

Philosophical Profiles

David Boonin

Professor and Chair of Philosophy at The University of Colorado Boulder

IN BRIEF

David Boonin is Professor and Chair of Philosophy at the University of Colorado, Boulder, where he has taught since 1998. He wrote his dissertation at the University of Pittsburgh under the direction of David Gauthier, who kindly assisted him in his rejection of Gauthier's own interpretation of Hobbes, work that led to *Hobbes and the Science of Moral Virtue* (Cambridge University Press, 1994). Thereafter his interests have taken him in pursuit of particular arguments that provoke him, beginning with Judith Jarvis Thomson's famous violinist analogy, which led to *A Defense of Abortion* (Cambridge, 2003) and more recently *Beyond Roe: Why Abortion Should be Legal Even if the Fetus is a Person* (Oxford University Press, 2019). His other books include *The Problem of Punishment* (Cambridge, 2008), *Should Race Matter?* (Cambridge, 2011), *The Non-identity Problem and the Ethics of Future People* (Oxford, 2014), *Dead Wrong: The Ethics of Posthumous Harm* (Oxford, 2019), and soon a book on sexual consent very shortly followed by a book on the ethics of artificial intelligence, telling us why we shouldn't worry *too much* about the coming Robopocalypse. He seems positively crestfallen if a position he defends is acceptable to the majority of his colleagues in academic philosophy.

DETAILS

Simon Cushing conducted the following interview with David Boonin on 22 August, 2022.

CITATION

Boonin, David. 2022. "Interview by Simon Cushing." *Journal of Cognition and Neuroethics* (Philosophical Profiles): 1–23.

David Boonin

a philosophical profile

SC: So, David Boonin, what first drew you to philosophy? What was the first, what was your first encounter with philosophy? What was the first issue? Maybe the first philosopher? What can you pick out as your origin story?

DB Yeah, well at the risk of sounding like a philosopher right off the bat, I think there were probably more than one first experience. So, can I give like a multi-part answer?

Absolutely.

Okay. Yeah. So I mean, I did have a pretty formative experience during my freshman year in college that I'll get to. But I think there were a couple of things before college that sort of primed me to get hooked on philosophy once I got there. I guess the first is my father was a philosophy professor. I don't know if you know, this. Was also a philosophy professor also at the University of Colorado, Boulder. So I grew up in Boulder, in the house of a philosophy professor. I don't think, looking back on it, I don't really remember talking about philosophy. I don't remember learning philosophical jargon. Probably the only philosophers I could have named, would've been philosophers that any high school student could have named. I have fairly vivid memories of when I was a fairly young kid, he would have some colleagues over for dinner, and I remember the experience of not understanding what they were saying and thinking that they were making up words as they went, and thinking it was some kind of game and I didn't quite understand the rules. So in that sense, I don't think I had sort of a head start in philosophy. But looking back on it, I would say, I think I was raised in a kind of Socratic household in the sense that I think we were always encouraged to think further about why we thought something. You know, he was always sort of asking questions. He's still always, if I go by to visit, almost always the first words out of his mouth are "Now David, I have a question." In that sense, I think I was kind of primed for philosophy.

The other experience was in high school. A friend of mine and I wanted to join the debate team, but it turned out that our high school didn't have a debate team. So we started one, but we didn't really fully know what we were doing. And we ended up participating in this form of debate called "on topic cross examination debate" where there's a topic a year. One year it was consumer products safety. We're supposed to do a bunch of research, and it was largely evidence-based. And then you show up and when you are on the negative side, the other team would present their case and you were expected to have facts at your disposal to try to rebut them. And then there were cross examinations like in a legal case. And we were generally woefully unprepared because we had just done relatively little research, and some of these teams would walk in with these moving trunks full of evidence cards, and so on. So we got pretty good at... the other team would give its case or whatever. And we typically didn't have any evidence we could spit out to rebut their claim that "the US government should introduce a new regulation of such and such." I don't know, "Are these things dangerous?" But we got pretty good at trying to pick holes in "your evidence card said such and such, but then

you said such and such, but what's the what's the reason for moving from this claim to that claim?" Or, you know, "you seem to be generalizing from one case," or, "that was a case where such and such was true, but this is a case where such and such." Yeah.

So laziness drew you to philosophy, as it does many people?

Yes. Laziness and ignorance I would say and I would say the lack of good funding from my high school. Maybe I shouldn't have said that last part.

I may use that when we advertise, for our major: "Are you lazy and ignorant? We have the major for you!"

Well, yeah, no. I mean, I have to say I did find it... we weren't like super successful, but we did well. But yeah, it was particularly gratifying to come out of one of those rounds. If we won anyhow, feeling like they have all these evidence cards and we had hardly any, but like we found holes in what they were saying and that was enough. And I think in retrospect, there was a little bit of blood sport to this form of debate. But I think I was particularly good at the cross-examination part. One of the members of the other team stands up, and then you try to punch holes in their argument, and then based on their responses, you try to keep it going. Yeah, So I think having the father is the philosophy professor always asking questions, and then this experience of seeing that you can do stuff and think through positions... I mean, evidence obviously is important, but there are things you can do that don't require evidence. You can find mistakes in reasoning that people are making, so forth. So I think I was kind of primed in that way. Anyhow, I got to Yale. My freshman year I was in this program called Directed Studies, which I think still exists there. It's a yearlong interdisciplinary seminar program. So you have a small seminar in philosophy, a small seminar in literature, a small seminar in history and politics. And they're all integrated. So, you know, when you're doing Plato in the philosophy class, you're doing Homer maybe in the literature class and Herodotus in the history and politics class, or maybe Thucydides. Anyhow. So I went to college planning to be an English major and my aspiration was to become a journalist. That was kind of my, my big ideal. But yeah, I think the first stuff we read in the philosophy class and we started with some of Plato's dialogues. I'm not sure if the *Protagoras* was the first one, but we read a couple of short ones and then some chunks of the *Republic*. And I think like in a lot of classes, Socrates is this kind of divisive figure and students either love him or hate him, and I just loved him. I just thought, "This is great! Look at what you can do." I think I wasn't very reflective at the time on what Plato was doing with Socrates. I think I was really just kind of looking at Socrates. And I don't think I was even particularly attuned to this whole notion of Socratic irony. I think, I think I have largely took it at face value. Look what he's doing. Like he cares about the truth. He sees the questions are difficult. He asks the experts and then he punches holes in their arguments.

He's Columbo, that's what I always used to say. I mean, he's disheveled. He acts like an idiot. And then, "I'm so stupid. Just explain it to me. I just have one question..."

Yeah. Yeah. Next time someone asks me this question, I'll have to add Colombo to my list as well.

“I went to college planning to be an English major and my aspiration was to become a journalist.”

Philosophical Profiles

I used to use that, but unfortunately none of my students know who the hell that is.

Well, I know who he is and maybe somebody listening to this will know. Yeah. So I think from that point on, I was I was pretty hooked.

[As to your question] I don't think it was a particular topic. I mean, I think the first paper I wrote was on something in the *Republic*, but, you know, we read a few different, whatever. There was stuff about politics, there was stuff about epistemology. It was really more, I think the methodology and just the sharpness and the crispness of kind of lining up: "Here's the question; here's an answer; here's a counterexample. Here's some attempt to finesse the counterexample. Here's why it doesn't work." Yeah, I was really hooked on that.

It is kind of a superpower really. I mean, if you hadn't become a philosopher, you would have become. Saul Goodman or something.

I know who that is too. I'd like to think. I wouldn't become at least a slightly less evil version of Saul.

Maybe his brother, Chuck.

I mentioned that my father was a philosopher. He previously had practiced law before he went into philosophy and my brother went into law. So between the two of us, we've kind of got the two angles covered.

So what did you write your dissertation on?

Yeah. Well, when I went to graduate school, my main interests were in history of philosophy. That was really what had kind of gotten me hooked and a lot of my undergraduate classes were in that area. And so my first year at Pitt, I took a seminar on Hobbes with Peter Machamer, who was in the history and philosophy of science department. That was a course that really focused primarily, almost exclusively, I guess on Hobbes' sort of natural philosophy stuff that I wasn't really familiar with. And then my second year, I took a seminar on Hobbes and Rousseau with David Gauthier. That was really focusing on the ethics and the politics. And I started seeing what I thought were pretty interesting connections between the natural philosophy and the normative philosophy. And seeing places where Hobbes seemed to be explicitly saying that that's how we should envision what he was doing. Gauthier wasn't particularly focusing on it. We were really just sort of doing the relevant parts of *Leviathan* and a few other works. Gauthier's book *Morals By Agreement* had been published fairly recently, and I started thinking--he was also a Hobbes scholar-- that he had his own kind of take on Hobbes. And I started thinking that Hobbes' moral theory was actually closer to Gauthier's moral theory than it was to Gauthier's interpretation of Hobbes' moral theory. Gauthier already took himself to be, and acknowledged himself to be, influenced by Hobbes. So my dissertation kind of grew out of those two thoughts, that, fundamentally, Hobbes was trying to ground a moral philosophy in his natural philosophy, as with Gauthier's own moral theory, but unlike Gauthier's interpretation of Hobbes' moral theory, there was this foundational thought that you can expect to do better in life if you internalize certain constraints on your behavior: honesty and promise keeping is kind of like the most straightforward example. And that ultimately what Hobbes thought was, I think,

“ I started thinking that Hobbes' moral theory was actually closer to Gauthier's moral theory than it was to Gauthier's interpretation of Hobbes' moral theory ”

more influenced by ancient thoughts about virtue than most people have thought. The upshot of my dissertation, which became my first book, was that actually Hobbes is best understood as a certain kind of virtue theorist in some ways not so different from other virtue theories. A virtuous person will flourish if they cultivate certain virtues. Some of the virtues were fairly traditional virtues things again, sort of like honesty and so forth. But the defense, the grounding of the virtues in flourishing, is much more minimalist and kind of materialist - like, you know, you'll have a better chance of surviving if you develop a reputation as an honest person, and the best way to develop a reputation as an honest person, is actually to *be* an honest person. And the best way to be an honest person is actually to have kind of come to cultivate the disposition to sort of tell the truth for its own sake. Keep your promises because "I'm a man of my word," that sort of thing. So that was my dissertation, and then my first book project after I graduated.

David [Gauthier] was great to work with especially since, like I said, my whole project was really sort of disagreeing with his interpretation - and disagreeing with other interpretations too - but I have a pretty horrific memory from graduate school. So I remember, I was helping in the graduate student lounge, which was four floors above main department office. And this was during the time that I was writing my dissertation. A student comes in and says, "Gauthier is looking for you." He was like running around the office, waving a copy of *Leviathan* and saying like, "Does anyone know where Boonin is?" And so I kind of hustle down the stairs, and I'm really nervous that he's going to have found some passage that just does me in, or whatever. But what he was excited about was that he had found a passage that *supported* my interpretation. It was one that I hadn't actually worked into the draft at that point. So he was just a delight to work with. And to me kind of the model of the philosopher who's developed his own strong views, but, you know, he's in it for the honest debate.

You are perhaps most well-known for writing on abortion. What first drew you to the topic?

So that's an interesting question. I remember pretty vividly becoming aware of the fact that I didn't have an opinion about abortion. This was during my first year in graduate school and a group of us were walking, I don't know, someplace on campus. And somebody was passing around a petition. A sort of pro-choice petition asking for signatures. And all my friends signed it. And I found myself thinking, "I don't really know that I want to sign." Like, I don't find this obvious. I haven't really thought much about it. But I had become a vegetarian my senior year in college. So I was kind of at least open to the thought that maybe beings that people don't think of as having moral status, maybe it turns out that they do, and maybe like that's enough to make certain things wrong. I was agnostic. I became aware that I really hadn't thought much about it. I suspect that would have continued to be the case for some time, but my second year in grad school, I was TA'ing for a political philosophy course. It was kind of history of political philosophy. Tamara Horowitz was the teacher, and I think she was lecturing on Rousseau. So "here's something Rousseau said about rights," or whatever. Then she sort of said as an aside or sort of an example, I don't remember exactly, but she just gave kind of a thirty-second summary of Judith Thomson's argument in her article called a "Defense of Abortion," which I had never encountered. I don't remember what the connection with Rousseau was supposed to be. But again, it was really just kind of an aside. Well, you imagine

“ I remember pretty vividly becoming aware of the fact that I didn't have an opinion about abortion. This was during my first year in graduate school and ... somebody was passing around a ... sort of pro-choice petition asking for signatures. And all my friends signed it. And I found myself thinking, "I don't really know that I want to sign." I don't find this obvious. ”

Philosophical Profiles

yourself in this situation: you wake up in the hospital, and you find that there's this violinist who mysteriously has become bodily connected to you, and they need to use your body for nine months to go on living. And Thomson says, as you know, surely the violinist has a right to life, but that doesn't give them a right to use your body. Right to life is not the same thing as the right to use another person's body. And she says "Thomson uses that to argue that abortion is permissible." And she says something to connect it to Rousseau. And then she goes back and keeps talking about Rousseau. And I remember just sitting there thinking, "I guess it turns out abortion is permissible after all!" So to me it was kind of a lightbulb moment. It was like, "oh yeah, there's a difference between the right to life, and the right to be kept alive, or the right to use someone else's body." And I hadn't previously thought about that. But that thought experiment convinces me that that's true. And now that I can see that it's true, you can just bypass this whole debate about whether the fetus is a person or whether it has a right to life. Maybe if you just see a fetus sitting on the street you shouldn't, like, shoot it in the head or something. But all you need to know to resolve the abortion issue is, you know, even if it has the right to life, that doesn't give it the right to use a pregnant woman's body if she doesn't want it to. But if the fetus uses her body, that's her right. End of story. So yeah, it was kind of a lightbulb moment, and I became interested in the issue. Not because of the *issue*, but because of the *argument*. And that's been true about pretty much everything I've written on. I've almost never gotten interested in a topic first and then thought, "let's find out what the arguments are about this topic." It's almost always been I encounter some argument. I think "wow, that's a weird argument." "That seems like a good argument," or "that's a puzzing argument."

That's strange, because you write on topics that people get very excited about, like racism, abortion, punishment, all of these things that people are crusaders about. And it's interesting that you are sort of above the fray.

I wouldn't know if I would say I am above the fray, but I guess I would say there are topics that I think are important social issues, where, at least so far, I haven't really seen interesting philosophical arguments or issues. Maybe because some of them do turn more heavily on empirical assumptions or whatever. So no, I mean, I tend to want to read around on issues that I think students will be interested in, typically, where there's something kind of socially interesting about the subject. But as far as what I end up writing about, it's almost always like discovering an argument first, and then that sort of driving the story. Anyhow. So I got to Georgetown at one-year job at Georgetown that turned over for a second year, that was when I finished turning my dissertation into a book. And in the meantime, I had been teaching these intro ethics courses that were about twelve or thirteen weeks of history of philosophy. And then a little bit of applied ethics at the end. And each time it was like so many more students kind of got energized and started talking when we got to the applied ethics stuff. And so it was right around the time that I was finishing the Hobbes project, I sort of felt like I want my research to be in alignment with my teaching, but I'd like to do a lot more teaching in this applied ethics area. And I remembered the Thomson article. And I think by that point I had read, Don Marquis's piece ["Why Abortion Is Immoral," *Journal of Philosophy*, 1989]. It was a little reckless, I think. I had no particular reason to think I could write a book about abortion that I hadn't really studied. But I just thought like, I'm going to pivot away

“That's been true about pretty much everything I've written on. I've almost never gotten interested in a topic first and then thought, "let's find out what the arguments are about this topic." It's almost always been I encounter some argument. I think "wow, that's a weird argument." "That seems like a good argument," or "that's a puzzing argument."”

from history and into Applied Ethics. And abortion was the main the main project that I worked on then.

Do you think your analysis of *McFall v. Shimp*, the court case that you use in our first book on abortion and in particular in this one, yeah, has influenced anybody? Because it seems like Thomson would have good reason to throw up her hands and say, "Was nobody listening, because the debate's still is: life begins at conception." That's all that people fixate upon. And she said, "Look, move past that. You're ignoring the rights of another person." That is the great thing about the violinist case, as she says at the beginning. "Women have been saying it's my body. That's, you're ignoring that point. You seem to think that the only thing that is worth discussing is whether or not the fetus is a person, but you're forgetting that the fetus is located inside somebody who has rights. And let's see how these rights affect whether or not abortion should be permissible." And as you say in this case, *McFall v. Shimp*, certainly what the judge said, and the decision seems to very much support that stance. I always thought the genius of Thomson's article is that it sets two conservative intuitions off against each other. There's the anti-abortion view, but then there's the libertarian view of "it's my body!", which we saw in full flow during the masking debate. Clearly that is a strain that has not gone away and if anything has gotten more intense and yet nobody picks up on this. Nobody even tries to engage with Thomson except maybe Christopher Kaczor or a few Catholic writers, but it hasn't entered the mainstream.

Yeah. I think that's largely true. So, [my book] *A Defense of Abortion* is written in a pretty standardly academic style. There's lots of footnotes, there's a fair amount of jargon. A lot of like, "here's ten responses to this objection, and here's seven objections to this response," and so forth, and covers a lot of ground besides just the stuff about Thomson. And in my aspiration with the *Beyond Roe* book was to write something more accessible to undergraduates, general readers, and so forth.

You certainly timed it well.

I guess that's one way of putting it. Anyhow. So I guess I was gonna say two things. One is, in teaching Thomson's argument over the years, and I'm sure lots of other people who've taught Thomson's argument have had the same experience, just the weirdness of the violinist case. I think kind of gets in the way of a lot of discussion, or maybe even giving it a fair hearing.

Let alone the People Seeds and the giant baby!

Yeah, no one knows who Henry Fonda is. I mean, if they don't know Columbo, they certainly aren't going to know Fonda.

***Twelve Angry Men*. I taught Thomson last semester to a bunch of high school kids and they were all wearing sweatshirts for their production of *Twelve Angry Jurors* (because we can't call it *Twelve Angry Men* anymore), and I was able to say, "Henry Fonda stars in the best movie version of that."**

For many years then you could have said, well, you know, he was Jane Fonda's father, and then students don't know Jane Fonda.

Philosophical Profiles

Even Bridget Fonda vanished.

It was a very brief period where you could say Bridget Fonda. Jane Fonda hasn't been for a while. In any event, I guess what I was going to say was, at some point when I was working on the *Defensive of Abortion* book, was when I came across the McFall/Shimp reference. I'm certainly not the first person to have made the connection to Thomson's argument. I guess, for people who are watching this: Mcfall gets sick and needs bone marrow, and the evidence seems to be that his cousin Shimp is the only person who could provide it. And then Shimp initially seems to agree and then he backs out. And then Mcfall is going to die. And so he tries to get an injunction from the courts to force Shimp to let Mcfall, have some of his bone marrow. And yeah, the judge says, in so many words, "this is ridiculous!"

I like the vampire analogy that he uses.

Yeah, yeah, there's, there's quite a bit of emotives wordings, including reference, I think, at least indirectly to the Nazis, and so forth. In any event, it's a clear, concrete case. It's a real-life case. And anyone who has the intuition that you'd have the right to unplug yourself from Thomson's violinist is going to have the intuition that Shimp has the right...you know, it was selfish of him, but morally speaking, is within his rights if he doesn't want to let McFall use his bone marrow. So I started using that example in teaching and I participated in a handful of public debates and I started using that example instead of the violinist example. And yeah, I guess my experience, at least in teaching, is *McFall v. Shimp*, helps to ground the argument in a case that doesn't get them worried about some of these questions about super science fiction, thought experiments and so forth. So the motivation behind the *Beyond Roe* book was to, in a sense, re-present Thomson's argument in terms of a legal right rather than the moral rights using *McFall v. Shimp*, but then also to address a bunch of questions that I don't think Thomson at least directly really addressed at all, which would be questions about sort of abortion restrictions. So even if you agree that you should have the right to unplug yourself or a right not to give someone bone marrow. Might there still be various sorts of restrictions that would be justified. And I became interested in thinking that through. But to get back to your original question, has it had any impact outside of the students in my class? I've gotten a handful of emails from people who've read the book and said that they found it pretty powerful. I've also gotten emails from people who have objections and they want to hear from me and I write back and tell them what I think. So probably a *tiny* little bit of impact.

It is strange that, that just seems to, I mean, in philosophical circles, Thomson's article is one of the big articles of the 20th century, but we haven't moved past the debate about whether or not the fetus is a person. Even if you think that she can't avoid the topic, at least you have to respond to her, but people just weren't.

I haven't looked at the recent Supreme Court ruling, but several days after it came out, a friend of mine who had sort of gone through it, sent me a list, or at least a partial list of the pro-choice philosophers who are referred to in a footnote in the opinion and made a point of saying: "No references to Boonin!" So now, if I was writing with the hope of influencing the Supreme Court - which I suppose, in my dreams, maybe I'm in part

“ I haven't looked at the recent Supreme Court ruling, but several days after it came out, a friend of mine who had sort of gone through it, sent me a list, or at least a partial list of the pro-choice philosophers who are referred to in a footnote in the opinion and made a point of saying: "No references to Boonin!" ”

doing... But no, that that certainly did not happen. I've written two books on abortion - maybe third time's the charm.

"Listen up, People," you should call the third one. Okay. You have so many books that I'm going to move on and ask you about *The Problem of Punishment*. Another view for your books you defend the claim that it is morally impermissible for the state to punish people for breaking the law. That sounds pretty radical. Say something about that.

Okay. Well let's see. I guess I'll say a couple of things. So part of the purpose of the book is to get the reader to think that there *is* a problem where they might not initially think there's a problem. And the way I think of the problem is in terms of a general practice. We assume that we have a bunch of just and reasonable laws [and] we have a bunch of just and reasonable procedures for deciding which people have broken the laws [and] which people haven't. We draw this line between the people who've broken the laws and the people who haven't, and we treat that line itself as morally relevant in a very particular way. Everybody on the offender side of the line, it becomes permissible, not just to do things that are very harmful to them, but to do things in order to harm them. That's the *point* of doing these things. And these are things that we don't ordinarily think it's okay for governments to do so. So when I say it's not morally permissible for the state to punish people for breaking the law, the claim is that you sort of you frame the problem in terms of "what's the justification for treating this particular line as a morally relevant line?" Then in the book I go through kind of consequentialist responses, various forms of retributivist responses, other kinds of responses (moral education theory, social communication theories, self-defense theories), and so forth. And the conclusion I reach is that in each of those cases; I mean typically there's a variety of problems. But one problem that I think comes up pretty much over and over again is they end up being overinclusive and typically also sort of underinclusive. So they end up justifying inflicting the suffering not just on the people on the offender's side of the line, but at least some people on the non-offender side of the line, and typically also they have difficulty explaining *why* it is permissible to punish everyone on the offender side of the line. So the claim that the practice is impermissible is the claim that the mere fact that someone has broken a just and reasonable law by itself doesn't provide a rationale for intentionally harming them. In a sense all I've done is kind of articulate what the thesis is, but I've [also] tried to build in what some of the motivation was. Once you come to see punishment as an institution that stands in need of justification - it seems to me - the kind of justification that it requires is precisely the kind that would show that there's a morally relevant difference between everyone on one side of that line and everyone on the other. And I think it turns out that there isn't.

I once had to take over teaching a class on punishment at the last minute, and one thing that I got out of that is that *Clockwork Orange* is a great work if you are a retributivist. Because the argument of *Clockwork Orange* is that "Ludovico's Technique" is not punishment. Ludovico's Technique is the method by which Alex, the antihero is made incapable of violence. Famously in the movie they prop his eyes open and make him watch movies of things like Nazi atrocities and inject him so he feels sick. And it succeeds, it makes him no longer a criminal. But the clear message

“We draw this line between the people who've broken the laws and the people who haven't, and we treat that line itself as morally relevant in a very particular way. Everybody on the offender side of the line, it becomes permissible, not just to do things that are very harmful to them, but to do things *in order to* harm them... And these are things that we don't ordinarily think it's okay for governments to do so. ”

of the book is that this is monstrous and that you cannot do this. And in fact *actually* punishing him in a way that is just standard punishment, you know, imprisoning him in a pretty brutal prison system, is better than that for Kantian reasons, I assume. Respect for his free will. So, what do you say to the retributivist who who says, "On the contrary: the alternatives to punishment, if we had this method of making criminals into good people, we still shouldn't do that because that in some sense does not respect their autonomy."

Boy there's a lot going on in that question. I guess I'll say two things. The first is there are different versions of retributivism. There's kind of the sort of basic sort of desert-based versions of retributivism - fairness-based versions of retributivism. And they might have different things to say about some of these cases. In my book I'm primarily, when I'm critiquing these views, trying to show that the views have implications that the proponents of the views themselves don't accept. So again, it's kind of my Socratic instinct. But as far as this question of what the alternative is and what would be better? So, as I recall, I'm not quite sure about this. The language of "restorative justice" was already sort of rampant at the time I was writing the book, but I don't think it made it into the book, as I put it in terms of "restitution," but in the last chapter I addressed the question: "If we can't have punishment, what are we allowed to do?" And I focus on victim restitution as an example of something that I think we can do. And in the case of violent criminals who do pose a danger to society, if there really isn't an autonomy respecting way of helping them to make themselves better, or like drug rehabilitation or something like this, then I say there may well be cases, maybe a number of cases, where we're entitled to protect ourselves from them by locking them up. So we do that with people who are found to be criminally insane - if they're found to be a danger to themselves or to society. But we don't take it that we're *punishing* them precisely because we don't take it that we're doing something *in order to harm them*. I taught this course on punishment for several years where I was working on the book right. So like the first many weeks or whatever I'd be like "I'm anti-punishment and now I'm going to shoot down this theory and shoot down that theory." Then, when we get to the end, often a lot of the students were really disappointed when I was like, "Oh yeah I still think it's okay to lock people up. But it's just not OK to *punish* them." But I guess I'd say a couple of things. One is that, in purely theoretical terms right, if it turns out that my form of the view of punishment abolitionism is still consistent with locking up people who are a danger to society then as long as you're doing it in a way that doesn't involve doing it in order to harm them... The way we do quarantine would be another sort of case: if somebody really is highly, highly contagious with a highly infectious and extremely dangerous disease, I think most people take it that it can be permissible to quarantine them...

[Makes "haven't you been paying attention to the anti-masking, anti-shutdown movements in this country?" face.]

Yeah. Okay. So I'll say a couple of things about that and I'll try and back up. So, first of all I did say *most* people. But, look. I could be wrong, but I suspect that with most people who would be opposed to mandatory quarantine in the COVID case, it would be because they're skeptical about whether COVID really is so dangerous or whether it really is so

“ In the case of violent criminals who do pose a danger to society, if there really isn't an autonomy-respecting way of helping them to make themselves better.. then I say there may well be cases, maybe a number of cases, where we're entitled to protect ourselves from them by locking them up ”

contagious. But I think if you stipulated a case where it just was very clear, everybody who gets within a hundred feet of this person just drops dead within hours or whatever, you'd be entitled to prevent that person from getting within one hundred feet of you. In any event, on the assumption that people accept that, then again, in the case of people who are found not guilty of a violent crime by proof of insanity or something, I think most people are okay with the idea that even though we don't hold them culpable and so we're not punishing them, it is permissible for us to confine them. So, part of what I wanted to say is that at a theoretical level if it turns out that my position ends up in a certain sense being not so different (at least with respect to violent criminals) from the practice that people were looking to defend, then that can be a good thing right? So we can say, "look you don't need punishment to get the thing that you really wanted, which was protection from violent criminals." But then, in all sorts of other cases, locking people up it will turn out isn't justified at all. But the other thing, I just feel like when we had time to go through this in class, I think it did resonate with most people, is that there are all sorts of differences (at least in principle) between the way we treat people when we put them in a prison and the way we treat people when we unfortunately have to confine them because of quarantine or because of mental illness or something like that. So there are all sorts of enjoyments that we deliberately deprive people of in prison because part of the point is we want them to be miserable there. And there would be no reason to do that if the only reason that we were confining them was to protect ourselves. So I do think there would actually be pretty substantial practical differences between what incarceration would look like under a non-punitive regime and what it looks like under a punitive regime. And so in that sense, I think that it's not such a trivial difference. But at the same time to get back to the original point, I think your question, I think it gives a fine answer to the retributivist. We don't have to lock people up in under nasty conditions and make them suffer in order to protect ourselves. Maybe that's something we'd have to do if we thought the state had the right to make sure everybody gets what they deserve. But I don't think on reflection, we think that.

I'm thinking of another Seventies movie - *One Flew Over the Cuckoo's Nest*. Because once you have a model where you say, "I'm not punishing you - I'm doing this for the good of society" then it's easy to view you as a patient, or even as a disease, and then you can be dehumanized in that respect. I mean, sure, prisoners are dehumanized but at the same time you have the right to attorney, you have a right to a parole hearing, these kinds of things. Whereas if you are being put there because you are a danger to society, then it's kind of an open-ended sentence. So, it could be that somebody has committed one minor felony, but they decided well we were lucky to catch them early. And in fact, this person is a psychopath and therefore nothing would prevent them from doing monstrous things. They could be imprisoned for a long time, maybe even an indefinite period. Whereas somebody who committed an atrocity but at some sense could show, "oh, that was a one-off thing"... Then it looks like the proportionality principle would be at risk. Retributivists tend to push that they're the only ones who can really account for proportionality of punishment.

Yeah. I'm not quite sure what to say because the line of argument that I think you're sketching is kind of consequentialist. You know, "if we give the state the authority to make decisions of this sort, maybe it's sufficiently likely that this will encourage them

“ There are all sorts of differences (at least in principle) between the way we treat people when we put them in a prison and the way we treat people when we unfortunately have to confine them because of quarantine or because of mental illness or something like that ”

Philosophical Profiles

to view these offenders as patients or in some sort of dehumanizing way and then as a result to mistreat them in one way or another." And I think that that might turn out to be true. But I think that's it's arguably more true that if you authorize the state to treat these people as people who deserve to suffer then that's what they get. So, the retributivism theorist... Okay. Pick a case like physical assault or sexual assault or blackmail or something. I don't know. At least the *desert-based* retributivist, they're just going to say, "well in principle, this person should be made to suffer in a manner or to a degree that's proportionate to the suffering that they inflicted on their victim or victims." So, they can say that. And then in some sense they can just make it true by definition that proportionality will be respected. But again if we're going to have an apples to apples comparison and if when we're looking at the restitution view we're kind of looking at it more like a *realpolitik* point of view, I think we would say the same thing. If you give the state and the judges the authority to make decisions about how to treat people based on their intuition about how much suffering this person deserves, I think you can get equally serious, if not if not more serious consequences. On the other hand if you make an apples to apples comparison by keeping them both at the *theory* level as opposed to the "real world" level, then I suppose one way of putting it is that on the non-punitive approach proportionality, is more like in self-defense cases, right? So it's not a matter of "how bad is this person, how bad is their character?" it's "how significant is the threat that we're protecting ourselves from and what can we do to mitigate it" and so forth. Right? In practical terms it's going to be difficult to distinguish between the person who's been rehabilitated and the person who hasn't. But again that's true on any kind of rationale for a criminal justice system. But in principle, if there are cases where we really feel that we've tried every humane and autonomy-respecting intervention and all the evidence points in the direction that this person is going to keep doing this - Hannibal Lecter or something like that - then yeah it's, it's regrettable, but... I want to say two things. The first is (and this is maybe the more important), the arguments that I make against punishment do not commit us to saying we're not allowed to keep this person confined, because the arguments against punishment are all arguments against the intentional infliction of harm. And that's not what's going on here. The second thing is I'm inclined to say and that would be okay to do under those circumstances. If there's someone else who thinks that wouldn't be okay either, then I think the upshot of my position would be, we shouldn't have punishment and we shouldn't *even* have preventive detention. That's not my view, but that would be a possible view.

When I was a kid there was there was a comic called *2000 A.D.* in England. The most famous character in this is Judge Dredd. And one of the things that happened in his strip that I thought was kind of intriguing was, they had a system where, instead of imprisoning people, they had machines that aged you. So, they put you in the chair and they pressed the button and they aged you by ten years. And that was your sentence. So, you served your sentence by just losing ten years. But that wouldn't really give you much time to think about your crimes.

So moral education theorists would definitely oppose that. But yeah I just um yeah that's interesting. You know if it didn't involve any suffering. I think the retributivist this might have to say that's an improvement over the current system.

Okay. Moving onto another book. How about *The Non-Identity Problem and the Ethics of Future People* (Oxford, 2014)?

This is exhausting. Yeah go ahead.

It's your own fault for writing so many books. For those unfamiliar with it, could you explain the non-identity problem? Why it is so important and what is so novel about your approach to it?

Yeah, the non-identity problem. I think the easiest way to approach it is through a particular example. Maybe I'll start with this. This is the example I mentioned in the preface to the books. I read about this one. During the time I was working on the book, New York City came out with a series of public service ads that were designed to try to discourage teenage pregnancy. I read about it because they ended up being controversial for reasons that are not relevant to the non-identity point. But anyhow, one of them, was a big sort of billboard in the subway, and there's this little boy, a little toddler. And as I recall, he's literally crying, or at least he's got these very sad eyes. And the wording says something to the effect of: "My chances of dropping out of high school are twice as great because you had me as a teenager." He's kind of "*j'accuse!*" to the teenage mom. The idea is supposed to be there's apparently a lot of statistical evidence that on average people who are born to teen mothers, teen parents, but mother is the one who gets the blame here, do less well in life by all sorts of relatively uncontroversial measures of human wellbeing. And so when you see this ad, you're supposed to think, "oh this poor little boy. He's been victimized by his mother. If only she had waited ten years. Finished high school and college. Gotten a good job before she had him, then his chances of success in life would be much greater." Right? And so that's why, so shame on you, you shouldn't become a teenage mom. Okay, so what I like about that example is that given the bigger time gap, is that it should become pretty clear as soon as you start thinking about that example. that if she waited ten years before conceiving, she wasn't going to have *that* kid. She would have just had *some other* kid. And that other kid would have had the great chance of graduating from high school. So not only did her having this kid as a teenager not reduce *his* chances of graduating from high school but really that the only way he could possibly graduate from high school would be if she conceived as a teenager. Ten years later, it would have been a different sperm, a different egg, and we think a different person would have been brought into existence. And that's where the non-identity part comes in. So the idea is many people, at least in the case of the teenage mother, have a pretty strong intuition that it's wrong for you to conceive a child at such a young age at least given these empirical assumptions about how much harder that's going to make it for your child. But then you realize that if she had waited ten years, it's not like that would have made things better for *this* kid. This kid just wouldn't have *existed*. And it still seems like he gets a life that's worth living. It's just not like as good as the life that this other kid would have had. Then you ought to be led to the view that she hasn't really harmed her child. [Whereas] the message that this billboard was clearly trying to convey is that she's harmed this child. That's why he's so sad. His chances of making it through high school had been greatly reduced. But it seems like, no, because of this non-identity feature of the case, she hasn't harmed him. And there's no particular reason to think she's harmed anyone else. So it kind of seems

“ It should become pretty clear as soon as you start thinking about that example. that if she waited ten years before conceiving, she wasn't going to have that kid. She would have just had some other kid... So not only did her having this kid as a teenager not reduce his chances of graduating from high school but really that the only way he could possibly graduate from high school would be if she conceived as a teenager. ”

Philosophical Profiles

like on reflection her action hasn't harmed anyone. But then you can join that claim with what seems to be a pretty common widespread element of, I guess I refer to it as sort of commonsense moral thinking, "no harm, no foul" view: if my act doesn't harm you it doesn't wrong you. If it doesn't harm *anyone*, then it's just not wrong, full stop. Then, it seems like her act isn't wrong. So that's one kind of example of the non-identity problem. You can spell it out a little bit more formally. You've got a set of assumptions that yield this conclusion. Her act didn't make the kid worse off, but if it had harmed the kid, the act would have made the kid worse off. So, the act didn't harm the kid. It doesn't harm anyone else [either]. It doesn't wrong anyone. If it doesn't wrong anyone, then it's not wrong. So the act's not wrong. So that's one sort of example. As far as the significance of the problem, once you start focusing on how fragile the circumstances of conception are you start seeing this problem in all sorts of contexts. So, a common argument against incest, at least heterosexual procreative incest, is that it increases the risks of various sorts of genetic abnormalities. But again, if a brother and sister conceive a child, and the child ends up with some kind of a genetic problem, it's not as if that kid would have been better if they had conceived with other people. *That* kid wouldn't have existed. Virtually every argument against developing reproductive human cloning technology also depends on the claim in one way or another that producing a child by cloning will harm the clone, either by causing physical problems, psychological problems, or social problems. All of those arguments go out the window if this non-identity problem remains unresolved. And then there are social-level versions of the non-identity problem. Climate change is a common example that people refer to. If a society at a large scale is choosing between two different policies and one would make things considerably better in the far future than the other, it's quite plausible that the choice between the two is going to have an impact on who meets who in the future, who pairs up with you in the future, who you conceive when, and so forth. Eventually you're just going to have two different populations of people in the future, and then you get this same sort of potential problem. One other area that I'll mention: arguments about reparations for historical injustice. Many of them can also potentially be undermined by this problem. Slave reparations is a classic case. At least on some arguments in defense of slavery reparations, the contemporary black American is said to have been harmed by the institution of slavery. But at least in the case when they're descendants of slaves, it's pretty plausible to think that if slavery hadn't existed they never would have been conceived by the pair of ancestors that at some point was brought together by this institution. Had the institution not existed those ancestors wouldn't have come together, and so forth.

That's the non-identity problem. That's the significance of it. Plus, I just think it's intrinsically interesting even if it didn't have all these potentially significant implications. As far as what's distinctive about my position: basically what I do in the book is, so in the first chapter, outline the problem. And I defend three requirements that I think a successful solution to the problem has to have. It amounts to saying, you have these five premises that each initially seem plausible [but] they seem to lead to an implausible conclusion. If you're going to solve the problem and avoid this conclusion you need to pick a premise and give a reason for rejecting it that's not *ad hoc* or some sort of independent reason for rejecting it. The reason for rejecting it has to be strong enough that I can't just tinker with the argument to get around the counterexample and then

“ Slave reparations is a classic case. At least on some arguments in defense of slavery reparations, the contemporary black American is said to have been harmed by the institution of slavery. But at least in the case when they're descendants of slaves, it's pretty plausible to think that if slavery hadn't existed they never would have been conceived by the pair of ancestors that at some point was brought together by this institution. ”

still get you to the conclusion of the non-identity argument. But it can't be so strong that it avoids that implication but then commits you to other implications that are even more problematic than that. Then basically what I do in the next five chapters is, one by one I go through the five premises that generate the problem. And in each case I consider solutions that people have offered where they've tried to show, "here's which premise you should reject and here's what's wrong with it." And I try to show in each case that at least one, usually more than one, of these requirements is not satisfied. And typically the problem is you can just revise one of the premises to get around the problem, but the revised premise still does the work of getting you to the seemingly implausible conclusion. Other times the problem is that it commits you to even more implausible implications. Then in the last chapter, this is what's distinctive about my position - not unique: a couple of other people have taken roughly the same view - essentially we just bite the bullet and accept the implication. What I try to do that I think *is* distinctive in the last chapter is, I try to give some cases to sort of massage your intuition so that it ends up not seeming so obvious to you that the act *was* wrong in that cases that generated the problem. That's definitely what's most distinctive about my view.

Have you read anything about ectogenesis? There was a bunch of articles of varying quality on the topic in the journal *Bioethics* recently, and what they mean by it is having the fetus gestate outside the body - basically artificial wombs. Like in *Brave New World* I guess. I don't know if it was Peter Singer first, but somebody suggested that this would be cutting the abortion Gordian Knot, because it would be basically getting what abortion gets for the mother without killing the fetus. And what they discovered is that, instead of this being a compromise both sides could agree on, both sides *hated* it.

Well, I think you may have just given me the idea for my third abortion book then.

Actually what the debate has turned into is whether or not the parent or both parents together have the right to kill the fetus at any point, given the possibility of ectogenesis. Does this mean that abortion should no longer be allowed because you don't have to *kill* the fetus to *get rid of* the fetus? But it occurs to me that this is assuming that the process of gestation cannot be paused. But by the time we've developed got this technology of ectogestation, presumably we will have also developed better freezing techniques. Just as we can now freeze zygotes, maybe then we'll be able to put embryos in suspended animation. And then, while they were frozen, court battles could and would be waged to decide their fate. But to bring it back to the topic in hand, maybe freezing would be a solution to the non-identity problem: you can conceive the child when you're a teenager but it isn't born then. So it *would* be the same kid that's born later!

I have to admit I was waiting to see how this is going to tie back. I think that might be the kind of solution that, as you said, makes nobody happy and that's usually a good reason for a philosopher to defend it. As you were speaking, though, I was thinking of another case that I didn't mention, where this kind of solution wouldn't work. So this is cases where a deaf couple is using IVF and then they wish to select in favor gestating the embryo that will become a deaf child. Many people have the intuition on the assumption

“What I try to do that I think is distinctive in the last chapter is, I try to give some cases to sort of massage your intuition so that it ends up not seeming so obvious to you that the act was wrong in that cases that generated the problem.”

Philosophical Profiles

(which obviously not everybody shares) that deafness impairs a person's quality of life in some nontrivial manner. Many people have the intuition that it's wrong for the deaf couple to do that. And obviously that's a case where it doesn't matter how long it's frozen before it's brought to term if it's such that genetically it's going to be deaf then it's going to be deaf. To be honest, that was one of the cases where biting the bullet didn't seem so difficult to me, but where most solutions denied any problem will try to in one way or another get you the results that they do actually harm the child, even though the child isn't worse off than it would otherwise have been, or that they have *wronged* the child even if they don't *harm* the child, or that the act is wrong in some nonperson affecting way. And then my response is look, it just turns out that it isn't wrong.

Hasn't the argument that sort of assumes the non-identity problem been used in court in wrongful life suits?

Yes, it has. I remember you asked earlier if I'm on the internet much and I think the answer is: somewhat more than I should. There was a case that came out in the last year or so that someone brought to my attention: I believe it was in England but I'm not sure. It had to do with someone who was conceived with some sort of genetic condition, I believe. I don't remember what, but in any event, what was striking was why the judge found in favor. I believe it was a woman who was suing her mother or the doctor for not having averted the birth or something. And what was striking was the judge's comments, as I recall, explicitly acknowledged this non-identity feature of the case. That had the woman not conceived, this person would not have existed at all. And the judge didn't seem to be saying, again I don't really remember the particular condition, but the judge didn't seem to be saying like "This condition is so horrible that they're stuck with a life that's like worse than no life at all." I mean there might be such conditions, but it was nothing like that. So yeah, it does come up in cases, and it seems to me that they just don't see the non-identity rather than that they have the wrong kind of distinctive solution to it.

Then of course there's David Benatar's position that it's better never to have been.

Well I you know I think that's a great book [i.e., *Better Never To Have Been: The Harm of Coming Into Existence*, Oxford, 2008]. I happened to have been one of the referees for the manuscript. And I was just completely unfamiliar with antinatalism and hadn't read any of Benatar's work before then. And I was pretty taken by it, I guess I would say. So, I think it's quite an intriguing argument. I do have a published response. It's a little bit detailed and I'm not sure I could sort of pull enough of it off the top of my head, but I have a great deal of respect for that book. And I'm not convinced, but I don't think it's making any kind of like simpleminded mistakes.

A few years ago I was teaching a seminar. And the students got to pick different things to read. And the kid that read that just had this visceral *hate* reaction to this argument. I was kind of surprised that it produced this response because this was kind of a slacker kid. This is a kid drifting through life and it was like, "No! How dare you say my life is not worth living!"

Yeah. I'm trying to think. I taught that a couple of times and I haven't really had that experience. I've certainly had the experience of students thinking that it's obviously wrong. And many of them think that it's obvious that his argument commits him to the conclusion that you should kill yourself. I think it's quite clear that it doesn't, although it takes a little bit of time to go through that. People found it sort of irritating. Students who already anticipate that they would like to have children can take it personally, in a way that not much else that we that we assign tends to do.

Well. Talking of arguments I think are an encouragement to kill yourself, I am sympathetic to the Epicurean view on death. Clearly, given your statement in your book *Dead Wrong: The Ethics of Posthumous Harm* (Oxford, 2019), you are not. Convince me.

Well, let's back up a minute. So what do you mean by the Epicurean view of death?

That my death is not a harm to me. And, I will assume that that implies that you cannot harm the dead.

Okay, then I guess that's the part that I would challenge. I'm pretty sure that the argument in the book is neutral with respect to the question of whether death harms you. This is not meant to be the whole argument but I suppose just as a quick analogy, I don't think *falling asleep* harms you. I don't think you are harmed while you're asleep. But I think things can happen while you're asleep that *are* a harm for you even if they never impinge on any of your conscious experiences. So I think if a peeping Tom were peeping on you while you were asleep, if you were naked or whatever, even if you never found out about it, I think that could make your life go worse, but that certainly doesn't commit you to saying that sleep is a harm. So, in roughly the same way then I guess I would say the same thing about death. So, the main argument that is supposed to motivate the view that posthumous harm is not possible, is that once you're dead nothing can impinge on any of your conscious experiences. But in the same way that while you're asleep or even while you're awake something might happen that you're just not aware of and it could still make your life feel worse. I think the same can be true of death.

But the big difference between sleep and death is that you don't exist when you're dead and you will never be conscious again. So, somebody sees you and you're asleep, presumably the arrow of causation works so that it could make your life worse after you wake up. Whereas this cannot happen when you're dead.

So as I said I didn't mean the analogy with sleep to be providing the whole argument. I think that was just meant to motivate the thought that the claim that something doesn't harm you doesn't commit you to the claim that you can't be harmed while it is happening. Sleep doesn't harm you but you can be harmed by something that happens while you sleep. Death doesn't harm you. But that was just meant to motivate the thought that that's not enough to get you the view you couldn't be harmed after death.

[Thomas] Nagel, for example, seems to be of the view that you can be harmed without it ever affecting you. You could be harmed by events outside of our light cone - or whatever it's called - that literally could not ever have a causal influence on you. If

“ The main argument that is supposed to motivate the view that posthumous harm is not possible, is that once you're dead nothing can impinge on any of your conscious experiences. But in the same way that while you're asleep or even while you're awake something might happen that you're just not aware of and it could still make your life feel worse. I think the same can be true of death. ”

your reputation was besmirched in a parallel universe, or whatever. That's not what you are saying. It sounds like you're saying that there has to be some kind of effect you for you to be harmed?

No I wasn't saying that. So maybe I should back up a little bit. All right. So I think I mentioned this earlier like I don't typically get interested in topics *per se*. I come across arguments. So I was actually in the process of developing a sexual ethics course, and I was trying to find different topics in sexual ethics that might fit into a course. And so I was reading just some sort of surveys. And there was one that was started to talk a bit about necrophilia and I thought, okay, I don't want to talk about that.

I was just wondering, "surely he can't be getting into necrophilia!"

But then, it was kind of like my experience with Thomson as a TA back in grad school. It referred to this article by George Pitcher ["The Misfortunes of the Dead," *American Philosophical Quarterly* 21 (1984)] and just gave like maybe a two paragraph summary of the argument. And I thought "that actually sounds kind of interesting..." And so then I check out the Pitcher article. I read it and reread it and mulled it over for a couple of weeks. And then, yeah - totally derailed. I had been planning to work on a project on sexual ethics. This was a number of years ago and I'm kind of getting back to it now. So, I was not previously interested in posthumous harm. It wasn't at all on my radar. There are in fact interesting social issues like posthumous organ donation. Posthumous embryo changes. But this argument was interesting, and the reason Pitcher's argument was interesting to me is precisely because he says something about the question that you just asked me that struck me as substantially different from what I had thought defenders of posthumous harm had to say. So you said the case where you're asleep, there's an existing subject who's being harmed. Let's say if somebody is peeping on them and we can assume for the sake of argument that's a harm to them. But once you're dead, there's no subject. So the traditional response those who defend posthumous harm, I came to see after I read into this more, is to try to come up with some remnant of you that still exists after you're dead. So they'll say something like "well your *biological* life might be over, but your *biographical* life continuous" or "your reputation still exists" or Feinberg has said your *interests* can still exist even after you don't exist. As if there could be these sort of disembodied interests. Okay, so what's distinctive, and what I found interesting about Pitcher's argument is, he doesn't claim that the subject of harm is a corpse. Or the pile of ashes. It's not the dead person, right? He claims that the subject of harm is the living breathing person who existed prior to their death. And so I found this puzzling but I sort of worked it through and eventually I was convinced. And so the basic idea, in the simultaneous harm case, like the peeping Tom case, is supposed to be, if we do have the intuition that the person being peeped on is being harmed, even if they're not aware of it, the best explanation is going to be something like that *right now* they have a *desire* that that not be happening. And the frustration of that desire makes things worse for them even if they're not aware of that the desire is being frustrated. So that's the first part of the argument that I tried to defend in the book. Anyhow, if you accept that view that the move from sort of simultaneous harm to backwards causation harm, if you want to look at it that way, is that most people also have at least some forward-looking desires about what will happen after their death. Many people have desires

“What I found interesting about Pitcher's argument is, he doesn't claim that the subject of harm is a corpse. Or the pile of ashes. It's not the dead person, right? He claims that the subject of harm is the living breathing person who existed prior to their death.”

about what will become of their remains. Their estate. If they were artists who have produced things, they have desires about what will become of them. The idea is that if at some point subsequent to a person's death one of those desires is frustrated, what that means is at the time the person was living, they were having a frustrated desire rather than a satisfied desire. Someone maybe thought that they were in the process of successfully completing a project where the project involved something happening after they die. But it turns out during their life they were unsuccessfully doing that. Because it turns out that after they die the project is thwarted. Something like that. That's the basic idea behind Pitcher's argument. And what I tried to do in the book is, somewhat similar to what I did with Thomson, I tried to expand on the argument, look at all the objections that have been raised in the literature, and try to respond to them.

Yeah I like your choice of Clyfford's Still as to begin the book: the the fact that he willed that all his art be collected one place [The Clyfford Still Museum in Denver]. And of course wills in general are obviously the phenomenon to press. Because if you honestly think that dead people can't be harmed then why should we follow wills?

I think that Clyfford Still stuff is very striking. I didn't know any of this until I happened to visit the museum. And it wasn't just that he sort of had these terms in his will but he was very explicit that this was his primary motivation for creating his works of art in the first place. He was creating them during his life with the idea that they would then be essentially gathered in one place posthumously. With respect to us, I don't want to deny that there are other kinds of stories we can tell. I suspect you could have some kind of rule consequentialist principle that could do a pretty plausible job of defending the claim that society as a whole will go better if we have a practice of honoring these requests. So I'm not sure that we can't come up with other routes to justify at least some of the practices that involve the dead. So in that sense, my motivation was like, "this is a puzzling claim, this is an interesting claim, but I'm starting to become convinced." And then after a while, I think I am convinced. And one way to find out if you're convinced about something, for me anyhow, is to try to write a book defending it. And if I do then I think it means I'm convinced.

That's an expensive way in terms of time and energy. But, hey if you've got the time.

Well, and I'm not convinced of that many things.

That's true. So each one deserves its treatment. Okay, you told me you have you have two more books already just about finished. One on sexual consent and one on AI Ethics. Give me the elevator pitch for each.

Okay. But one is closer to being finished than the other. The sexual consent book is almost finished. I'm taking the Fall to make some final revisions in response to some additional feedback. So, the elevator pitch for that book is, given that getting valid consent to sex is important, it's pretty widely recognized that there are three main kinds of things that can invalidate consent: coercion, incompetence, deception. You can have sort of clear cases in nonsexual contexts where pretty much everyone will agree, like, yeah, another person said "yes" to something but it's not valid consent. You can't act on it because it was clearly coercion, or they're clearly not competent to consent,

“ The idea is that if at some point subsequent to a person's death one of those desires is frustrated, what that means is at the time the person was living, they were having a frustrated desire rather than a satisfied desire. Someone maybe thought that they were in the process of successfully completing a project where the project involved something happening after they die. But it turns out during their life they were unsuccessfully doing that. ”

Philosophical Profiles

or they were tricked into signing the document or whatever. But in each of those three cases there can be versions where I think it's less clear what we should say. For example, clear paradigmatic cases of coercion involve a threat to inflict significant physical harm. And so, one kind of *non*-standard case can involve threats to inflict *non*-physical harm. So A says to B "if you don't have sex with me, I'll reveal an embarrassing secret." Maybe a pretty important secret, like that they're cheating on their spouse. So it's a threat of significant harm, but it's not *physical* harm. Does that count as consent-invalidating coercion? Or again paradigm cases involve threats of serious harm: many people think it has to be harm so serious that like no reasonable person would resist it. But you can have cases where the threat is pretty mild. I don't know: I'll pinch you. It'll hurt for a couple of seconds, or cause you some minor embarrassment, or something. And I don't find it obvious that that's not supposed to count as coercion. And then, similar things with incompetence. You know, you've got different kinds of cases involving incompetence. And again I think, for example, with intoxication, there'd be cases where it's clear that someone is too drunk to give consent, and other cases where I think it's pretty clear having a few sips of wine is not enough to invalidate their consent. And it's not as easy as you might think to figure out like what principle to appeal to to draw the line between them. So the book is basically a discussion of a series of cases where it's clear at a *general* level that the kind of thing in question invalidates consent: that coercion can, incompetence can, deception can. But it's not clear what to say about a *particular* range of cases. And then I tried to figure out what the most reasonable conclusions are.

The AI ethics book is less far along but I did just finish teaching a graduate course on the main topics. So I think at least I have my elevator pitch ready. So that is focusing on four forms of machine learning, where three cases are already widely used by governments and the fourth case is under development for government use. And for each case there's really one distinctive moral problem that they raise. I guess I can mention them very briefly. So there's kind of the big data worry that massive scales of AI involve a violation of the right to privacy. So I have a chapter on CCTV-enabled mass public surveillance. And there the focus is on whether mass public surveillance violates the right to privacy. You have these predictive policing algorithms which are used to help police departments make deployment decisions: these are things that are already very widely used. They take historical crime data, they spew out predictions about where crimes are most likely to occur, what times of day or night they're most likely to occur. The police then make their deployment decisions in accordance with that. The concern there is the main concern that people have raised that I focus on, is various ways in which the algorithms become racially biased. Either they inherit the racial bias of the historical crime data or exacerbate other racial bias in various ways. The third chapter is on risk assessment tools. So these are all sort of data-driven algorithms. Actually, kind of related to some of the *Clockwork Orange* stuff we were talking about earlier, I suppose, but making predictions about who's a bigger risk to reoffend, flee from their trial, this sort of thing. And courts are increasingly using these to make decisions about, not just bail and parole, but even sentencing. And the main problem that I focus on there is opacity. So you've got these algorithms that are so complex... There was this famous case in the literature: Eric Loomis in Wisconsin. He's sentenced for a relatively minor crime, but he gets the maximum sentence, and the judge says a part of the reason for the maximum sentence is because this algorithm, "COMPAS," said that he was at a high

“ So the book is basically a discussion of a series of cases where it's clear at a *general* level that the kind of thing in question invalidates consent: that coercion can, incompetence can, deception can. But it's not clear what to say about a *particular* range of cases. ”

risk to reoffend. And Loomis's lawyer says, "well, we'd like to see why the algorithm thinks this." And in this case the problem is that the algorithm is proprietary. It's a secret formula for private companies and they shouldn't have the right to look at it.

But then the court shouldn't have the right to use it.

Well, that's that's what Loomis said, but he lost in court. So, there the claim is that the opacity of these algorithms violates what said to be a *right to an explanation*. That's the kind of core term that you come across. So, mass public surveillance, predictive policing, risk assessment tools. These are all widely in use to try to address crime in one way or another. And each one has raised a pretty substantial moral objection. So I will say this tentatively because I haven't finished the book yet, but the way the book is written right now, I end up rejecting all three of those objections, including the right to an explanation objection. You're just asking for the elevator pitch now, so I probably can't give you more than that, but I think it turns out that none of these objections are successful. And then the final chapter is a bit different because it's on autonomous weapons systems which are under development. These sort of killer robots don't really exist right now and some people think that they won't, but to the extent that we could develop weapons systems that are autonomous in the sense that they make their own decisions about which targets to engage and how to engage them - and when I say "make their own decisions," decisions that are not just preprogrammed into them but that are based on lessons that they've come to on their own as a result of acquiring information that they acquired on their own, well after their programmers activated them and so forth. There's this very interesting argument to the effect of that, if we use one of these things and then it kind of goes haywire and decides to like massacre a bunch of innocent civilians, nobody can be held responsible for the deaths of these innocence civilians. Robert Sparrow has maybe the best-known article in the ethics literature called "Killer Robots." And in that chapter I argue against *that* objection too.

You're fine with the coming dystopia. You welcome our new robot overlords.

Well yes. Actually the working title is: *Two Cheers for Our New Robot Overlords: AI Ethics and The State* or something like this. But what I want to do in the book is just set aside all the empirical objections, just assume these algorithms would reduce crime a bit or assume that these autonomous weapons systems would at least somewhat reduce civilian casualties and so forth, and just look at whether these particular moral objections are satisfactory objections. And, again, it's a work in progress, so maybe don't quote me quite yet, but I'm on track to be defending the view that in all four cases, if the empirical stuff works then it's okay for the state to do all this stuff. I find myself as surprised as you look. I definitely had that experience teaching this material this last semester. I will say, I think of myself as someone who has very strong civil libertarian instincts, as someone who ought to be pretty predisposed to accept all these objections. And yet I find when I think about them, I'm not.

So somebody needs to come up with one that convinces you.

Yeah. I mean that's that's one way to look at it. I mean yeah. Or it's possible that there's a lot of sort of hysteria over this. It's weird: I have to say as I get older now I feel like this

“ The working title is: *Two Cheers for Our New Robot Overlords: AI Ethics and The State...* But what I want to do in the book is just set aside all the empirical objections, just assume these algorithms would reduce crime a bit or assumed that these autonomous weapons systems would at least somewhat reduce civilian casualties and so forth, and just look at whether these particular moral objections are satisfactory objections. And... I'm on track to be defending the view that in all four cases... it's okay for the state to do all this stuff. ”

Philosophical Profiles

is the one area where I would have expected younger people - they've grown up they're digitally literate and all this sort of thing - to be more sympathetic to [these uses of AI]. We're already using computers for all *these* things, what's the big deal about using it for [these other things]? And then it'd be the old fogies being like, you know, "this is totally unacceptable!" But, at least so far, my students are generally much more resistant to all this than I am.

But it's old people who are using Facebook, who have Alexas and other smart speakers, because it's easy, whereas the young people say, "Hell no - Zuckerberg's not getting *my* information!"

Yeah it's interesting. I will mention that in the case of the mass public surveillance one of the reasons that I was interested in this is because I can remember having this discussion even when I was a college student. So, I was a young person visiting a friend in New York City. And he's also very civil libertarian. And we somehow got into some discussion about whether it would be okay if the cops could inspect your backpack before you walked into Central Park or something. And he thought it was just obvious that that would be totally unacceptable. You know - "it's public property, anybody can go there," and whatever. I feel like I can remember all these years back, and this is partly what motivated me to take on this topic now, is thinking that I just didn't find that so obvious. You know, "I have the right to walk into this park without proving that there is not a bomb in my bag." I don't know. So some of this stuff, I think even young people - at least for me even as a young person - I think I would've been skeptical. But who knows?

Regarding the consent thing, I seem to recall that Hobbes - to tie it back to your dissertation - Hobbes had a very loose notion of consent. Doesn't he say that if you're held at gunpoint and you consent to give someone your purse, then that consent is binding?

Is it okay if I admit that I don't remember. It sounds like the kind of thing he would've said.

I wrote my dissertation on the social contract, and I remember thinking that if you look at a lot of fairy stories, like the story of Rumpelstiltskin, the idea that there's binding consent is ridiculous. I argued that a Lockean social contract just wouldn't work because, the kind of circumstances you're in in the state of nature, no consent would ever be binding.

I'm not sure if I can pull this off off the top of my head but one of the cases I am talking about in the sexual consent book involves so called "third-party" coercion. There's an interesting little literature - it was kind of stagnant for a while, but the last year or two there's been a little bit of a revival. But these are cases where A threatens to harm B unless B gives consent to C to do something. So like an example to tie it back to the abortion case, it could be the boyfriend threatens the girlfriend and says, "I will beat you up unless you go get an abortion." But, of course, to get an abortion she has to give consent to the abortion provider to perform the abortion. So A threatens B into giving consent to C. Picture a case where the abortion provider knows that the only reason the woman is consenting is because her boyfriend will beat her up if she doesn't consent. But suppose she refuses to tell the abortion provider who the boyfriend is.

There's no way for the abortion provider to prevent this. The question in the literature on third-party coercion is whether, even though two-party coercion uncontroversially invalidates consent, three-party coercion might nonetheless *not* invalidate consent. In that section I do end up defending the view that it *doesn't* invalidate consent. Again, a somewhat surprising view that I initially became convinced of by reading a, as far as I can tell, almost completely uncited four page medical ethics medical journal article from the 1980s. It seems to me that that might be relevant to the issue you're raising. So, if one holds the view that I'm defending, that third-party coercion does not invalidate consent, I think it could turn out that, let's say you're the sovereign, and the only reason I give my consent to obey you is because someone else is threatening to kill me and this is the only way that I can avoid that. I think then I would be committed to the view that that could be valid consent. I'd have to think a bit more about that.

Well, you've got more books in you clearly.

At least those two.

Well, thank you very much David. That was very fun. And you clearly are not afraid of being a contrarian. You go where the argument leads.

Well thank you very much. This has been a pleasure. And yeah the contrarian thing I guess, I will say, When I wrote *A Defense of Abortion*, I felt good about that book. I thought I had done a nice job. I mean obviously, it's a contentious issue. Obviously many many people disagree. But you know in the world that I inhabit, the world of academia, the vast, vast majority of people are pro-choice and many of them kind of almost unreflexively... And so I think to some extent it was a little bit less rewarding because the people in my world... it wasn't so clear that many people really felt like we needed a whole book defending this, or whatever. So yeah, I don't feel that I'm deliberately contrarian, but there has been something rewarding about finding arguments where "this seems right to me," and then seeing it's not just that there are other people out there that don't agree, but plenty of people in *academic philosophy* don't agree.

It's a good way to get cited a lot, that's for sure. Because everybody wants to write a reply to you because, "oh I hate *that* argument."

Yeah. I mean I guess that's working well for David Benatar. I don't see myself as in David's league as a contrarian but I can aspire to be.

“ I don't feel that I'm deliberately contrarian, but there has been something rewarding about finding arguments where "this seems right to me," and then seeing it's not just that there are other people out there that don't agree, but plenty of people in *academic philosophy* don't agree. ”