PHI - Case Study 1

**Case Study - Predictive Policing**

Charles Richardson

Group 11  
(Modified submission)

1. Priming question: **Predictive policing systems identify places and people who are at greatest risk of crime. This is a type of decision that has traditionally been made by humans that is now being “outsourced” to computer systems. Are there decisions that you would only ever trust a human police officer (and never trust an algorithm alone) to make? If you would only trust a human officer with certain decisions, what do you think is special about human decision-making? If you *would* trust an algorithm with any decision, how accurate would it need to be in order for you to trust it?**

The decision-making process of humans is a highly sophisticated and intricate matter, involving the consideration of various complex factors, such as the context of the situation, the novelty of the circumstances, and social cues from others. However, algorithms are limited in their decision-making capabilities as they are dependent on the quality of data they are trained on. Furthermore, while humans can be held accountable for the outcomes of their decisions and provide justifications for them, algorithms are often opaque, or “black box” making it difficult to explain how they arrived at their conclusions (Boonin, page 162). As such, it is critical to have human judgement at the center of decisions with significant implications for people's lives, such as those that arise in law enforcement or mental health crises. While algorithms can contribute to improved accuracy, they cannot match the empathetic, adaptive, and flexible decision-making abilities of humans in dynamic and unpredictable situations.

2. **Is the bias in predictive policing described above an example of preexisting bias, technical bias, or emergent bias? Maybe it’s more than one? Do you think that a predictive policing algorithm could escape the “feedback loop” problem? If so, how? If not, why not?**

The use of algorithms in decision-making has raised concerns about the potential for biases to be incorporated into the system, resulting in preexisting and emergent bias. As described by Nissembaum and Friedman, preexisting bias originates from societal or individual biases that are then incorporated into the algorithm through biased data (page 334). This is exemplified by the algorithm in the provided example, which focuses on black areas due to historically biased data. In contrast, emergent bias arises from the use of the algorithm and changes in the population after the system is created. In this case, if the algorithm targets hot spots, there would be a higher arrest rate in those areas, thus reinforcing the algorithm's bias through a feedback loop. While there are methods to mitigate the effect of this loop, such as using different traits based on race to determine a target value, it is crucial to acknowledge that these methods still rely on data from racially biased police practices, which is an issue that needs to be addressed independently. As Hellman suggests, we need to actively work towards promoting fairness and equality in decision-making processes to reduce the potential for biases in algorithms (page 853).

The statements above are examples of preexisting and emergent bias. In Nissembaum and Friedman’s article on page 334, they define preexisting bias as originating independently from the algorithm, such as from society or an individual. Then, the bias is incorporated into the algorithm, either consciously or unconsciously, through the biased data being fed into the system. Preexisting bias applies to the situation above since the system had made the decision to focus on black areas due to historically biased data it had been fed. In the same article on page 335, the authors describe emergent bias as arising from using the algorithm and from changes in the population after the system is completed. Since these hot spots are being targeted due to the algorithm, there would be a higher arrest rate within those areas. Data of higher arrest rates would be fed back into the system and would re-affirm the algorithm’s bias. This leads into emergent bias since a feedback loop eventually emerges after using the algorithm for some time.

We do not believe it is possible for an algorithm to completely escape from the feedback loop. The algorithm itself would not be able to recognize that it is in a feedback loop since it is not programmed to do so, so it cannot break out of it. It is possible to mitigate the effect of the loop by increasing the fairness of the algorithm. On, Hellman proposes using different traits based on race to determine a target value. An option could be to use this method with the population to determine how much to patrol an area, and this would account for historical disadvantages black people have, thus increasing fairness. However, this cannot completely escape the feedback loop since it was the result of data from racially biased police practices, which is an issue that needs to be addressed independently.