Meeting 8: Culture, Computation, Morality

CS198: The Poetry of Computer Science, the Computer Science of Poetry
Philosophy of Computation at Berkeley
pocab.org

November 27th, 2017

1 Weapons of Math Destruction

Algorithms are seeping into every corner of society. Is this good or is it bad? Can algorithms solve all problems? Or are there some domains where algorithms ought not to be used? For example, should a teacher be fired based on an algorithm? Is there *any* such computable algorithm that *guarantees* some ϵ probability of correctness, and if so, can we prove that? Or, can we prove it wrong?

Cathy O'Neil[1] is a mathematician and a bluegrass musician. She went to Berkeley for her undergraduate degree, received her Ph.D. in mathematics from Harvard, and taught at MIT as an assistant professor for a number of years. Then she left academia to become a quant at D.E. Shaw. After a few short years there, she grew disillusioned, and quit.

> I had gone into finance thinking I was making the market more efficient, and now I was trying to make money off of people who were saving for retirement. I started thinking of us as junk-yard dogs, scavenging off of the financial system's scraps.
>
> (Interview in *The New Yorker*, "Bluegrass and Big Data"[2])

Increasingly concerned about the impact of algorithms on society, she is a prominent activist against the overreach of technology. Here is an excerpt from her most influential book, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*:

> In 2007, Washington, D.C.'s new mayor, Adrian Fenty, was determined to turn around the city's underperforming schools. He had his work cut out for him: at the time, barely one out of every two high school students was surviving to graduation after ninth grade. (...) Fenty hired an eudcation reformer named Michelle Rhee as chancellor of Washington's schools.
>
> The going theory was that the students weren't learning enought because their teachers weren't doing a good job. So in 2009, Rhee implemented a plan to weed out the low-performing teachers. (...) Rhee developed a teacher assessment tool called IMPACT, and at the end of the 2009-10 school year the district fired all the teachers whose scores put them in the bottom 2 percent. At the end of the following year, another 5 percent. (...)
>
> Sarah Wysocki, a fifth-grade teacher, didn't seem to have any reason to worry. She (...) was getting excellent review from her principal and her students' parents. One evaluation praised her attentiveness to the children; another called her "one of the best teachers I've ever come into contact with."
>
> Yet (...) Wysocki received a miserable score on her IMPACT evaluation (...) This left the district with no choice to fire her. (...)
>
> This didn't seem to be a witch hunt or a settling of scores. Indeed, there's a logic to the school district's approach. Administrators, after all, could be friends with terrible teachers. So

[1] mathbabe.org
[2] https://www.newyorker.com/magazine/2016/10/10/bluegrass-and-big-data

> Washington, like many other school systems, would minimize this human bias and pay more attention to scores based on hard results: achievement scores in math and reading. The numbers would speak clearly, district officials promised. They would be more *fair*. (...)
>
> [However], attempting to reduce human behavior, performance, and potential to algorithms is *no easy job*. (...)
>
> The model itself is a black box, its contents a fiercely guarded corporate secret. This allows consultants (...) to charge more, but it serves another purpose as well: if the people being evaluated are kept in the dark, the thinking goes, they'll be less likely to attempt to game the system. (...) But if the details are hidden, it's also harder to question the score or to protest against it. (...)
>
> After the shock of her firing, Sarah Wysocki was out of a job for only a few days. She had plenty of people, including her principal, to vouch for her as a teacher, and she promptly landed a position at a school in an affluent district in northern Virginia. So thanks to a highly questionable model, a poor school lost a great teacher, and a rich school, which didn't fire people on the basis of their students' scores, gained one.
>
> (*Weapons of Math Destruction*, Cathy O'Neil)

2 The Judgment Algorithm, Revisited

In our last meeting, we considered the following argument:

> Assume that humans are universal Turing machines, that is, a Turing machine able to execute any Turing machine whatsoever. From this, we can assume that a human H is an *arbitrary* Turing machine. Now suppose there exists a Turing machine J such that $J(H) = i$ where $i \in S$ and S is a well-ordered set of numbers. Also assume that J looks at the output of H – the output of an arbitrary Turing machine – to compute the output i. So J can be used to compare humans, such that if $J(H_1) > J(H_2)$, H_1 is more "worthy" than H_2. But H is an arbitrary Turing machine, and by the uncomputability of the halting problem, we know that J cannot know if H even halts or not! Therefore, no such J exists.

A forceful objection is that humans are not universal Turing machines. This is very plausible: if we assume humans are universal Turing machines, we would be saying that anyone can do anything anyone else can do. Which would mean that anyone could become like Newton, Einstein, or Elon Musk. Now that seems like a version of a elementary school pep-talk session – "you can do anything!". Call it naïve or hopeful, but it is not totally plausible.

But notice that, in the argument, we did not *need* humans to be universal Turing machines. We only needed them to be *arbitrary* Turing machines. This is a weaker requirement. When we say that each person is an arbitrary Turing machine, it does not imply that anyone can do anything anyone else can do. It means something more like, each human's computation (and mental contents) is unique, and no computation from the outside can reduce that human's experience to something short of reproducing that entire computation.

Still, why should we believe that? Maybe humans aren't Turing machines at all. Humans are bounded things. We will all die one day, and when we die, the lights go out, it's all over.[3]

Francis Bacon, an English philosopher in the 16th century, had the same problem. His time was when empiricism started to develop, and doubts about God also started to develop. To convince empiricists that they should believe in God, he had to argue that it is rational to believe in God. To this end, he proposed the following payoff matrix:

	Believe in God	Believe in no God
God exists	Heaven: infinite payoff	Hell: negative infinite payoff
God does not exist	No payoff	No Payoff

[3]For some reason, "dying" intuitively corresponds well with "halting". But is this correspondence justified? Recall Hofstadter's argument that his deceased wife, Carol, *is literally in his brain*, because she has left a lasting mark – a lasting set of Turing machines – on him.

So, regardless of whether God exists or not, one can expect a higher payoff by believing in God, and should, rationally, believe in God. Ponder this matrix for a bit. Does it make you believe in God? Why or why not?

Our situation is somewhat similar. Should one believe that humans are arbitrary Turing machines, or should one believe that none are?[4] We can construct a similar payoff matrix:

	A believes humans are arbitrary TMs	A believes humans are not arbitrary TMs
All humans are arbitrary TMs	A is an arbitrary TM (A is free)	A is not an arbitrary TM (A is not free)
No humans are arbitrary TMs	A is not an arbitrary TM (A is not free)	A is not an arbitrary TM (A is not free)

(Where A is a human)

First things first: if no humans are arbitrary Turing machines, A is automatically not an arbitrary Turing machine, and so A is computable, A is not free. These are the bottom right two boxes. So the only interesting case is when all humans are in fact arbitrary Turing machines. If this be the case, and A believes that all humans are universal Turing machines, then A is an arbitrary Turing machine, A is uncomputable, A is free. This is the center box. This is great.

However, if A believes no humans are arbitrary Turing machines, and all humans are in fact arbitrary Turing machines, something interesting happens. *Because* A believes humans are not arbitrary Turing machines, it *causes* A to become *not* an arbitrary Turing machine. Let me explain just what I mean. So far we have been vague about what it means for A to "believe" something, but what, exactly does that mean? Well, to believe something is to be engaging in some thought, to be having some mental content. And, as disciples of the Church-Turing thesis, we assume that there exists a Turing machine for any mental content, that any mental content can be described as a Turing machine. So A's process of believing *is* the execution of a Turing machine. In fact, it is the execution of a *wrong* Turing machine! We know it is wrong because this is a Turing machine that says an aribtrary Turing machine is not an arbitrary Turing machine, in other words, that an arbitrary Turing machine can be computed; in other words, this is a Turing machine that purports to be solving the halting problem. And we know no such Turing machine can do that correctly. Because the belief is a *wrong* Turing machine, the belief can actually be described by a sub-Turing machine, such as a finite state machine. As a concrete example, a racist who bristles at the sight of people with a skin color c can be described by the following very simple program:

```
if c is seen:
    bristle
```

Because A is having a wrong belief, we know A is executing a simpler program. This lets us describe A as a simple machine; it lets us *compute* A. Therefore, A's belief causes A to be not free.

Does this payoff matrix suffer from the same mistake as Bacon's? Why or why not?

3 Self-Driving Cars That Kill People

Suppose a self-driving car must kill person A or person B. Obviously, this is not a realistic situation. But for the sake of argument, let's suppose that the car knows with certainty that if it takes some action e_1, person A will die, and if it takes some other action e_2, person B will die. There are no other available actions; it must take action e_1 or e_2.[5]

If our Judgment Algorithm J exists, the solution is simple: save A if $J(A) > J(B)$, and save B if $J(B) > J(A)$. If we can prove that J does not exist, the solution is also simple: flip a coin and kill A with $\frac{1}{2}$ probability, kill B with $\frac{1}{2}$ probability. Implicitly, this says that all people have equal moral value.[6] So the question is... does J exist, or not? What is your final verdict?

[4] A third position is possible, where one believes some people are arbitrary Turing machines, but some are not. This position is addressed, and shown to be incoherent, in section 5, "Why You Shouldn't Judge Just Anyone".

[5] Not doing anything is itself an action. For example, maybe e_1 is the action where the car chooses to not do anything.

[6] What if the car knows it can kill either person A or two people B, C? In this case, the probabilities can be multiplied, so

4 If Humans Were Arbitrary Turing Machines...

O.K., whatever, let's say humans are arbitrary Turing machines. How does that change anything? Well, for one, it shows that the Judgment Algorithm, J, does not exist. So it means a self-driving car should flip a coin instead of analyzing a person and outputting a "value" of the person to decide whom to kill. But how does it effect our day-to-day, non-self-driving-car-owning lives?

It can effect us like the following:

- One cannot rationally hold a static, unchanging idea of a person in one's head. Suppose Nic is very very mad at Tharis. In this case Nic is apt to *compute* Tharis, focusing only on her one or two attributes and thinking that those attributes describe her completely. Nic may be tempted to put Tharis down, that is, compute Tharis and give her a lower score than what Nic would give himself. This is just a strange way of saying that, if Nic were mad at Tharis, Nic might try to get over it by saying something like, "I'm better than her. I'm not giving her another chance. She's not worth my time." But if Nic remembers that all humans are arbitrary Turing machines, Nic will realize that he is *wrong*, that he is not being *free* in his anger towards Tharis. Then he will examine why exactly he is not free as such. He may write a journal entry, cry for hours on end, or talk to Tharis. He will do anything to stop being not free, even if it takes a very long time and a very heavy effort. Over time, he may come to forgive Tharis. They may even get married.

- One cannot rationally feel worthless about oneself. In the stressed mind of a Berkeley student, the following oscillating thought pattern, or something like it, may often occur:

  > I got a bad grade on the exam. No, no, no, I'm a failure! I'm so much worse than all my peers. John next door has an internship at Google, where I got rejected from. Why am I such a failure as a human being? ...But, after all, I go to Berkeley! That must mean I'm, like, the top 0.1% in intellect. I've been smart all my life. My mom tells me so. I didn't get an internship at Google, but I got one at Microsoft. Microsoft is the up-and-coming company, anyway, and Google is becoming evil. I don't even *want* to be at Google. ...But, I got rejected from Google, so I'm a failure. ... But I'm a national merit scholar... But I only got a 2260 at the SAT, while John got a 2400 ... But....
  >
  > (Stressed Berkeley Student, circa 2017)

  What is the common thread in almost every single sentence of the above thought? It involves the *computation of the uncomputable*. The student compares oneself to John, then compares oneself to the rest of humanity, then compares the perception of working at Microsoft versus that at Google, and so on. But if the student can only remind him/herself that humans are arbitrary Turing machines, and thus uncomputable, he can escape this thought pattern.[7]

- At the same time, if one believes humans are arbitrary Turing machines, one must commit to the idea that one is not "better" than anybody else. Some of us may have built our identities around being "smarter", "nicer", "humbler", or, in any case, "better" than others. If you are such a person, a painful examination of your identity may be required.

Just to drill in the point that there is every reason to believe humans are arbitrary Turing machines, and thus that there is every reason to stop computing any human you may be currently computing, a formal proof is waiting in the next section.

---

the car kills A with probability $1 - \frac{1}{2^2}$ and B with probability $\frac{1}{2^2}$. Three people? $\frac{1}{2^3}$. And so on. This way, our intuition that a car should almost definitely kill one person over a million people is safely preserved – the chances that it kills a million people is almost nil.

[7] This thought pattern is somewhat similar to Niezsche's *ressentiment*.

5 Why You Shouldn't Judge Just Anyone

Proof of proposition: A human H ought not to be computing an arbitrary human H'.

Assumption 1. *The Church-Turing Thesis is true: everything that is physically computable is computable by some Turing machine.*

Definition 1. *A human H is a thing that does computation and is in the physical world.*

Explanation: In other words, a human H is an automaton to which the Church-Turing Thesis applies; for each thought process of human H, there exists a Turing machine.

Definition 2. *A human H is free if and only if H is uncomputable.*

Corollary: A human H is not free if and only if H is computable.

Definition 3. *We say a human H "ought not to be" executing some Turing machine M in the case that H is not free if H is executing M.*

Proposition 1. *A human H is at most Turing-complete.*

Proof: This follows from Assumption 1, that the Church-Turing Thesis is true.

Proposition 2. *There exists no Turing machine M that computes the output of an arbitrary Turing machine A.*

Proof: This follows from the undecidability of the halting problem.

Proposition 3. *A human H cannot compute an uncomputable function.*

Proof: By Proposition 1, H cannot compute any function no Turing machine can compute. No Turing machine can compute an uncomputable function. Therefore H cannot compute an uncomputable function.

Definition 4. *An automaton S is said to be "stronger" than an automaton W if and only if the functions W can compute is a strict subset of the functions M can compute. Conversely, W is "weaker" than S if and only if S is said to be "stronger" than W.*

Explanation: This definition exists purely for the sake of linguistic convenience. In each subsequent proposition, replace "stronger" or "weaker" with the formal definition here.

Proposition 4. *For some automaton M, if M is computing the output of an arbitrary Turing machine A, M is either stronger than or weaker than a universal Turing machine.*

Proof: By Proposition 2, no Turing machine M computes the output of an arbitrary Turing machine A. Therefore, if M computes the output of an arbitrary Turing machine, M is not a Turing machine. In particular, M is not a universal Turing machine. There are two possibilities for M. (1) M is Turing-complete and has extra computing capabilities. For example, M may be a universal Turing machine with a halting problem oracle. (2) M is sub-Turing-complete, that is, there are Turing machines which M cannot simulate. Therefore, in this case, M is either stronger than or weaker than a universal Turing machine.

Proposition 5. *If a human H is computing the output of an arbitrary Turing machine A, H is weaker than a universal Turing machine.*

Proof: By Definition 1, a human H is an automaton. By Proposition 4, if an automaton H computes the output of an arbitrary Turing machine A, H is either stronger or weaker than a universal Turing machine. By Proposition 1, a human H is no stronger than a Turing-complete machine. Therefore, H is weaker than a universal Turing machine.

Proposition 6. *If a human H is computing the output of an arbitrary Turing machine A, H is computable by some Turing machine.*

Proof: By Proposition 5, if a human H is computing the output of an arbitrary Turing machine A, H is weaker than a universal Turing machine. Therefore, H is a sub-Turing-complete machine.

Lemma 1: There exists a Turing machine that can compute the outcome of any sub-Turing-complete machine. *Proof is left as an exercise for the reader.*

By Lemma 1, if H is a sub-Turing complete machine, H is computable by some Turing machine.

Proposition 7. *If a human H is computing the output of an arbitrary Turing machine A, H is not free.*

Proof: By Proposition 6, if a human H is computing the output of an arbitrary Turing machine A, H is computable by some Turing machine. By the corollary to Definition 2, if H is computable, H is not free.

Proposition 8. *If a human H is computing a free human H', H is not free.*

Proof: By Definition 2, a human H' is free if and only if H' is uncomputable. By Proposition 2, H' is at most Turing-complete. Because H' is uncomputable, H' must be at least Turing-complete. Therefore, H' is exactly Turing-complete. To compute the output of a Turing-complete machine is tantamount to computing the output of an arbitrary Turing machine. By Proposition 7, if a human H computes the output of an arbitrary Turing machine A, H is not free. Therefore, if a human H computes the output of a Turing-complete machine H', H is not free. Therefore, if a human H computes a free human H', H is not free.

So far, so good. However, at this point, one problem remains. A human H may compute some H' and simply claim that H' is not free, therefore H is free. But how should H know if H' is free or not? We show that there is no Turing machine to do just that. This lets us squeeze out a stronger result: *If a human H is computing an arbitrary human H', H is not free.* This formalizes the intuitive dictum, "all persons are innocent until proven guilty."

Proposition 9. *There is no Turing machine M that takes as input an arbitrary human H and outputs whether H is free or not.*

Proof: Suppose such a Turing machine M exists. Then M takes as input an automaton H and outputs whether H is an arbitrary Turing machine or not. If H were an arbitrary Turing machine, M could not know if H halts or not. If H were a sub-Turing-complete machine, then M can run H until it halts. Any Turing machine that halts can be simulated by a sub-Turing-complete machine. If H were to halt, H can be simulated by a sub-Turing-complete machine. Therefore M is equivalent to the solution to the halting problem. Therefore no M exists.

Remark: Clearly, there exists a Turing machine M that takes as input a human H with a specific semantic description – namely, that H is computing a free human H' – and outputs whether H is free or not: that Turing machine is described by Propositions 1-8. We may gain such a semantic description about H through, for example, something H has said or done. However, we are talking here about an *arbitrary* human H that may or may not possess this semantic description.[8] We have shown that, in this general case, there exists no such M.

Proposition 10. *If a human H is computing an arbitrary human H', H is not free.*

Proof: By Proposition 10, no Turing machine M exists that takes as input an arbitrary human H and outputs whether H is free or not. The rest of the proof mirrors the structure of the proof to Proposition 8.

Proposition 11. *A human H ought not to be computing an arbitrary human H'.*

Proof: This follows from Definition 3 and Proposition 10.

[8]Interestingly, by Rice's Theorem, there is no Turing machine that gives us any such semantic description.

6 Summary

Moral philosophy is about what one ought to do. Over the semester, I have tried to describe a worldview, a view of the world in terms of computation. I tried to show how this worldview would change how you think, how it would change what you believe you ought to do. So, this was a course about moral philosophy. Ironically, to talk about moral philosophy is, to some, something one ought not to do. It is obscene. It is like lifting up a skirt, or failing to zip up the frontmatter after urinating. Why talk about moral philosophy? Why lift the curtain? We are doing just fine under its backdrop.

The problem is when the hegemonic moral philosophy is not sufficient, and the cracks of its insufficiency cause some people to not do so fine. Falling through the cracks, they become either depressed or story-tellers. This course is a story about my almost lifelong confusion with, and embarrassment by, two hegemonic moral philosophies: the American one, and the Korean one. I have been confused for a long time, a large part of which I attribute to the fact that I moved back and forth from the USA to Korea multiple times. I lived there half my life, and lived here half my life[9]. I have been confused on what I *ought to do*, over and over again, in large part because the cultures of the two countries, or more precisely their moral philosophies, are almost contradictory.

The thesis is that they are not contradictory, that there is a particular way to reconcile them, and that the resulting moral philosophy is the one that ought to become hegemonic. This is not a new idea. Many comparative philosophers have spilled lots of ink over it.[10] The new idea is just one of perspective. The new idea is that ideas from theoretical computer science, and philosophy of computation in general, can be interpreted in a particular way to yield the aforementioned thesis.

At the same time, this course was about my fiancé. She has fibromyalgia, a disease, which, among other things, amplifies pain, and translates mental stress to bodily pain. The mind-body problem is a strange thing to have a vendetta against, yet I have just that vendetta. As the poet, rape victim, and subsequent fibromylagia sufferer Amy Berkowitz[11] says, "I've found that some fibromyalgia patients themselves refuse to believe the mind-body connection because they don't want to think "it's all in ther head". ... Trauma is nonlinear". In this course, I attempted to give her another voice, formalizing her poetic word "nonlinear" with the formal and precise, and thus befitting a different audience, word "uncomputable".

This is where the thesis extends to issues of social justice. Since moral philosophy tells people what they ought to do, it is precisely the failure of the hegemonic moral philosophy which causes people to do what is ostensibly wrong. This failure gives the social justice warrior reason to fight. The patriarchy is a system that valorizes men for their ability to perform computation, and makes women the object of that computation. Therefore women lose their voice; they are deemed incapable of producing anything new, anything uncomputable, because they are incessantly computed. Racism is the computation that takes a syntactic feature of a person and outputs a semantic feature of the person. Therefore the objects of racism lose their voice; they are deemed incapable of producing anything new, anything uncomputable, because they are incessantly computed. Colonialism functions in much the same way. However, these oppressive systems – the patriarchy, racism, colonialism – are *not* arbitrary Turing machines. Which means they are not uncomputable. In fact, they are readily identifiable Turing machines which may be described by a sub-Turing-complete machine. Which means they are computable. Which means we can compute the perpetrator of those oppressive systems. Which means they are not free. Which is why it is possible to stop those oppressive Turing machines on their tracks, and why it is our imperative to reason with the perpetrators so that they may also become free.

But if you have already accepted the thesis, there is no reason to read any of that jargonic talk. Using computer science to talk about moral philosophy is a sort of perversion. In a sense, all philosophy is a sort of perversion. As a smartypants once said, the purpose of philosophy is the dissolution of philosophy. I know at least a dozen grandmothers and grandfathers, most of them selling fish at a street market, who know everything this course can say and more. The audience I have in mind are the cynics, the highly educated, the "rationalists" who have retreated to their enclave, who refuse to believe anything that cannot be proven, who endorse things like utilitarianism, behaviorism, and *The Bell Curve*.[12] I believe I can change their minds

[9] This sentence is useful, becuase it is true whether I happen to be at the USA or Korea at the moment.

[10] For one of the first treatments on this topic, see *The Meeting of East and West: An Inquiry Concerning World Understanding* by F.S.C. Northrop.

[11] She wrote a book called *Tender Points*. In a sense, her book says exactly what I have said the entire semester, and more.

[12] *The Bell Curve: Intelligence and Class Structure in American Life*, by Hernnstein and Murray of Harvard and MIT, is a

because they are rational, and rationality is an admirable ontological property. Rationality, for all its faults, does one job very well: when proven wrong, it clips off, however much it hurts, that irrational cancerous outgrowth, the misapplication of ego.[13] What this course has tried to do is to show that the Modern Scientific World View, and its moral philosophy, which purports to be based on rationality, is utterly irrational. I tried to show this using something every "rationalist" would agree as a method for achieving rational truth: theoretical computer science.

That is not to say that this course could *prove* that the "rationalist"'s moral philosophy is wrong, and could change their philosophy accordingly. Nothing can do that. While the mathematical proofs in this course are sound, this course is primarily about interpretations of those proofs. And interpretations are not proof-proof. But as Wittgenstein may remind us, *whereof one cannot speak, thereof one must be silent.*

The course originally had two titles. The first title – "Culture, Computation, Morality" – was picked because the course is about how the three intertwine. The course is about the derivation of culture and morality in terms of computation. A central problem in comparative philosophy is the problem of translation. Given two cultures with wildly contradictory moral philosophies and wildly different languages, how can one know what their philosophies even mean? While many methods have been established, the method I have pushed in this course is to use the language of theoretical computer science as a sort of universal language. I have tried to describe what that looks like.

The current course title – "The Poetry of Computer Science, the Computer Science of Poetry" – apparently makes no sense, as many people have told me: what do poetry and computer science have anything to do with each other? As is known, poetry is *subjective*, and computer science is *objective*. Poetry is *private*, and computer science is *public*. And there seems to be an irreconcilable gap between the two. This course argues that this perceived irreconcilability is mistaken.[14] When I say there is no irreconcilable gap between the subjective and the objective, it is not to say that one reduces to the other. Computer science does not reduce to poetry, nor does poetry reduce to computer science.

The alert student would be right to be severely creeped out by any attempt at such reduction, because in the Modern Capitalistic World, the primary application of computer science is to *compute people*. Giant technological conglomerates predict, determine, and follow their users' every flick of a hand, every toss of a foot. So the alert student is right to have a gut abhorrence against mixing poetry with computer science, if it means *yielding* poetry to computer science, if it means that nothing is sacred, if it means that our entire humanity can be subject to computation, manipulation, monetization. But that is precisely the opposite of what I wish to say.

Capitalism facilitates the computation of persons. And sometimes, there is nothing wrong with that: it can enable the fast reduction of those evils easily "computed away", that is, evils that can be destroyed with an efficient algorithm, such as the lack of food, water, and shelter, in short, basic necessities. As the old Korean proverb goes, a stocked granary – a computable good – is the basis for humanity – an uncomputable good. On the other hand, capitalism's overconfidence in the power of computation seeps into where it should not, such as love, education, and how to "make the world a better place", in short, abstract human goods. These are problems which cannot be solved with any efficient algorithm, which, foolishly, are again and again tackled by such. In Silicon Valley, billions of dollars are spent every day, by some of the most educated, most purportedly rational people in the world, irrationally trying to solve uncomputable problems through computable solutions.

But the theory of computation has already provided us with what exactly can, and cannot, be computed. The central thesis of this course, reiterated, is that to live the good life is to compute exactly what can be computed and to not compute exactly what cannot be computed. The thesis connects to existing ideas in moral philosophy through a close isomorphism to Kant's moral philosophy, which, as many comparative philosophers have established, is closely isomorphic to Confucius's moral philosophy. The thesis is therefore also the convergence of Western and Eastern philosophy. In its shortest form, it says: *don't judge anyone.*

book that engages in what I believe is "scientific" racism. It examines IQ levels across different ethnicities and provides social policy guidelines based off that.

[13] Actually, I am not so sure about that. I might have too much faith in rationality. But I can only try.

[14] Again, this is not a new idea. Heidegger and Wittgenstein, among others, have argued for something similar. See *Being and Time* and *Philosophical Investigations*.