, but also to algorithmic systems trained on endogenous data. Evidence from predictive policing and criminal risk assessment tools shows that algorithms trained on selected crime data can reproduce and amplify existing biases, precisely because selection is embedded in the data-generating process (Arnold et al., 2021, 2025; Lum & Isaac, 2016). In this sense, both human and algorithmic decision-makers operating in endogenous data environments can inherit and propagate bias when selection is neglected.

A natural response to these inference errors is to improve information provision or rely on algorithmic advice to correct biased beliefs. Our results suggest that such approaches are insufficient when the underlying problem is representational rather than informational. Officers do not neglect reporting rates simply because they lack access to them; rather, selection does not enter their mental model of the inference problem, so the information is neither demanded nor used. Consistent with this interpretation, making reporting rates salient does not eliminate neglect, and algorithmic recommendations that correct for selection are overridden when they contradict officers' inference rules. Similar patterns have been documented in other high-stakes settings, where judges and physicians override algorithmic advice, limiting its bias-correcting potential (Angelova et al., 2025; Mullainathan & Obermeyer, 2022). Together, these findings suggest that interventions targeting information or computation alone are unlikely to succeed unless they also address how decision-makers represent and reason about the data-generating process.

Our findings also point to several directions for future research. First, while we isolate selection neglect at the belief-formation stage, an important next step is to study how distorted crime predictions translate into outcomes in the field. Linking belief formation to enforcement behavior or judicial decisions would help quantify the downstream consequences of selection neglect across the criminal justice system.

Second, future work could explore interventions that target representation rather than information, such as alternative framings of data, training focused on the data-generating process, or institutional designs that reshape how selection enters decision-makers' mental models. Finally, extending the analysis to other professional contexts where data are selectively and endogenously generated, such as healthcare, hiring, or finance, would help assess the generality of selection neglect and identify the environments in which it is most consequential.

References

Aggarwal, P., Brandon, A., Goldszmidt, A., Holz, J., List, J. A., Muir, I., Sun, G., & Yu, T. (2025). High-frequency location data show that race affects citations and fines for speeding. *Science*, *387*(6741), 1397–1401.

Ali, S. N., Mihm, M., Siga, L., & Tergiman, C. (2021). Adverse and Advantageous Selection in the Laboratory. *American Economic Review*, *111*(7), 2152–2178.

Ang, D., Bencsik, P., Bruhn, J., & Derenoncourt, E. (2025). Community Engagement with Law Enforcement after High-Profile Acts of Police Violence. *American Economic Review: Insights*, *7*(1), 124–142.

Angelova, V., Dobbie, W., & Yang, C. S. (2025). Algorithmic Recommendations and Human Discretion. *The Review of Economic Studies*, rdaf084.

Araujo, F. A., Wang, S. W., & Wilson, A. J. (2021). The Times They Are A-Changing: Experimenting with Dynamic Adverse Selection. *American Economic Journal: Microeconomics*, *13*(4), 1–22.

Arnold, D., Dobbie, W., & Hull, P. (2021). Measuring Racial Discrimination in Algorithms. *AEA Papers and Proceedings*, *111*, 49–54.

Arnold, D., Dobbie, W., & Hull, P. (2025). Building Nondiscriminatory Algorithms in Selected Data. *American Economic Review: Insights*, *7*(2), 231–249.

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932.

Ba, B., Knox, D., Mummolo, J., & Rivera, R. (2021). The role of officer race and gender in police-civilian interactions in Chicago. *Science*, *371*(6530), 696–702.

Barron, K., Huck, S., & Jehiel, P. (2024). Everyday Econometricians: Selection Neglect and Overoptimism When Learning from Others. *American Economic Journal: Microeconomics*, *16*(3), 162–198.

Bhuller, M., & Sigstad, H. (2024). Feedback and Learning: The Causal Effects of Reversals on Judicial Decision-Making. *The Review of Economic Studies*, rdae073.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2023). How People Use Statistics. *NBER Working Paper*.

Brayne, S. (2020, November). *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press.

Brundage, M., Little, A. T., & You, S. (2024). Selection Neglect and Political Beliefs. *Annual Review of Political Science*, *27*(Volume 27, 2024), 63–85.

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLOS Computational Biology*, *17*(3), e1008863.

Che, Y.-K., Kim, J., & Mierendorff, K. (2024, September). Predictive Enforcement.

Chen, M. K., Christensen, K. L., John, E., Owens, E., & Zhuo, Y. (2023). Smartphone Data Reveal Neighborhood-Level Racial Disparities in Police Presence. *The Review of Economics and Statistics*, 1–29.

DANE. (2023). *Encuesta de convivencia y seguridad ciudadana (ECSC)* (tech. rep.).

DellaVigna, S., & Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting*. *The Quarterly Journal of Economics*, *122*(3), 1187–1234.

Dube, O., MacArthur, S. J., & Shah, A. K. (2024). A Cognitive View of Policing. *The Quarterly Journal of Economics*.

Enke, B. (2020). What You See Is All There Is. *The Quarterly Journal of Economics*, *135*(3), 1363–1398.

Esponda, I., & Vespa, E. (2018). Endogenous sample selection: A laboratory study. *Quantitative Economics*, *9*(1), 183–216.

Farina, A., & Herman, C. (2025). Communicating with Data Generating Processes: An Experimental Analysis.

Feigenberg, B., & Miller, C. (2025, March). *Class Disparities and Discrimination in Traffic Stops and Searches* (Working Paper No. 33629). National Bureau of Economic Research.

Ferrazares, T. (2025). Shift Structure and Cognitive Depletion: Evidence from Police Officers.

Gonçalves, F., Jacome, E., & Weisburst, E. (2025). Community Engagement and Public Safety: Evidence from Crime Enforcement Targeting Immigrants.

Gonçalves, F., & Mello, S. (2021). A Few Bad Apples? Racial Bias in Policing. *American Economic Review*, *111*(5), 1406–1441.

Gonçalves, F., Mello, S., & Weisburst, E. K. (2025, August). *Selection Bias and Racial Disparities in Police Use of Force* (tech. rep. No. w34175). National Bureau of Economic Research.

Holz, J. E., Rivera, R. G., & Ba, B. (2023). Peer Effects in Police Use of Force. *American Economic Journal: Economic Policy*, *15*(2), 256–291.

Hübert, R., & Little, A. T. (2023). A Behavioural Theory of Discrimination in Policing. *The Economic Journal*, *133*(655), 2828–2843.

Jehiel, P. (2018). Investment Strategy and Selection Bias: An Equilibrium Perspective on Overoptimism. *American Economic Review*, *108*(6), 1582–1597.

Jin, G. Z., Luca, M., & Martin, D. (2021). Is no news (perceived as) bad news? An experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, *13*(2), 141–173.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119.

Koehler, J. J., & Mercer, M. (2009). Selection Neglect in Mutual Fund Advertisements. *Management Science*, *55*(7), 1107–1121.

Lum, K., & Isaac, W. (2016). To Predict and Serve? *Significance*, *13*(5), 14–19.

Malmendier, U., & Shanthikumar, D. (2007). Are small investors naive about incentives? *Journal of Financial Economics*, *85*(2), 457–489.

Mullainathan, S., & Obermeyer, Z. (2022). Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care. *The Quarterly Journal of Economics*, *137*(2), 679–727.

National Academies of Sciences, Engineering, and Medicine. (2023, May). *Reducing Racial Inequality in Crime and Justice: Science, Practice, and Policy*. National Academies Press.

Owens, E., Weisburd, D., Amendola, K. L., & Alpert, G. P. (2018). Can You Build a Better Cop? *Criminology & Public Policy*, *17*(1), 41–87.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745.

Prendergast, C. (2021, February). 'Drive and Wave': The Response to LAPD Police Reforms After Rampart.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search Dynamics in Consumer Choice under Time Pressure: An Eye- Tracking Study. *The American Economic Review*, *101*(2), 900–926.

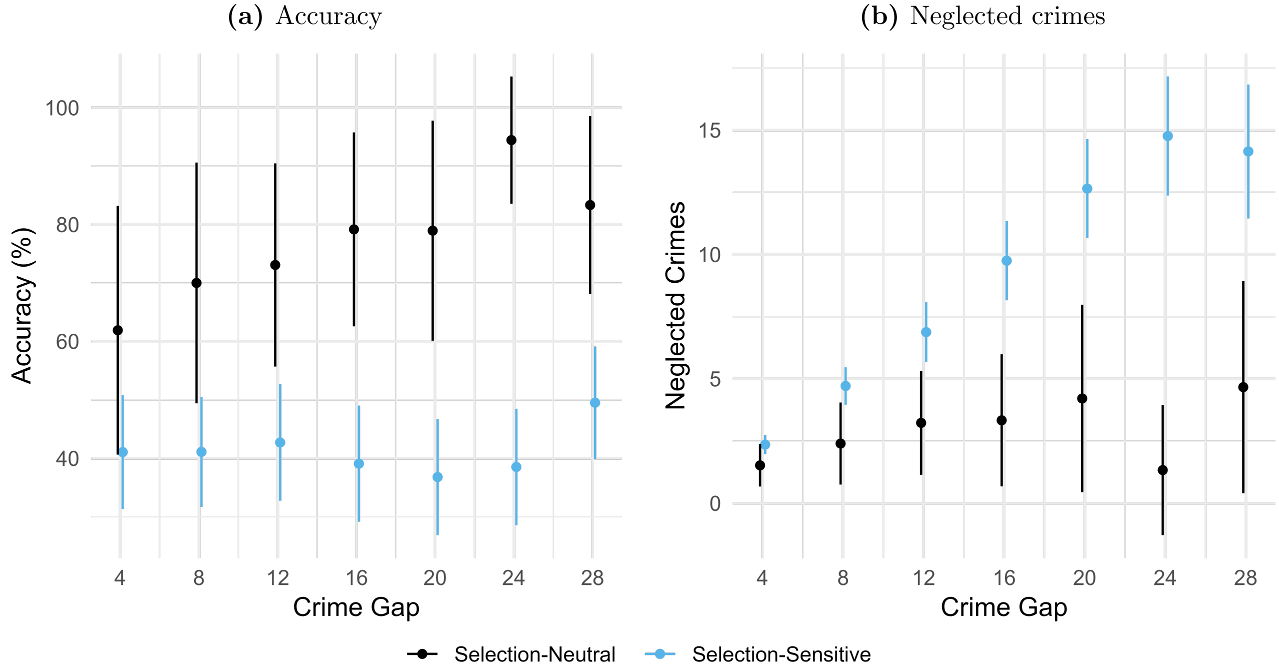
Schwartzstein, J. (2014). Selective Attention and Learning. *Journal of the European Economic Association*, *12*(6), 1423–1452.

Vomfell, L., & Stewart, N. (2021). Officer bias, over-patrolling and ethnic disparities in stop and search. *Nature Human Behaviour*, *5*(5), 566–575.

A Appendix

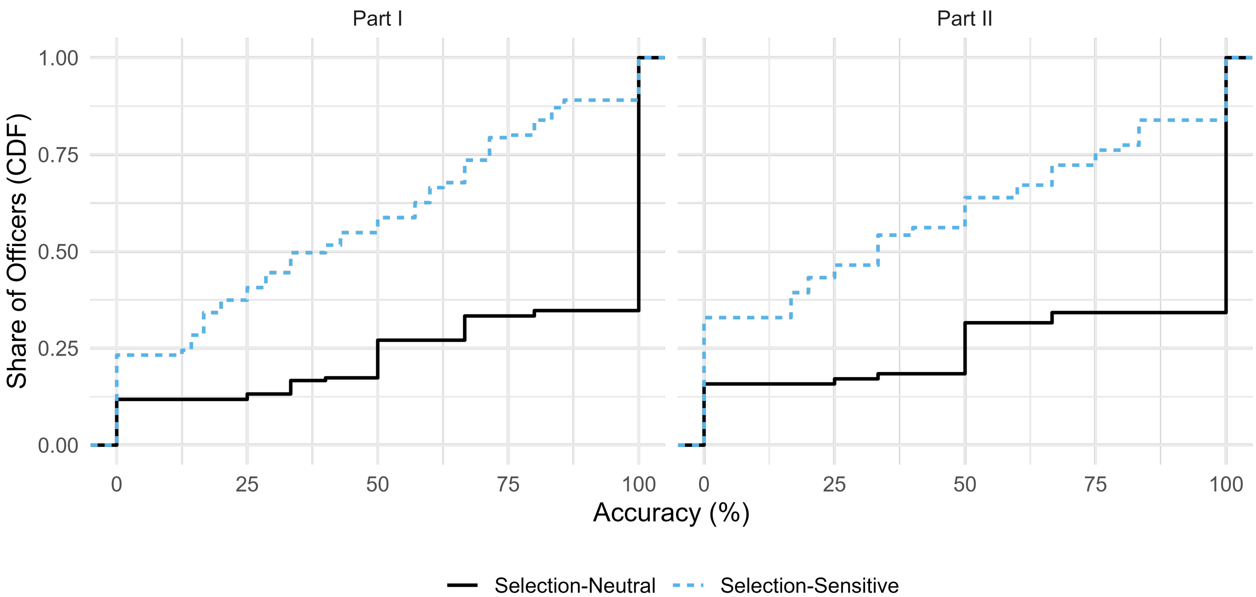
Additional Figures and Tables

Figure A.1: Accuracy and neglected crimes by crime gap and prediction type



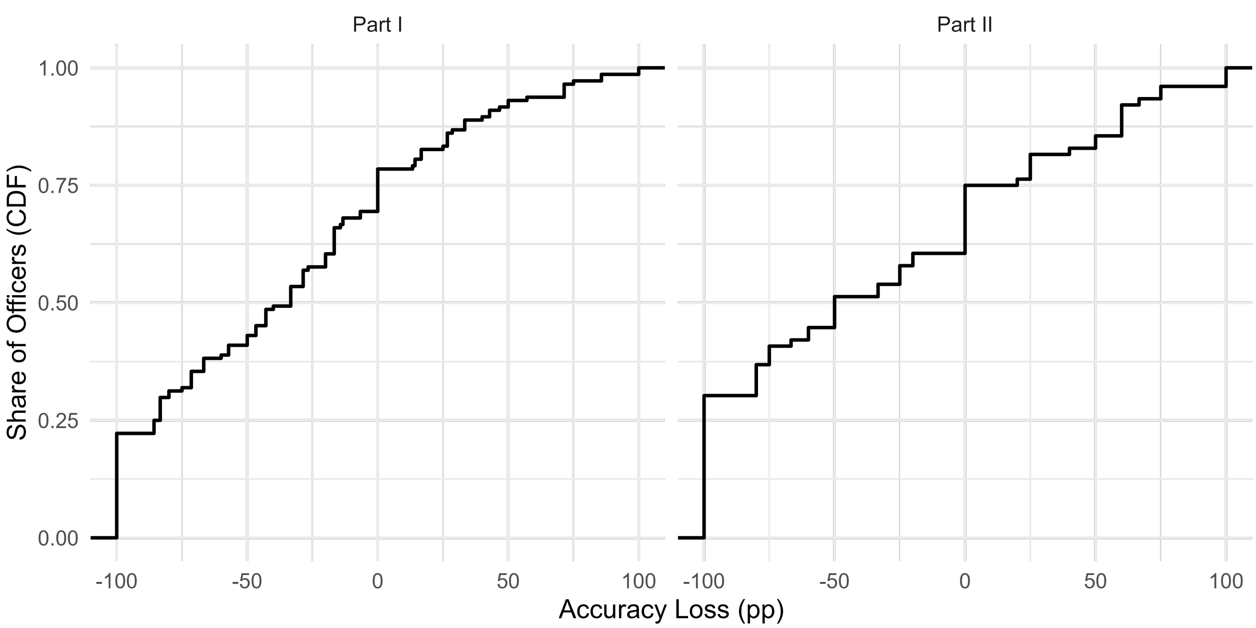
Notes: This figure disaggregates Figure 2 by crime gap, this is, the difference in crimes between neighborhoods at each prediction. Panel (a) shows the average prediction accuracy, by crime gap. Panel (b) shows the average number of neglected crimes —the difference in crimes between the optimal prediction and the actual one, by crime gap. Neglected crimes can be interpreted as the crime gap multiplied by the inaccuracy rate at that gap. Both panels further divide the data by whether adjusting for selection was needed to make the accurate prediction. Selection-neutral decisions, where adjusting for selection doesn't change the optimal prediction, are depicted in black; selection-sensitive decisions, where adjusting for selection does change the optimal prediction, are depicted in blue. Error bars in both panels represent the 95% confidence interval of the mean, with standard errors clustered at the officer level.

Figure A.2: Accuracy and neglected crimes by crime gap and prediction type



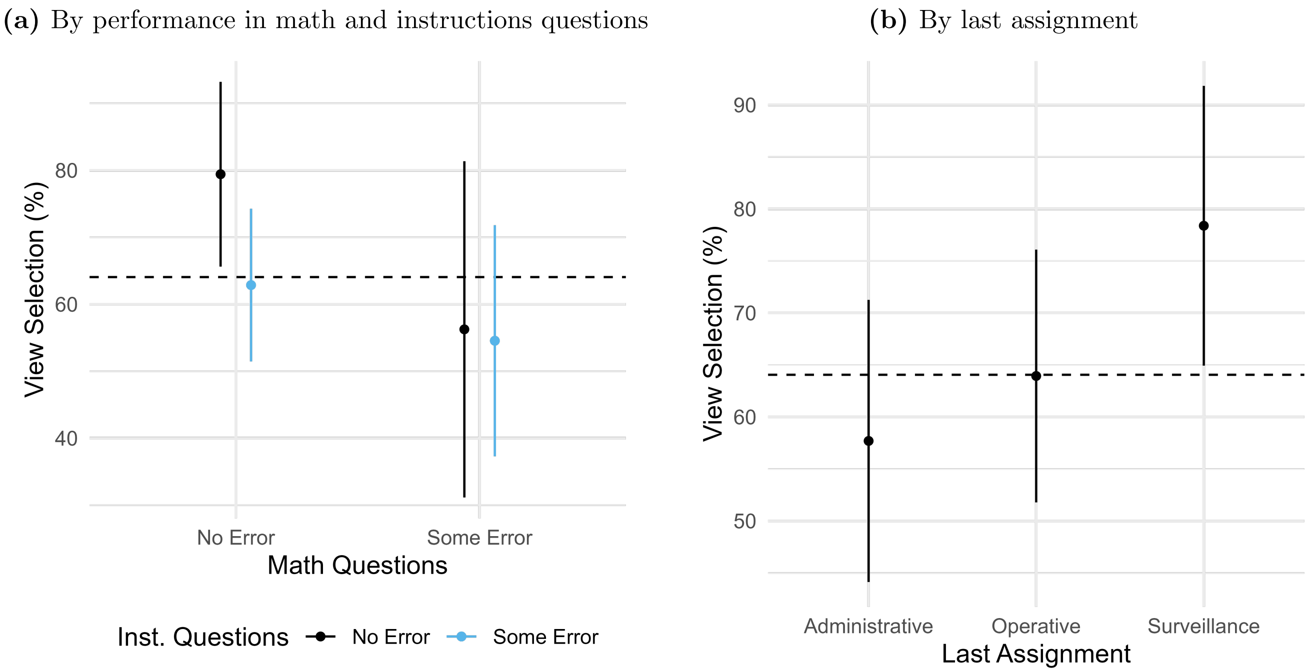
Notes: This figure shows the cumulative distribution of officer-level accuracy by type of prediction. Both panels divide the data by whether adjusting for selection was needed to make the accurate prediction. Selection-neutral decisions, where adjusting for selection doesn't change the optimal prediction, are depicted in black; selection-sensitive decisions, where adjusting for selection does change the optimal prediction, are depicted in blue. The left panel uses all eight predictions from Part I, the right panel uses the first six predictions from Part II.

Figure A.3: Distribution of accuracy loss due to selection neglect



Notes: This figure shows the cumulative distribution of officer-level loss in accuracy due to selection, this is, the difference in prediction accuracy in selection-sensitive relative to selection-neutral predictions. Negative values indicate officers who make less accurate predictions when these are selection-sensitive. The solid black line uses all predictions from Part I. The dashed blue line uses the first six predictions from Part II.

Figure A.4: Choice to View Selection



Notes: This figure shows the share of officers who choose to view selection for the last prediction of Part I. Panel (a) divides the sample by performance in the math and instruction comprehension questions. The left column shows officers who answered all math questions correctly, while the right column shows those who made at least one error. In black, we show officers who answered all instruction comprehension questions correctly, while in black we show those who made at least one error. Panel (b) divides the sample by the last assignment of the officer before entering the academy. Note that those with a Vigilance assignment were in charge of predicting crime and assigning patrols. In both panels, error bars represent the 95% confidence interval around the mean. The dashed line shows the sample-level mean at 64%.

Figure A.5: Accuracy Distributions in Selection-Sensitive Decisions

Notes: This figure shows the distribution of officer-level prediction accuracy in selection-sensitive decisions. Plots in the first row show predictions from Part I, where data selection is less salient. Plots in the second row show prediction from Part II, where data selection is made more salient. Columns separate the sample between officers who chose *not* to view the reporting rate for the last prediction of Part I (left, in black), and officers who chose to view selection (right, in blue). When comparing across columns, we notice the strong link between choosing to see selection and prediction accuracy in Part I, where selection wasn't salient. Once selection is made more salient in Part II, officers who hadn't chosen to view it in Part I (left column) don't really improve accuracy but become more likely to systematically neglect it. For officers who already were paying attention to reporting rates, the increase in salience doesn't seem to have any effect. In both cases, salience doesn't mitigate neglect. While bottom-up salience increases the likelihood that selection comes to mind, it does not substitute for having the correct adjustment rule.

Table A.1: Prediction accuracy across decisions

	Accuracy					
		Part I			Part II	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.779	0.673	0.828	0.770	0.667	0.821
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Selection-Sensitive	-0.358	-0.282	-0.418	-0.362	-0.323	-0.440
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)
View Selection		0.165			0.155	
		(0.007)			(0.111)	
Selection-Sensitive × View Selection		-0.119			-0.047	
		(0.183)			(0.704)	
Use Calculator			0.081			0.002
			(0.275)			(0.983)
Selection-Sensitive × Use Calculator			0.270			0.361
			(0.020)			(0.002)
Observations	1,240	1,224	784	930	918	588
Dependent variable mean	0.507	0.507	0.532	0.467	0.472	0.515

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Accuracy is a dummy that equals 1 if the prediction of which neighborhood had more crime was accurate. Selection-Sensitive is a dummy that equals 1 if adjusting for selection is needed to make an accurate prediction. View Selection is a dummy that equals 1 if the officer chose to view the reporting rate for the last decision of Part I. Use Calculator is a dummy that equals 1 if the on-screen calculator was used for that prediction. Columns (1) to (3) use the eight predictions from Part I, with Column (3) restricting the sample to those who chose to view selection in the last decision of Part I. Columns (4) to (6) use all predictions in Part II, excluding the last six predictions for participants randomized into an algorithm treatment other than *Control*. Column (6) restricts the sample to those who chose to view selection in the last decision of Part I.

Table A.2: Neglected crimes across decisions

	Neglected Crimes			
	Part I		Part II	
	(1)	(2)	(3)	(4)
Constant	9.880	10.466	3.026	1.741
	(0.000)	(0.000)	(0.000)	(0.034)
Selection-Sensitive	11.273	-8.718	6.367	-0.997
	(0.000)	(0.001)	(0.000)	(0.285)
Crime Gap		-0.011		0.081
		(0.799)		(0.228)
Selection-Sensitive × Crime Gap		0.516		0.459
		(0.000)		(0.000)
Observations	1,240	1,240	930	930
Dependent variable mean	18.435	18.435	8.353	8.353

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Neglected Crimes measures the prediction error in each decision: the difference between the highest number of crimes across the two neighborhoods and the predicted neighborhood. Note that this variable takes a value of zero if the prediction was correct, and the crime gap between neighborhoods if the prediction was wrong. Selection-Sensitive is a dummy that equals 1 if adjusting for selection is needed to make an accurate prediction. Crime Gap measures the difference in crimes between the two neighborhoods. Columns (1) and (2) use the eight predictions from Part I. Columns (3) and (4) use all predictions in Part II, excluding the last six predictions for participants randomized into an algorithm treatment other than *Control*.

Table A.3: Effect of crime and reporting gaps in prediction probability

	Predict A			Predict C		
		Part I			Part II	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.469			0.499		
	(0.000)			(0.000)		
Crime Gap	0.002	0.003	0.001	0.003	0.003	-0.001
	(0.000)	(0.000)	(0.098)	(0.011)	(0.007)	(0.674)
Reporting Gap	0.003	0.004	0.003	0.003	0.003	0.003
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)
View Selection × Crime Gap			0.002			0.006
			(0.062)			(0.009)
View Selection × Reporting Gap			0.000			0.000
			(0.700)			(0.980)
Fixed-effects						
Officer	No	Yes	Yes	No	Yes	Yes
Observations	1,224	1,224	1,224	918	918	918
Dependent variable mean	0.469	0.469	0.469	0.487	0.487	0.487

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Predict A is a dummy that equals 1 if the neighborhood from city A was predicted to have higher crime (Columns (1) to (3)). The dependent variable Predict C is a dummy that equals 1 if the neighborhood from city C was predicted to have higher crime, using predictions from Part II (Columns (4) to (6)). Crime Gap is the difference in crimes between the two presented neighborhoods at each prediction. Reporting Gap is the difference in reporting rates between the two presented neighborhoods at each prediction, in percentage points. View Selection is a dummy that equals 1 if the officer chose to view the reporting rate for the last prediction of Part I. Columns (1) to (3) use the eight predictions of Part I, including the gap in demographic decoys as controls. Columns (4) to (6) use the first six predictions in Part II. Columns (2), (3), (5), and (6) include officer fixed effects.

Table A.4: Accuracy and neglected crimes by selection, robustness

	Accuracy				Neglected Crimes			
	Part I		Part II		Part I		Part II	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.783	0.753	0.786	0.804	10.321	11.048	3.000	2.579
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.024)	(0.001)
Selection-Sensitive	-0.502	-0.364	-0.482	-0.423	16.620	11.322	8.184	7.172
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
No Math Errors	-0.006		-0.025		-0.683		0.042	
	(0.921)		(0.790)		(0.800)		(0.978)	
Selection-Sensitive × No Math Errors	0.211		0.178		-7.839		-2.680	
	(0.020)		(0.166)		(0.052)		(0.195)	
No Inst. Errors		0.070		-0.115		-3.093		1.509
		(0.221)		(0.248)		(0.236)		(0.346)
Selection-Sensitive × No Inst. Errors		0.036		0.197		-0.898		-2.602
		(0.685)		(0.114)		(0.816)		(0.204)
Observations	1,240	1,240	930	930	1,240	1,240	930	930
Dependent variable mean	0.507	0.507	0.467	0.467	18.435	18.435	8.353	8.353

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Accuracy is a dummy that equals 1 if the prediction of which neighborhood had more crime was accurate (Columns (1) to (4)). The dependent variable Neglected Crimes measures the prediction error in each decision: the difference between the highest number of crimes across the two neighborhoods and the predicted neighborhood (Columns (5) to (8)). Note that this variable takes a value of zero if the prediction was correct, and the crime gap between neighborhoods if the prediction was wrong. Selection-Sensitive is a dummy that equals 1 if adjusting for selection is needed to make an accurate prediction. No Math Errors is a dummy that equals 1 if the officer answered all math questions correctly. No Instruction Errors is a dummy that equals 1 if the officer answered all instruction comprehension questions correctly on the first attempt. Columns (1), (2), (5), and (6) use the eight predictions from Part I. Columns (3), (4), (7), and (8) use the first six predictions in Part II.

Table A.5: Heterogeneity in choice to view selection

	View Selection		
	(1)	(2)	(3)
Constant	0.577	0.576	0.281
	(0.000)	(0.000)	(0.607)
Operative	0.062	0.068	0.007
	(0.503)	(0.466)	(0.939)
Surveillance	0.207	0.200	0.119
	(0.035)	(0.044)	(0.251)
Reported Data Reliability		0.086	0.112
		(0.299)	(0.226)
Observations	150	150	147
Dependent variable mean	0.653	0.653	0.653

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable View Selection is a dummy that equals 1 if the officer chose to view the reporting rate for the last decision of Part I. Operative is a dummy that equals 1 if the officer's last assignment was operative. Surveillance is a dummy that equals 1 if the officer's last assignment was in surveillance —predicting crime and assigning patrols. The reference group is officers whose last assignment was administrative. Reported Data Reliability measures the officer's perception of how reliably reported crime data captures underlying crime, with higher values implying higher reliability. All Columns use data at the officer level, as there is only one choice to view selection. Column (3) also includes as controls some officer-level variables: their gender, age, and years of experience in the police.

Table A.6: Heterogeneity in accuracy loss due to selection

	Accuracy					
		Part I			Part II	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.788	0.782	0.788	0.732	0.732	0.754
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Selection-Sensitive	-0.428	-0.358	-0.426	-0.378	-0.338	-0.400
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Operative	0.002		0.007	-0.009		-0.030
	(0.971)		(0.917)	(0.921)		(0.688)
Surveillance	-0.038		-0.050	0.004		-0.034
	(0.609)		(0.465)	(0.972)		(0.706)
Selection-Sensitive × Operative	0.131		0.118	0.102		0.120
	(0.168)		(0.208)	(0.433)		(0.283)
Selection-Sensitive × Surveillance	0.082		0.104	0.015		0.064
	(0.485)		(0.343)	(0.915)		(0.611)
Report Reliability		0.115	0.124		0.278	0.281
		(0.082)	(0.068)		(0.000)	(0.000)
Selection-Sensitive × Report Reliability		-0.255	-0.262		-0.394	-0.406
		(0.010)	(0.008)		(0.000)	(0.000)
Observations	1,216	1,224	1,216	1,224	1,236	1,224
Dependent variable mean	0.512	0.511	0.512	0.467	0.462	0.467

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Accuracy is a dummy that equals 1 if the prediction of which neighborhood had more crime was accurate. Selection-Sensitive is a dummy that equals 1 if adjusting for selection is needed to make an accurate prediction. Operative is a dummy that equals 1 if the officer's last assignment was operative. Surveillance is a dummy that equals 1 if the officer's last assignment was in surveillance —predicting crime and assigning patrols. The reference group is officers whose last assignment was administrative. Report Reliability (Reported Data Reliability) measures the officer's perception of how reliably reported crime data captures underlying crime, with higher values implying higher reliability. Columns (1) to (3) use the eight predictions from Part I. Columns (4) to (6) use all predictions in Part II, excluding the last six predictions for participants randomized into an algorithm treatment other than *Control*.

Table A.7: Selection Neglect across Selection Environments

	Accuracy		
	(1)	(2)	(3)
Constant	1.000	0.565	1.000
	(0)	(0.000)	(0.000)
Selection-Sensitive	-0.591	-0.167	-0.600
	(0.000)	(0.260)	(0.000)
Asymmetric	-0.235		-0.455
	(0.000)		(0.000)
Selection-Sensitive × Asymmetric	0.230		0.449
	(0.006)		(0.034)
Endogenous		0.132	-0.345
		(0.324)	(0.003)
Selection-Sensitive × Endogenous		-0.214	0.236
		(0.236)	(0.146)
Endogenous × Asymmetric			0.523
			(0.006)
Selection-Sensitive × Endogenous × Asymmetric			-0.460
			(0.127)
Observations	930	330	330
Dependent variable mean	0.467	0.455	0.455

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Accuracy is a dummy that equals 1 if the prediction of which neighborhood had more crime was accurate. Asymmetric is a dummy variable that equals 1 if the officer was randomized into the *Asymmetric Selection* treatment, where neighborhoods from one of the two cities have a 30 percentage points higher reporting rate in expectation than the other city. It equals 0 when the officer is in the *Symmetric Selection* treatment, where reporting rates for both cities are drawn from the same uniform distribution, and thus are equal in expectation. Endogenous is a dummy variable that equals 1 if an officer is randomized into the *Endogenous Selection* treatment, where predicting one city to have higher crime increases reporting rates in that city by 15 percentage points in the next prediction. It equals 0 when the officer is in the *Exogenous Selection* treatment, where predictions don't affect future reporting rates. Column (1) uses the first six predictions of Part II. Columns (2) and (3) use the last six predictions in Part II, excluding officers randomized into an algorithm treatment other than *Control*.

Table A.8: Effects of selection environments

	Chose High Rate	Repeat Prediction	Persistent Prediction	
	(1)	(2)	(3)	(4)
Constant	0.496	0.532	0.056	0.000
	(0.000)	(0.000)	(0.153)	(1.000)
Systematic Bias	0.104		0.150	0.200
	(0.006)		(0.063)	(0.134)
Endogenous		0.146	0.201	0.214
		(0.028)	(0.014)	(0.065)
Systematic Bias × Endogenous			-0.037	-0.148
			(0.774)	(0.486)
Observations	930	330	155	55
Dependent variable mean	0.549	0.609	0.232	0.164

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Chose High Rate is a dummy variable that equals 1 if the city with the higher average reporting rate was predicted to have higher crime. The dependent variable Repeat Prediction is a dummy that equals 1 if the city predicted to have higher crime is the same as in the previous prediction. The dependent variable Persistent Prediction is a dummy that equals 1 if an officer predicted the same city to have higher crime across the last six decisions of Part II. Systematic Bias is a dummy variable that equals 1 if the officer was randomized into the *Systematic Selection Bias* treatment, where neighborhoods from one of the two cities have a 30 percentage points higher reporting rate in expectation than the other city. It equals 0 when the officer is in the *Uniform Selection Bias* treatment, where reporting rates for both cities are drawn from the same uniform distribution, and thus are equal in expectation. Endogenous is a dummy variable that equals 1 if an officer is randomized into the *Endogenous Selection* treatment, where predicting one city to have higher crime increases reporting rates in that city by 15 percentage points in the next prediction. It equals 0 when the officer is in the *Exogenous Selection* treatment, where predictions don't affect future reporting rates. Column (1) uses the first six predictions of Part II. Column (2) uses the last six predictions in Part II, excluding officers randomized into an algorithm treatment other than *Control*. Column (3) aggregates the last six predictions of Part II at the officer level. Column (4) repeats this analysis but excluding officers randomized into an algorithm treatment other than *Control*.

Table A.9: Standardized response times by decision features

	Standardized Response Time					
	Part I			Part II		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.085	0.272	0.192	-0.114	0.298	0.071
	(0.167)	(0.038)	(0.114)	(0.119)	(0.135)	(0.683)
Accurate	0.167	-0.411	-0.481	0.245	-0.468	-0.434
	(0.031)	(0.003)	(0.000)	(0.004)	(0.025)	(0.021)
Selection-Sensitive		-0.400	-0.365		-0.444	-0.317
		(0.003)	(0.003)		(0.032)	(0.080)
Selection-Sensitive × Accurate		0.752	0.713		0.855	0.566
		(0.000)	(0.000)		(0.001)	(0.009)
Use Calculator			1.747			1.323
			(0.000)			(0.000)
Observations	1,240	1,240	1,240	930	930	930
Dependent variable mean	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Results from a linear regression with $p-$values in parentheses and standard errors clustered at the officer level. The dependent variable Standardized Response Time is the z-scored response time for each prediction. Accurate is a dummy that equals 1 if the prediction of which neighborhood had more crime was accurate. Selection-Sensitive is a dummy that equals 1 if adjusting for selection is needed to make an accurate prediction. Use Calculator is a dummy that equals 1 if the on-screen calculator was used for that prediction. Columns (1) to (3) use the eight predictions from Part I. Columns (4) to (6) use the first six predictions of Part II.