Kynan Lewis, Daniel Kagramanov
Prof. Sadoghi
ECS 188
06/10/2025

Math & Morality: Where Do We Draw The Line?

In the modern day, computers have settled into every part of our world. Whether it be at home, at work, or in schools - things that used to be done by pen and paper are now done by programs. In many cases, computers have taken up positions that used to be filled by people; after all, computers are much faster and reliable when it comes to manipulating data. These are called mathematical models, which take all sorts of different pieces of information and use them to make decisions. However, when the lives of human beings are held in the hands of a machine's determinations, it raises some concerns.

For better or for worse, mathematical models dictate many aspects of our lives today. They may dictate our ability to receive aid, our ability to get a job, even our ability to get loans. Weapons of Math Destruction, a book by Cathy O' Neil, sheds some insight on the fact that mathematical models our government creates impacts our lives, sometimes in a negative way. Inherent biases exist within the creation of these models, and as such, leave the door open for a lot of criticism. The question begs to be asked then: Is it ethical to leave it up to a set of predetermined variables to make decisions about a person's situation?

Opacity, Scale, and Damage. These are the three categories regarding the models O' Neil spoke about, nicknamed WMDs. Opacity relates to how transparent a model is. What this means is that the average person will not understand what's going on behind the scenes. These models will work "...invisible to all but the highest priests in their domain: mathematicians and computer scientists" (O' Neil, pg. 3). The concern here is that people do not have a chance to understand how they're being judged, and what they can do to help themselves out. Scale refers to how large of a population these models affect. Models that decide a person's eligibility to receive financial aid can affect millions of people all across America. And finally, damage. This reflects the potential to harm a person's livelihood.

A perfect example of a WMD that O' Neil highlighted was Sarah Wysocki, a teacher in D.C. who got fired in 2011 unjustly by a mathematical model called IMPACT. IMPACT was a model seemingly created to address the learning gap in underperforming schools. The notion was that each teacher would receive a score based on how well their students would do on standardized tests leading into the next year, which represented their 'teaching ability'. Sarah was praised by students and parents alike for how dedicated she was to her classes, yet scored

tremendously low on the IMPACT evaluation due to her students not scoring well, and she was deemed a bad teacher. She proceeded to get fired with no room for appeal. The critique that O' Neil proposed was that this model, like others, was too binary. She brought up an example that these students in less privileged neighborhoods may have extraneous factors that could harm their test scores. Life at home, life at school, how they're feeling, can all impact a student's mental state, and subsequently tank their test scores. If they take an exam one year in a position where they're stable, and then face some sort of crisis the next year, the IMPACT evaluation dictates that it's the teacher's liability. If all of the students love Sarah, yet perform poorly on the test, that doesn't necessarily mean she should be faulted for being a bad teacher.

Herein lies the problem. Even if you create variables upon variables, trying to take every situation into consideration, you cannot fully predict the impact a model can have. IMPACT wanted to address the issue of poor teaching in order to shape a better future for the students, and improve their experience within the school system. But as you can see, it had unintended consequences arising from some overlooked possibilities. This is where we get to the ethical dilemma behind these models. The implementation of these algorithms is supposedly objective: if the math is what's making all of the decisions, then it should eliminate any biases from decision making, right? Think for example, job screenings. A recruiter who may be tired at the end of the day from rummaging through stacks of papers might gloss over a resume because she wants to clock out on time, and pass up a perfectly good candidate in their carelessness. On the other hand, an algorithm has been designed to just find just the keywords, and give those the most priority. It wouldn't possess the same potential for "human error" that the recruiter does with their biases. But now, the recruitment process isn't about actually having the right experience, it's about ticking all of the boxes. This same candidate who has the right experience may not be very familiar with writing resumes, and write one that's not the same format as the machine expects, thus getting rejected. While this may not be as impactful as other algorithms, the point still stands: every choice will be binary, with no room for context or rational judgment. The possibilities for statistical outliers and variation are largely disregarded.

Furthermore, the assumption that mathematical models eliminate biases might be incorrect. After all, algorithms are created by people, and it's possible that the creator's biases are ultimately reflected in their program. Additionally, an algorithm might inadvertently create its own biases based on the data that it receives. We see this in people being judged on factors that aren't true to their actual situation, which will hurt them in the long run. For example, people from worse backgrounds who are being sent to jail may get more time on their sentence purely because they live there. One of the models that the judicial system uses in order to measure the risk of recidivism (how likely a criminal will commit a crime again) is called the LSI-R (Level of Service Inventory - Revised). It's a questionnaire that the inmates are given, and based upon the answers it receives, will determine how likely they'll reoffend. O' Neil highlighted the issue with the questionnaire by giving a hypothetical: "Ask a criminal who grew up in the suburbs about the

‘first time you were ever involved with the police’, and he might not have a single incident to report… while black and Latino males between the ages of fourteen and twenty-four … accounted for 40.6 percent of the stop-and-frisk checks by the police” (O’ Neil, pg. 26). More often than not, they will receive worse sentences purely based on their living experience. This attests to the fact that people from poorer areas are affected disproportionately by these models to people from better socioeconomic statuses.

This is how prejudice manifests within these math models. Although it might seem that 0’s and 1’s cannot be prejudicial, by virtue of the fact that their judgment of people can affect them disproportionately and unfairly, it is inherently biased. Furthermore, by being completely binary, the algorithm takes an objectivist approach to decision-making, where there is a “right” decision to be made in each case, without consideration for context or nuance. However, the disregard for context can often be harmful, as seen in our previous examples, and, due to the presence of bias, it ultimately fails to be truly objective. Wouldn’t it be more fair to judge from a relativistic perspective, rather than a flawed objective one?

However, this brings us to a problem. How would we efficiently sort through all of this data without algorithms? There’s not nearly enough manpower or time for the government to actively go through every single person’s situation, one by one, go through all the important background details, speak with the person to figure out aspects of their lives, and then decide where to go from there. It’s just not feasible. So we run into a catch-22: we need these models for their efficiency, even though they can be very flawed. Is it still ethical to use large scale math models in order to regulate these processes, knowing that the oversight and updates to these systems are not up to par?

If we look at things from a utilitarian point of view, employing the use of algorithms is the correct thing to do. Even if a small number of people end up being harmed because of them, the benefit outweighs that. Since not using algorithms means a loss of efficiency in government operations, using them is a better option, and indeed, ethical. The argument that they’re simply too useful to give up is an undeniable one.

The employment of large-scale math-models also holds up under the scrutiny of contractualism. Being a citizen of a country means that you inherently agree to follow its rules and stipulations. Since the government has put these math models into use, then as per the “contract”, the citizens should abide by it. Even should it harm a few people, if the government determines that these models could still be reasonably justified, this is ethical under contractualism.

Existentialism, however, would look at it from another way. They, above all else, support an individual's right to freedom of choice, and that each person chooses whether they want to participate in something or not. These models, being created by other people, relinquishes an individual's right to choose. They cannot choose themselves whether or not they want this model to be applied to them, and the models affect them directly. As such, it is not ethical to continue to use these models, and they should be replaced with an alternative.

Lastly, from a Kantian perspective, these large mathematical models seem to be ok at first. After all, the intent of these algorithms is to benefit the masses, so it doesn't matter if it ultimately causes problems. However, if we look a little further, we see that it goes against a few principles: first, we shouldn't continue to use it when we know that it will cause problems. Secondly, the algorithm reduces the lives of people down to mere points of data, thus treating them as a means to an end and failing to uphold Kant's views of ethics. This only ends up bringing us back to the problem, how do we make the system more "personal" without employing the efficiency of algorithms?

Clearly these math models are too useful to give up, since it isn't possible to have humans review each and every case that comes their way. Even so, it's unethical to continue using them in their current state, when they can harm people without any room for appeal. In that case, is it possible to find a middle ground?

With enough overview and active revision, these models wouldn't be as bad. If people regularly monitor the data, filter it, find statistical inaccuracies, and tweak the model based on this new information, it would help reduce the negative feedback loops that they induce. Furthermore, the math models need to be more transparent. Currently they are opaque, with no way for the average person to understand them - it would help significantly if the designers were to convey the details of how these models operate, what to look out for, and how to challenge them effectively. Lastly, there needs to be room for personal appeal. As we've seen, it's dangerous to completely remove the human element from decision making by reducing people to numbers, and generalizing their situation. Although this cuts down on possible biases, it also means that people who are outliers get grouped up with others and categorized incorrectly. While the vast majority of cases can be handled fairly by math models, there should be a way for people to have their situation reviewed by a person, who can parse the situation in ways a machine is incapable of. These changes would help to strike a balance between efficiency and fairness, reintroducing the relativism that machines are incapable of, and making the employment of large scale math models better aligned with the ethical principles that we've discussed.