[MUSIC PLAYING] All right.

I hope you all had a good break-- spring break, as they call it, uncreatively.

I certainly had one.

Didn't do any work.

That's for sure.

Why would they put this philosophy class in an engineering building?

Hold on one moment, please.

I'm speaking, and I will get to you when it's your turn.

I am Professor Breck Knowles.

This is Advanced Ethics.

You should be upper-class philosophy majors.

That should be all right in here.

Can you please wait your turn?

Please.

This is the textbook.

Now, this textbook covers some of the most important topics across ethics, edited, of course, by yours truly.

[CHUCKLES] I expect you all to have it, and you can pull that-- none of you have it.

[SIGHS] The bookstore must have really messed up this time.

No.

When you do get it, make sure you get the new latest edition.

I don't get my cut for used books.

Now, if you turn the page 432, that's my chapter.

Could you read?

[CLEARS THROAT] Chapter 18, "Bias is Bad," by Brack Know-less.

Knowles!

It's Knowles.

[SIGHS] You surely don't know how to impress your professors.

You read.

When investigating the societal implications of bias within the framework of non-hierarchical dependencies, we must integrate that the taxonomy of bias and inequity are combinationally intertwined.

First and foremost, bias affects both the sub- super-conscious-- Apologies for being late.

Computer systems ask me to help fix a glitch that happened this morning and affected which classrooms a few professors are sent to.

Hi.

I'm Sophie Muller.

I don't think we've met.

That's what I was trying to tell Professor Know-less, that he's in the wrong classroom.

My apologies.

The problem with the software upgrade this morning, but it's all fixed.

This is not the advanced class?

No.

That's what I was trying to tell you.

We are in engineering ethics.

Well, I always am supposed to teach the advanced class.

I don't want to be stuck with beginners.

But where am I supposed to go?

Let's see.

Knowles?

Your class is all of the way back in the philosophy floor in the subbasement of building seven, room 001.

A bit of a hike, I'm afraid.

[CHUCKLES] Always get stuck in the cave.

Excuse me.

Do you want your book back?

Thank you.

You engineers can make something that US philosophers deserve.

Just-- [LAUGHTER] [MUSIC PLAYING] How is everyone doing?

Oh.

I guess I didn't introduce myself to all of you.

I'm Sophie Muller.

I'll be taking over from Professor Baer, who has been asked to run an important project on the ethics of AI.

Did you have a good spring break?

Yeah.

That's good, but the registration list said 40 students.

Where is everyone else?

Yeah.

Well, the others went to go hear the president speak about the new administrative organizational structure.

He's changed it again.

Third time this year.

Mhm.

The university president.

Sounds interesting.

[CHUCKLING] But I thought students almost never go to such things.

Yes.

But since no one goes to these and the president didn't want to speak in front of an empty auditorium again, he offered everyone free food and drink and allowed them to break the rules and eat and drink in the auditorium.

He also hired a large steam cleaning crew, complete with industrial drying equipment to clean up right afterwards.

I see.

Oh.

I realized that I introduced myself but got distracted by our conversation and didn't have you introduce yourselves.

What are your names?

Alyssa.

I'm Patch.

My name is Sophie.

Alvis.

Glenda.

Pleased to meet you.

Well, nice to meet all of you too.

So, you were telling me about the event going on right now.

But why aren't you five there?

Yes, Sophie.

Well, professor, don't tell anyone else, but we're really dedicated to understanding the ethical issues of bias in AI.

In fact, instead of traveling over spring break, we just stayed here to discuss ethics altogether.

And we looked at the resources that we could find online.

Your dedication is admirable.

What did you find in your research?

Well, there are a bunch of videos and articles online, but-- But what?

We found them unsatisfying.

Why is that?

Alyssa.

Well, not everyone found them unsatisfying.

I thought they made some good points.

Well, you weren't able to explain away their failings.

Were you?

No, but you weren't able to give a better approach.

Were you?

Hold on, you two.

Always bickering.

We didn't answer professor's question.

Professor Muller, we found them unsatisfying because they didn't give us an answer.

That's right.

They told us that it's a problem, and they said that it's a hard problem, but they didn't tell us how to address it.

It's more than that.

The videos on bias in AI are all the same.

They all say it's in the data set-- bias in, bias out-- and they keep recycling the same examples.

Some videos have better graphics or better speakers, but other than that, they're all the same.

[SCOFFS] So conformist.

Yeah.

And while they say that bias is in the data set, there seems to be no general solution.

They just indicate that for each algorithm the issue of bias needs to be researched.

Even for that, they don't give a systematic approach.

Yeah.

Even if you could choose the right data set and quantify it on that data set to ensure no bias in your training set for a given analysis, you have no way to know if that applies to all situations.

My point is that we all agree that bias is bad.

We simply need to address it in every single algorithm that we make.

Yes, I agree.

We don't want any bias in our AI.

We want everyone to be happy.

This is a good start, but perhaps it's best if we get to specifics.

Alyssa, what were some of the good points regarding algorithmic bias that you mentioned?

For example, algorithms are biased against underrepresented minorities.

Facial recognition doesn't work as well for Blacks as it does for Whites, especially Black women.

One common example shows that while it identifies White males nearly perfectly, it misidentified Black woman over 20% of the time.

Well, one algorithm gives 20% error and the other 30%.

They can't even be consistent in their bias.

If we can't nail down the extent of the problem, how do we fix it?

You could at least work to try to make sure that the data set is more representative.

Oh, and what about the algorithm that identifies Indian brides as actors?

Well, why should they complain?

Being identified as an actor is the greatest honor anyone can have.

Oh, very funny.

I like chicken sandwiches.

How would you like it if an algorithm identified you as a chicken sandwich?

OK.

I see why they call it an adversarial example.

In any case, it's not easy to address bias across the board because of skews in data sets due to historical inequities.

And those skews differ.

At least we know that there have been historical inequities that we need to address.

And while it may not be easy to do, we have to at least try, even if it's on a case-by-case basis.

Even if it is on a case-by-case basis, we're doomed to miss a lot of bias until long after everyone is equal.

And I for one am not that patient.

Well, this problem is so important that even if it takes a very long time and even if we have to do it algorithm by algorithm, that is what we should do.

If we all do this, everyone will be happy.

But no, many people won't be happy waiting for the bias to go away.

So, professor, this is where we got stuck over spring break.

It's just not practical to engineer every algorithm against bias.

Even if we could, we would end up eliminating it in a few dimensions.

But then we're bound to miss it in other important dimensions.

Above and beyond that, we can't even quantify the bias consistently since different studies on the same data set give different amounts of bias.

We know that we need to do something to address bias, but what the videos say will at best take a small step forward.

We know that others may be content with that, but it's not sufficient for us to feel good about ourselves.

No, it's not.

Ah.

You have reached an impasse, which, believe it or not, is often a necessary place to be before you can make progress.

I think, however, it's best if we take a step back and explore how these algorithms work and the roots of wanting to eliminate bias.

[MUSIC PLAYING] So you probably all know this, but just to remind you, the overall approach is simple.

You start out with a training set of data here together with descriptors, which we use to derive a model.

Then we input the data, and the model classifies it based on some criterion.

Obviously, the devil is in the details.

Based on this high-level approach, where do you see bias?

There's no bias here.

The algorithm-- say, a convolutional neural network-- objectively analyzes the data provided to it.

It's just math.

It is not just math.

We choose the training set.

We choose the algorithm.

Even if we think that we are not biased, we are subconsciously biased.

How do we actually choose them if everything is described by mathematics?

We don't really have a choice.

Duh.

The bias is already in the data due to historical inequity.

You said it yourself.

So then our choice is to note that and add an appropriate loss function.

Does everybody know what a loss function is?

Of course.

Yes.

Yeah.

A loss function is just the error function associated with an event, such as facial recognition.

It is what should be mathematically minimized.

Good.

Just to clarify, it shouldn't be simply minimized across the board but targeted, for example, so that the overall error in facial recognition is minimized with the constraint that it is equal among different groups.

But unless you have equivalent training sets for the different groups, the statistical errors won't be equivalent.

And you can't quantify the errors in advance to determine whether or not they will be equivalent.

In other words, you can't choose the loss function approach to solve the problem because the

fluctuations or randomness.

The problem is statistics.

We have to have a very large data set to average out randomness to acceptably small percentages and make sure that our data is representative.

Yes, but my point is that it's just not possible.

I agree that we could make the error rates in facial recognition algorithms equal for Whites and Blacks and for men and women.

But what about all the other racial and ethnic minorities?

And for that matter, what about all the different genders?

Even if we could agree on what groups need to have bias eliminated, we can't practically address them all.

Yeah, and that's not even how the field works.

No one's going to spend five or 10 years for each algorithm to eliminate bias.

For cutting-edge research, people just try to hit 80%, sometimes 90% accuracy for a given algorithm and then publish a paper on it.

The incentives are such that they can't spend years getting 99% or so accuracy across the board, at least for new algorithms.

There just isn't enough data.

Better to just get something out there, even if it's not so good, so you at least get credit for it.

If we could only eliminate randomness, our problems would be solved.

But randomness is good because it's the basis of choice, otherwise our minds would just follow a deterministic path based on our neural pathways and electrical signals.

The heart of it is quantum mechanics, which introduces intrinsic randomness.

Well, randomness doesn't give you choice, just chaos.

That's also the reason we have bias to begin with, because of random historical circumstances.

We want to eliminate bias, not just make it random.

If it is random, then maybe you will mitigate it for minorities, or maybe, at least for some, you'll find a way to augment it.

[SCOFFS] That is what I was saying, too.

So we seem to be stuck again.

Mathematics is the rigorous way to understand the world, but it doesn't give us choice.

In the world described by mathematics, choice is an illusion.

Professor, we ran into the same impasse when we had this discussion last week during spring break.

But how do we get out of it?

I think I can be of some help here.

So you all agree that it is desirable to eliminate bias, right?

Yeah.

Yes.

So you all think it is possible to mitigate bias depending on what actions you take?

Yeah.

Yeah.

So you do think that can make choices and not just randomly.

More well-grounded choices are better than less well-grounded ones, say, for addressing bias.

I don't think we really have a choice about things.

People might think that they do, but they don't really.

We can't choose where we are born or how smart we are, and we can't choose if we get hit by a car.

We think we are making a choice, but we really aren't.

It's like Sophie said.

It's all just an illusion.

Yeah.

Yeah, you're right.

Randomness doesn't help.

You've convinced me.

Wait.

Did I have a choice, or was I compelled to agree?

No.

Your neurons told you to agree.

Yes, exactly.

And randomness, a consequence of a chaotic quantum process, just makes what we do not predictable but doesn't mean that we choose what we do.

Right.

Well, you have a choice whether you're going to be biased or not.

Mhm.

Indeed.

It does seem that judgments involve choice.

But Alyssa, if you were given the full understanding of a situation, wouldn't you always take the best action?

Yeah, of course.

No.

You can't really choose whether anything is better or worse.

It's all just a value judgment.

You may be compelled to do something by others, but you never really freely choose.

Do you really believe that, or are you just playing devil's advocate?

OK.

A little of both.

But even if I am just playing devil's advocate, what's wrong with what I'm saying?

What's wrong is that your model of human life is inadequate.

It doesn't describe what it really is.

We're not just some balls moving around deterministically with random noise interspersed like a term in a Langevin equation.

Isn't that true, Professor Muller?

What a great discussion this is.

Sounds like you had a much better spring break discussing these questions than going to the beach.

[CHUCKLES] I think you're on the right track, Sophie.

Funny.

Coincidentally, that's my first name too.

At any rate, you have unveiled a great tension in modern science.

We are brought up to think that modern science is the way to describe the world, but it seems inadequate to describe human things.

Ethical questions cannot be reducible to mathematics-- similar to justice and beauty, hope and suffering, and, for that matter, even truth.

Yes, Glenda.

Aren't those things even more important than modern science?

I would say so.

No.

Those things are arbitrary or a matter of individual's interests-- what's in it for them.

That's the way to look at them scientifically.

Bias exists because people have conflicting interests.

Sure, if you look at it purely scientifically, but interest alone cannot be adequate since people are willing to sacrifice interests for passions and, more importantly, for things beyond themselves.

At any rate, do most of us really know all of our interests?

We all have preferences that go far beyond interests.

Perhaps we can say at best that those interests are intertwined with different approaches to the world.

Besides, aren't the most boring people the ones who say, I'm right, and, even if one can never make value judgments, I have my own values and I stick with them?

[CHUCKLES] Yeah, these are the most irritating and boring people.

So, now let's dig down a bit more into the mathematics of AI and the sorting problem.

Perhaps it's best to start with a specific example.

What's the reason for the bias in facial recognition that we discussed a little while ago?

A biased training set due to historical inequity.

OK.

So one of the issues is the training set.

The other issue is the algorithm, including its hyperparameters, all leading to how the classification is done.

This is why different algorithms give different error rates using the same training data.

Anything else?

Right.

So those are the two key issues.

However, within the choice of algorithm, there is the problem with classification.

Let me project this data set, here with points in red and blue.

You have many choices for putting the dividing surface.

But what if we choose the wrong one?

Some of the data may be misclassified, won't it?

You mean like this?

Yes, exactly.

So how do you choose the right line ahead of time?

That is a problem.

It could be even worse if your data set looks like this.

But then we don't have to stick with lines.

Why don't you draw a curve around them?

Like this?

Yes, exactly.

Now you've classified them all perfectly.

Well, wait a second.

If you do that, it works for the training set, but it may not work for any additional data sets.

In fact, it might be even worse than the lines because you're overfitting to the training data.

Do you mean like this?

Yeah.

That doesn't work.

Oh, yes.

I see it.

Exactly.

Remember this is a sorting algorithm, here in two dimensions with two descriptors, thus it is quite limited.

But even if you do it in a very large number of dimensions with a very large number of descriptors, you are still projecting a complicated system-- shall we say a natural system-- onto an artificial mathematical construct that necessarily leaves much out of the description.

Or if you want to stay in the realm of mathematics, you could say that due to a necessary incomplete set of descriptors you were fitting noise.

This allows you to get a perfect fit for your training set, but then you have an even worse fit for your data set moving forward.

Patch.

In other words, we have to accept the fact that there will be errors.

But at least we could distribute the errors evenly among groups.

That is what Alyssa said before.

We can try, but then I'm worried we won't include all the groups.

Even if we could, there'll be errors in the errors.

Unless we have a very, very large training set, which practically we never have, we won't be able to do that.

Yes.

The training set would have to incorporate all the data in the world.

You know, that reminds me of a story I once heard, I think from a South American writer, where someone makes a one-to-one scale map of their country with all the details included.

[CHUCKLES] Ha, ha.

Well, I grant the problem.

But then what do we do?

We all agree that we need to eliminate bias, but it seems so hard.

What if we could at least have a very extensive model that for all practical purposes captures a given situation?

Which I guess is actually impossible because it would require a complete and accurate model of all the objects in the whole world.

Yes.

As engineers, we can at least work to do that.

If the model is extensive enough, we can eliminate bias.

But then we'd be eliminating any choice that we might have had.

We would be controlled by an algorithm.

That sounds-- that sounds tyrannical.

Indeed.

Mathematization of human things distorts them so much that they become something which they are not.

This is actually the utilitarian approach.

Human things like joy, anger, justice, honor, beauty, and happiness are really not mathematizable.

We lose all of what is important in them when we mathematize them.

We could in principle try to do so and therein transform human reality.

Glenda.

That doesn't sound appealing.

It sounds dreadful.

Yeah, that wouldn't be good.

But we can at least do it partially, can't we?

Yes.

Maybe that would be a good solution.

Yes.

Even if those human things cannot be described by mathematics, it is still the only rigorous way to understand the world.

I'm not so sure anymore, but the bias is in the data set, and it is in the algorithms.

So we must be able to adjust our models to solve the problem.

Mustn't we?

Well, consider this-- our recognition of the problem of bias is not mathematical.

Mathematics comes only after such recognition.

Mathematics is something we impose on our broader understanding of the world.

We can use it to develop models, but they will always be just that, models which are necessarily incomplete descriptions.

I see what you're saying, I think.

Sophie.

But AI is based on mathematics.

So we have to fix this problem in the algorithms via fixing the algorithms, i.e.

by using mathematics.

Don't we?

Well, we can think of things this way, but that will lead us to a double whammy.

We cannot develop a non-arbitrary definition of the criteria for the training set, and, two, we cannot develop a non-arbitrary definition of the error target for the sorting algorithm.

Well, I knew that we were at an impasse, but I didn't think it was this big of one.

Professor, your class has made us more stuck than we were before.

So, what do we do?

Is there another way?

Indeed, Glenda, there is that.

It requires a reconsideration of our concerns and of the issue.

[MUSIC PLAYING] [DOOR CLOSES] Thank you.

Sophie.

So, what is the other way?

Well, let's think about it.

We just showed that because of the intrinsic incompleteness of models eliminating bias in all dimensions is mathematically impossible.

We could no doubt decrease bias among some groups to a certain degree, with an acceptably small but still significant margin of error.

Even if we got rid of bias in some predefined dimensions, we would still have it in other dimensions.

But that's precisely it.

We want equality in at least some dimensions, the important ones.

That is good.

How do you choose which dimensions are important?

Alvis.

One important dimension is wealth.

Good.

So even you five, the top engineering students in your class at the top engineering school, should get paid just as much as someone like Dean, the paper collector.

Well, he is a very nice fellow, if a bit strange.

But no, we think we should be rewarded for the contributions that we make to society.

But wealth should be distributed more equally.

That's right.

But how do you decide how equal it should be?

Perhaps the problem is once again that mathematics in its homogenization of kinds is not the right way to address the problem.

Let me ask you this-- is it really equality that you want, or is it justice?

Justice.

Yeah.

I'll go with justice.

Great that you two agree on something.

Glenda.

They should go hand in hand, justice and equality.

Yeah.

Isn't equality justice?

I think we need to add another consideration, since we are beginning to delve into the complex human questions underlying our definition and perception of bias.

So let me ask you this.

Which is most important-- equality, justice, or happiness?

Sophie.

I would say happiness.

Even if someone has more than you do, and even if they don't deserve it, it doesn't matter to you as long as you are happy.

Think of all the people born into rich families.

They're typically not happy since they don't need to have a goal, and therefore they don't have one.

I would still have a goal if I were born into a rich family.

Why do they need a goal if they're having such a good time?

Because people need to have a purpose to be happy.

And how can people be happy if they're not treated fairly?

So we have equality, justice, and happiness, but it is unclear which is most important.

Glenda.

Can't they be of equal importance?

Actually, no.

They are certainly interrelated.

But since they are at least somewhat different and since, as we saw, they can conflict, one of them has to be most important, has to take precedence over the other.

I think that justice is more important than equality.

After all, we don't want everyone to be the exact same, but still justice and equality are related.

People can be unequal because they're treated unfairly.

They can also be unequal due to luck, good or bad.

I don't think that happiness can be the standard.

We need equality, and that's why mathematics is still so important, because we can employ it rigorously to make everyone equal.

But if everyone is the same, wouldn't the world be drab?

I mean, what if all of us were the same?

We wouldn't even be able to have this dialogue, as we'd all be making the same point.

There would be nothing interesting, as everyone is reduced to a forced homogeneity.

Besides, how do you make everyone equal without violating their rights?

We're going to have to violate some rights to make people at least more equal.

But then maybe I guess that's not justice.

Yeah.

Isn't it the essence of tyranny when rights are violated in the name of justice?

Yes.

We have strong reasons not to want that.

We want equality, but we certainly don't want to violate people's rights to get it.

But what do we do?

Here we are at an impasse again.

How do we understand this conflict between equality and rights?

By the way, professor, what about that other way that you promised?

[MUSIC PLAYING] I was wondering when you would want to go to that.

Do you remember learning about modern natural right from Professor Baer before spring break?

Yes, a bit.

We studied Locke's state of nature and its connection with natural right, if that's what you're talking about.

Professor Baer also made the case that modern natural right is the basis for the American regime and the reason for this country's success.

Good.

And yes, that is exactly what I mean.

It's great that you remember reading Locke and about modern natural right being the basis for the American regime.

Oh.

Hello, Professor Knowles.

What are you doing here?

I hope I'm not bothering you.

I heard that you were discussing modern natural right, and I wanted to join.

The stuff you heard me spew earlier was a front that I have to put forward to my colleagues in the philosophy department, and some students who appreciate such things.

It's the only way I could keep a job, which I need to be able to pursue true philosophy.

That is refreshing to hear.

And yes, please do join us.

We were just starting on modern natural right.

So how does Locke view the problem of AI and bias?

Patch.

Locke would think it wrong because bias violates your natural right.

Yes, I agree.

That sounds right, but I'm not sure because we said that justice is more important than equality.

But it seems that equality is fundamental in Locke.

But then where is justice in Locke?

Let's take this step by step.

Does anyone remember the beginning?

But you said you were working on these questions during spring break.

Alvis?

Well, we were discussing them, but we never actually went back to the texts from class.

Yeah.

We just wanted to figure them out for ourselves.

I guess you disapprove, Professor Muller?

On the contrary.

You are a very impressive group to have stayed here over spring break and investigated these questions.

Any teacher would be thrilled to have even one student like you, and I have five.

But you said that you got stuck.

What was your plan to get out of your impasse?

Well, your class, of course.

[LAUGHS] Hush, Patch.

You seem not to realize that Professor Muller is in a most mild way rebuking us for not using the resources at hand.

Well, let's use them now.

Professor, please remind us of Locke's starting point.

Of course.

Let's turn to second treatise, chapter 1.

Professor Knowles, where do you think we should start?

I'd say the last section of paragraph one, starting with "he that." That is exactly what I was thinking.

Glenda, could you read the statement starting with "he that?" Of course.

He that will not give just occasion to think that all government in the world is the product of only force and violence must of necessity find out another rise of government.

Good.

So Locke is giving us two options of the origin of government, force or something else--maybe something else based on reason but not narrow or mathematical reason.

Yes, of course.

That is much better.

It would be terrible to think that politics is just about force.

I agree.

It should be about allowing people to be happy, as we said before.

But what does this alternative basis?

You can look at your books, you know.

Alvis.

It is the state of nature.

Good.

What is the state of nature?

Well, let's see.

I actually underlined this pretty extensively.

It is a state of perfect freedom and a state also of equality.

Yes.

I remember now.

But how can it be a state of perfect freedom if anyone can take your stuff and it's only up to you to protect it?

You don't remember?

That's the point.

There is no government, so we're all perfectly free in that sense.

But life is not desirable since we're all individually vulnerable.

That's why we form governments, for security.

Very good, Sophie, but you're moving too fast.

Remember Locke says but though this be a state of liberty, yet it is not a state of license.

Do you recall why that is?

Oh, hold on.

I have it right here.

Because of a law of nature.

Yes, that's it.

Laws of nature are based on reason.

Yeah.

We talked about that before spring break.

Now I remember.

It's dumb to waste stuff, including your own life.

What about these people that waste their money on luxuries, expensive cars and things?

Yeah.

I'd be content with a cheap rundown car.

Why should someone buy a Ferrari when most of us can't afford an old Kia?

Does anyone recall Locke's answer to this?

Perhaps, once again, we are moving too fast.

Let's go step by step.

What is the state of nature like?

Alvis.

In addition to freedom and equality, there is a law.

Good.

What else?

There is punishment.

In paragraph seven, he says everyone has a right to punish the transgressors of that law to such a degree as may hinder its violation.

Very good, Sophie.

I think we will see that our objection to bias is itself based on a sense that there is a law of nature.

How does Locke characterize this statement?

He says in paragraph nine that this will seem a very strange doctrine to some.

Excellent.

Does it seem strange to you?

Well, not any more strange than the state of nature.

Why is the state of nature strange?

So nature is what you experience when you go on a hike or go camping.

What Locke is describing never really existed.

What do you think existed in its place?

So humans evolved from humanoids, and they lived in tribes.

They weren't by themselves.

Right, Patch.

But what is the basis of the rule of the tribe leader?

Power.

But isn't this exactly what Locke said at the beginning?

We can have tribal rule based on power, or we can have an alternative, which Locke provides.

Besides, don't you think human beings' natural state is freedom and equality?

That all makes sense.

But then why does Locke call it a very strange doctrine?

To some.

See, I think that will find that what is required is not so much a new doctrine, say a learned latent structure for algorithmic bias mitigation or domain adaptation, but a reflection on what doctrine some find strange and others do not.

What's strange to us actually contradicts our own biases.

Psychologically, we tend to resist those things because they point out our biases to ourselves.

Oh, I see.

What we find strange is an indication of our own bias.

That's what Locke is saying.

Thank you, Professor Knowles.

And I realize that he defends this doctrine by arguing that governments can punish foreigners who break a law, which is related to the example in paragraph 14, when he says that by the state of nature exists today, since all princes and rulers of independent governments all through the world are in a state of nature.

Very good, and then he brings in the next paragraph the weight of the writings of Hooker, or shall we say the judicious Hooker.

And he ends the chapter with the idea of consent.

So political society is made legitimate only by consent.

Well, since we are short on time, I'll have to move on to chapter five on property.

But explore when you can what he says about the state of war and slavery in the Skip chapters.

We will also come back to other subtleties, like why he calls the doctrine by everyone in the state of nature has executive power this strange doctrine in paragraph 13 and, for that matter, why there is no authority in the state of nature besides each individual.

So what are his key arguments on property?

Alyssa.

One is that each of us has a property in our own person.

That's in paragraph 27.

Good.

And later in paragraph 27, that we make things our property by mixing our labor with things from nature.

Very good.

Yeah.

I like the acorn example.

By picking up an acorn, we make it ours.

I guess that applies to fruit, too.

[LAUGHTER] And he says in paragraph 26 that the Earth is ours for the support and comfort of our being and that we're given reason to use it well.

That brings me back to my point from before.

Someone isn't using it well if they're rich and they hoard things.

Well, you can only hoard so much.

Like, who needs 10 Ferraris?

I do.

Ha, ha.

Ha, ha, yourself.

The point is that you cannot hoard because your things will spoil.

Yes.

That is why he talks about money somewhere.

Money doesn't go bad.

Yes, in paragraph 36.

But you'll see that he says in paragraph 31 that everyone can take as much as he wants from nature, but there is a limitation.

What is that limitation that prevents us from accumulating so much that things spoil?

Well, what is also present in nature besides things you can take?

Oh, yes.

The law of reason.

It's stupid to hoard because then you're just wasting stuff, and that's against the law of reason.

But once there is money, you can hoard it without spoiling.

So why isn't that wrong to do?

Because you put your labor into it?

But what if you're just lucky?

Or what if you inherited it?

OK if I answer this one?

It's because someone else put their labor into it, therefore they can give however they choose.

Excellent.

Thanks.

Do you recall how much Locke says labor increases the value of what is in nature?

100 times in paragraph 40.

1,000 times in paragraph 43.

Both 10 and 100 times in paragraph 37.

And since we mentioned paragraph 40, we should point out his famous statement in the subsequent one.

A king of a large and fruitful territory in America feeds, lodges, and is clad worse than a day laborer in England.

So whatever the factor of increase is, the point is not just that labor is almost all the value in property but that by protecting property we will all benefit with many good things.

Yes, like this university.

[CHUCKLES] Exactly.

I still don't think it's fair that some people have so much more than others.

But we all benefit.

Besides, if it weren't for this system, you wouldn't even be able to get that broken-down Kia that you dream of.

Oh.

Ha, ha.

[LAUGHTER] But even if the pie gets bigger and bigger, some people will still control others because they have a bigger piece, making those people unhappy.

Or do you mean less free?

If so, it seems that AI doesn't solve the problem at all.

Of course, some people do think that it can, that mathematization of all things can lead to a sort of techno utopia.

Locke is giving us a choice.

We can be controlled by algorithms designed by smart but unworldly people working long hours in some basement-- or let's call it a cave-- or we can be allowed our own choice.

Of course, that choice should be the product of thoughtful reflection, particularly on our own biases.

But AI could give more freedom, couldn't it?

I see what you're saying.

Certainly, it can bring benefits in health care and transportation and make many things more efficient.

But intrinsically it can't solve the problem of bias even in itself.

Actually, it's even worse.

Those who develop and therefore control AI will build AI systems biased for their own benefit.

Hmm.

Hmm.

Having said that, I realized that we somehow seem to have gotten distracted from understanding how natural right can help us to address bias in AI.

And we only have a few minutes of class left.

I was wondering when you were going to realize that.

Well, you convinced us that all AI systems will be biased in some dimensions no matter what you do.

But then you also convinced us that property should be respected.

So, what do we do about bias?

Yeah.

And how should we punish people who write unacceptably biased AI systems?

Or reward them for making acceptably biased AI systems.

I mean unbiased AI systems.

[LAUGHTER] Yeah.

We can't reward or punish people, according to Locke's system.

On the contrary, we can reward them by buying their products if we like them or punish them

by not buying their products if we don't like them.

So transparency is the key?

Yes, if people are willing to pay for transparency.

You see, the law of nature addresses that too.

And in most contemporary approaches, there is an assumption, one which we have been addressing today too, that bias is bad and should be eradicated by technical mathematical means.

We now understand that a deeper reflection on the law of nature and the law of reason is needed, but we engineers needed to see the limitations of the mathematical or computer science approach first before we turn to a broader or deeper one.

Yeah.

I think I understand this now, but it still hasn't changed one thing.

I want a Ferrari.

[SCOFFING, CHUCKLING] The point is under the Lockean system we are all better off-- happier, we could say-- because algorithms with unacceptable bias will go away because no one will buy them.

You're starting to understand.

Earlier, you mentioned utilitarianism, and you showed not only that AI is tied up in it but that it doesn't treat the human things well.

But what I didn't really understand is that we still need to make utilitarian choices, like allocation of resources.

Could you show us how to think about that necessity within the realm of human things?

Yes, I can.

We did say that utilitarianism as a form of thought tends to give an incomplete account of human concerns and passions, no matter how much we strive to make the mathematical model complete.

But we did indicate that engineering is utilitarian and that it always strives to optimize some function.

How we should think about that for human benefit would be the subject for next class.

[BELL RINGS] [MUSIC PLAYING]

1075 00:44:40,010 --> 00:45:45,000