

Luke Salamone

Simon Benigeri

Renpin Luo

Final Report

Theme:

For the CPD to function as intended, the incentive system needs to be aligned with the public interest. Given the CPD record of misconduct, the existence of phenomena such as the “blue wall of silence”, and the appearance of impunity for misconduct within the force, we expect to see that police officer rewards, as defined below, are not aligned with the fulfillment of their roles.

We will analyze instances of career rewards and misconduct records, and the relationships between rewards and misconduct. We understand that police officers get yearly pay raises, so better indicators of reward are pay raise % for the same rank, promotions, and awards. We understand that many factors go into career advancement: successes, failures, misconduct, race, gender, unit, visibility of one's work, network, etc. However, we believe that by controlling for some of these factors, including through the analysis of “doppelgangers”, we can see how misconduct correlates with various measures of career advancement. The central question we seek to explore is: *Do officers who frequently engage in misconduct get rewarded more than peers who offend less often?*

Checkpoint 1: Relational analytics

For the first checkpoint, our team attempted to find out if there is a relationship between the rewards and penalties of the police department. We also looked into the number and severity of allegations by comparing the promotion and award rates of accused officers each year. Visual results have been included in the last section.

Before breaking down tasks by number of allegations and sustained allegations, we needed to establish a baseline to understand the basics of promotions within the Chicago Police Department (CPD). Therefore, we found out the proportion of officers getting promoted or

awards each year as the first step. As it shows in figure 1, the numbers fluctuate year by year, with the range from 0.13% in 2011 up to over 2% in 2016. In 2017 over 12% of officers were promoted, but this number seems anomalous. In most years, awards are extremely common. After 2003, awards ranged in frequency from over 38% in 2004 to over 99% in 2009.

For the next step, the team attempted to find out how the number of allegations or sustained allegations affects the officer's likelihood of being promoted each year. By comparing the results in figure 2 and 4, in general, except for certain years, officers with no allegation were promoted at a higher rate than those with more than one allegation. Due to the lower overall number of promotions in some years, there is a relatively higher likelihood of anomalous data in those years. Such results also confirmed the team's initial hypothesis that an increase in the number of allegations should correlate with a decrease in the likelihood of being promoted. It is important to keep in mind that some of these allegations were not sustained, indicating that the police department did not punish officers for this kind of allegation. Sustained allegations seem to be more harmful to an officer's likelihood of being promoted. However, in some years, officers with one or more sustained allegations appeared to be more likely to be promoted than those with no sustained allegations. Officers with two or more sustained allegations in a given year were extremely rare, and none of them were promoted in most years.

The team engaged in observing the relationship between awards and allegations. 45 different types of awards were given out each year varying in number and degree of importance. By comparing the results in figure 3 and 5, the probability of an officer receiving an award seems to be directly correlated with the number of complaints received that year. For example, in 2006, 62% of officers with no allegation received an award, but officers with 1, 2, 3, 4 or more complaints were awarded at increasing rates: 84%, 90%, 94%, and 96% respectively. By breaking down the data to include only sustained complaints, the results still hold. This trend seems to actually incentivize complaints.

Checkpoint 2: Visualization

For this checkpoint our group continued to look at the relationship between rewards and allegations of misconduct. We produced two visualizations in Observable which are linked below.

The first visualization¹ studied award and complaint records for officers that we considered to be doppelgangers, i.e. officers with similar ages, demographic backgrounds and career trajectories. To do this we used a calendar with squares representing numbers of complaint events and award events for each month. We chose to take a closer look at officers similar to Jason Van Dyke since he was the officer who in 2014 murdered Laquan McDonald by shooting him 16 times. Our visualization demonstrated that Van Dyke was never heavily awarded at any point in his career, but received complaints fairly regularly from month to month, especially towards the end of his career in the months leading up to McDonald's murder.

For the second visualization², we were interested in whether officers with more accusations of misconduct were awarded more often. To do this, we chose to use a grouped bar chart to analyze the Chicago Police Department unit by unit. For a given unit, we looked at TRR counts per capita per year, allegation counts per capita per year, sustained allegation counts per capita per year and award counts per capita per year. We chose units 9, 18, and 25 for their high amounts of use of force and misconduct, and unit 19 for its relatively low numbers by those metrics. Our results showed that awards generally do tend to correlate with allegations of misconduct, and it still holds for sustained allegations as well.

Checkpoint 3: Interactive Visualization

For this checkpoint we looked at the relationship between pay and misconduct directly. To do so, we created a histogram³ showing officer salaries when split into 30 bins. Because it is known that certain demographic factors such as race and gender can correlate with salaries, we allow users to filter by these factors to limit confounding variables. For each distribution we also display the number of salaries which are displayed, and the mean of the group of salaries as well. If incentives are properly aligned, we should see that officers who received allegations and sustained allegations in the previous year would have lower pay than officers who received no such negative attention.

In fact, we saw a relationship between reward and punishment which seems much more troubling. Officers with no allegations had lower salaries on average than the average of all officers, and lower than those with any number of allegations as well. This relationship continues to hold when filtering by race and gender as well.

¹ <https://observablehq.com/@lukesalamone/calendar-view>

² <https://observablehq.com/@lukesalamone/officer-allegations-and-awards-chicago>

³ <https://observablehq.com/@lukesalamone/officer-salaries>

One might also expect that sustained allegations would have a negative effect on salaries. After all, a sustained allegation is an official recognition that an officer committed wrongdoing, not just a citizen complaint. However, this does not seem to be the case. In fact, as one filters for salaries of officers with 1 or more allegations, the average salary is lower than for officers with 2 or more allegations, and 3 or more allegations is higher than 2 or more! It appears that the relationship between pay and misconduct is exactly backwards from the ideal.

Checkpoint 4 Machine Learning

For this checkpoint, we continued exploring the factors which affect officer pay in a given year. We considered 10 factors⁴ which might have an effect on the change in pay for an officer? For example, in a system where allegations and sustained allegations are taken seriously, one might expect that these factors play a large role in the decisioning for pay raises for the next year.

Our analysis used the 10 factors to try to predict percent change in pay year over year using a model called CatBoost gradient boosting for regression analysis. We chose this model because we liked its performance on the categorical variables in our dataset, race and gender. Catboost is also less prone to overfitting, which was an important factor when selecting the model to use.

Our first piece of analysis looked to predict the percent change in salary for an officer. On this task, our regression analysis resulted in an R2 score of 0.96, meaning that 96% of the variation in next year's raise could be predicted by the features used in our model. CatBoost provides a method called feature importance which sheds light on the effect of each of the features on the prediction.⁵ We found that the year, salary, and experience play the largest role (98.7 out of 100) in determining how much raise an officer would receive in a given year. These factors account for the lion's share of the importance. It seems that the factors we were most interested in, allegations and sustained allegations, play little to no role in determining the change in wages for an officer, suggesting that they are largely ignored. Note that this runs counter to our previous analysis which showed that officers with more allegations tended to receive higher pay.

⁴ The 10 factors were salary, TRR count, honorable mention count, allegations count, sustained allegations count, experience (year - year joined), age (year - date of birth)

⁵ Feature importances were: year (40.58), salary (46.52), race (0.21), gender (0.14), trr_count (0.02), honorable mention count (0.34), allegations count (0.07), sustained allegations count (0.01), experience (11.60), and age (0.51). Importances are normalized and rounded so may not sum to 100.

For our second element of analysis, we again used CatBoost to look at officer pay as a time series problem, with salary varying by year. Rather than using year as a feature, we tried to use data from previous years to predict what would happen in the next year. We again trained a CatBoost model to determine this. This analysis had high RMSE and low R2 scores on our validation set.

Again we looked into the feature importances reported by our trained model.⁶ We again found that nearly all of the predicted value can be accounted for by salary, experience, and age. Together, these factors account for 97.35 out of 100 of the feature importance. This is similar to what we found in the previous CatBoost experiment. Again, allegations and sustained allegations were very low in terms of predictive power.

Checkpoint 5 Natural Language Processing

For this checkpoint we wanted to determine whether we could determine the category of an allegation based on features of the corresponding complaint report text. We used the FastText library to accomplish this checkpoint because it allowed us to get a working model quickly with a word vector representation and begin to classify text.

Our intention was to use a pre-trained fastText model as a base that we could fine tune but we were unable to implement this method with the resources available in Google Colab. Still, we were able to train our own supervised model on the narratives data and we used the fastText library to optimize it.

Our predictions were unconvincing, mainly due to significant degrees of class imbalance, which also negatively affected our analysis. Certain types of allegations (e.g. Operations and personnel violations) are much more common than others (e.g. First Amendment violations). This caused both overfitting and underfitting issues for us, where underrepresented classes were underfitted and overrepresented classes were overfitted.

Conclusion

Through our different experiments, we found that the relationship between pay and officer misconduct as well as awards and misconduct seem to encourage additional misconduct, and does not seem to be corrected for in subsequent years even after official investigations substantiate allegations against officers. This surprising finding suggests that the Chicago Police Department should reevaluate the way that pay is determined, since it now seems to be

⁶ Feature importances were as follows: salary (70.82), race (0.33), gender (0.20), TRR count (0.23), honorable mention count (1.42), allegation count (0.38), sustained allegation count (0.08), experience (24.71), age (1.82).

encouraging future misconduct. We found that this troubling relationship holds even when holding race and gender constant, and that different units in Chicago seem to have much different internal dynamics with regards to misconduct and award rates. Finally, we found that complaint narratives are an interesting but ultimately incomplete method of analyzing complaint categories.

Future Research

For checkpoint 1, it would be helpful to add sustained complaints and use of force events into our longitudinal visualization. This would add context to the award and complaint events. It may be the case that some officers have many use of force events without accompanying awards, which would bolster the conclusion that the officer in question is unusually aggressive. For the second visualization which analyzed unit by unit, it may be helpful in the future to exclude certain types of awards or look at certain types of awards (e.g. honorable mention awards) in more detail.

Results

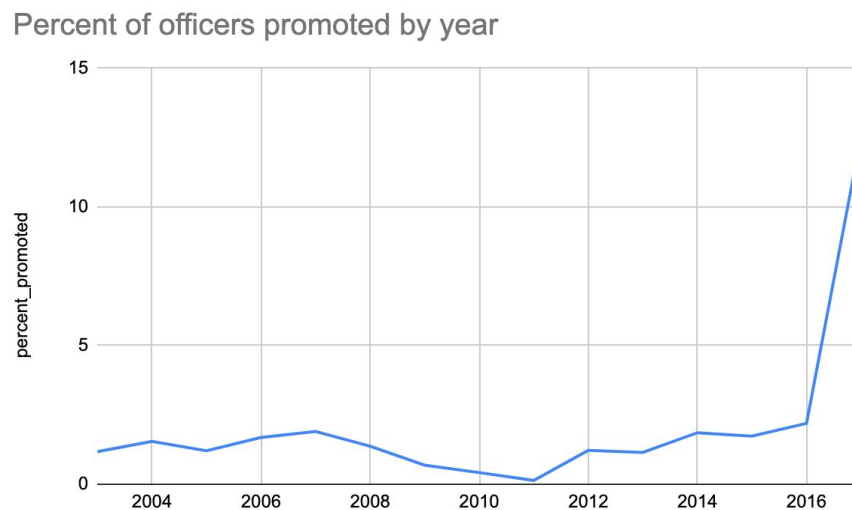


Figure 1

Percent of officers promoted by number of allegations by year

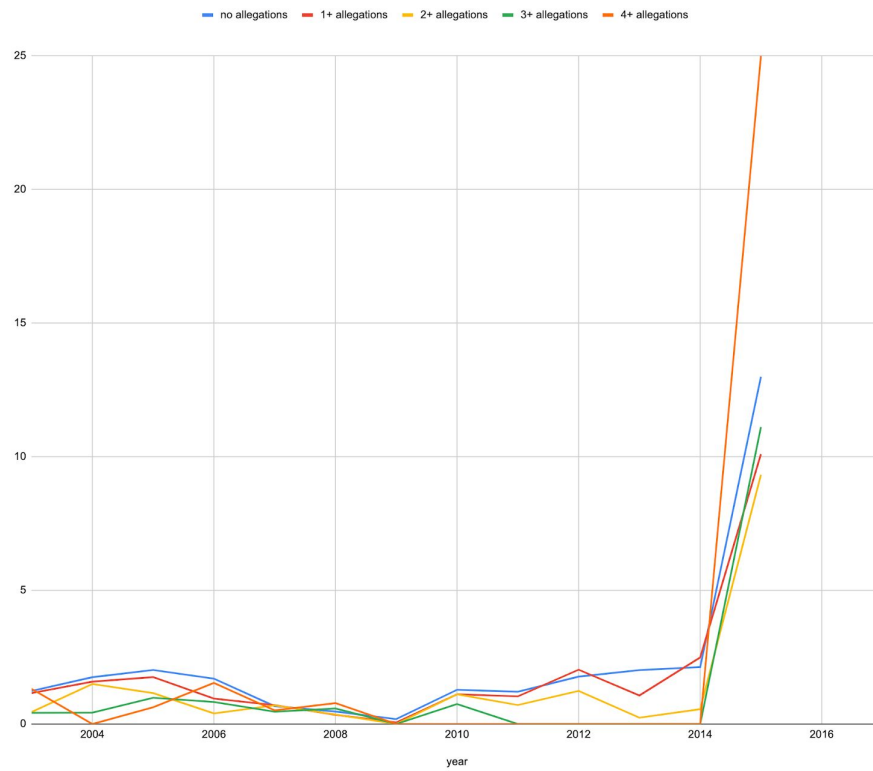


Figure 2

Percent of officers receiving awards by number of allegations per year

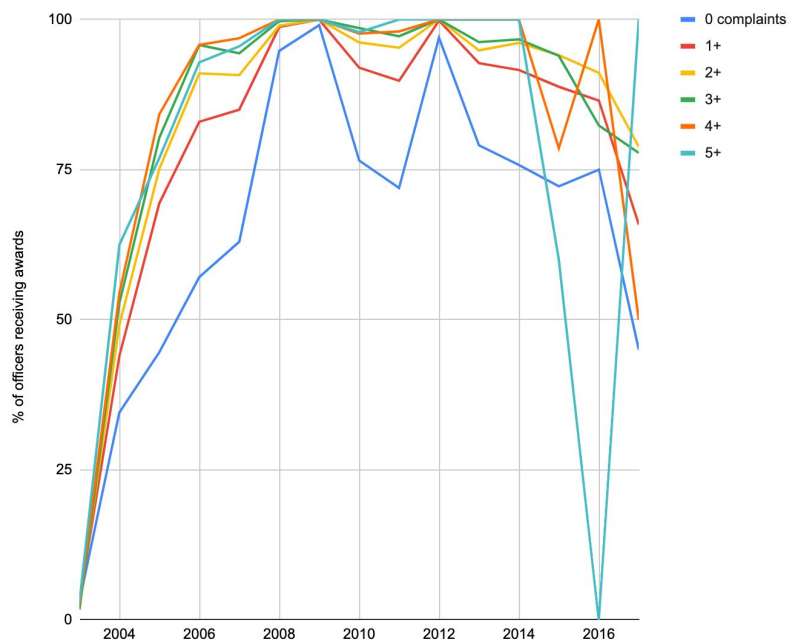


Figure 3

Percent of officers promoted by number of sustained allegations by year

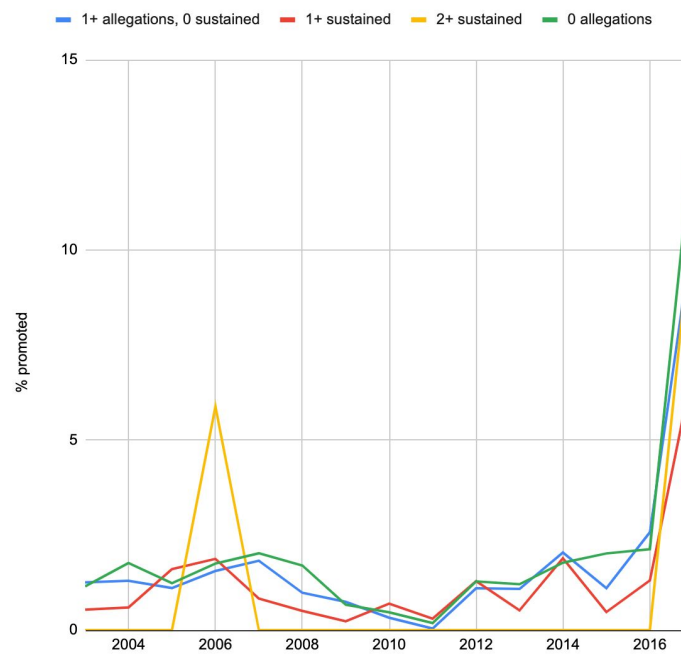


Figure 4

Percent of officers receiving awards by number of allegations sustained by year

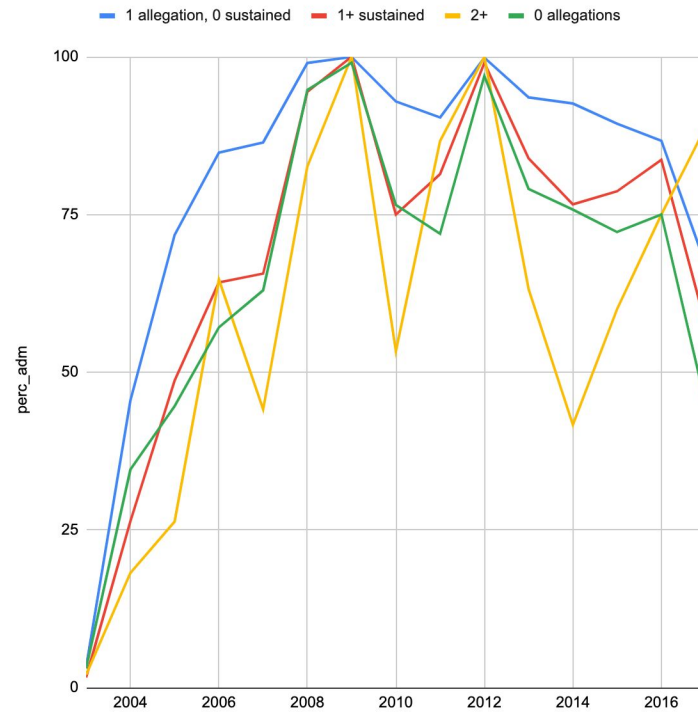


Figure 5