# Philosophical Profiles

# Pamela Hieronymi
## Professor of Philosophy at UCLA

**IN BRIEF**

Pamela Hieronymi is Professor of Philosophy at the University of California, Los Angeles. She has written about forgiveness, blame, belief, reasons, agency, responsibility, free will and ethics. She has one book forthcoming (*Freedom, Resentment, and the Metaphysics of Morals*, Princeton) on perhaps the most influential article of the second half of the twentieth century, and another one under revision (*Minds That Matter*) that brings together most of the threads in her work so far. She believes humans almost unavoidably do wrong, but that that doesn't excuse them from blame for so doing. However, the good news is that blame on her conception is not a form of punishment—it is not to be confused with "guilt tripping." She also defends a version of the moral contractualism of T.M. Scanlon (whose *What We Owe to Each Other* (Harvard, 1998) she is acknowledged within, and which she is responsible for featuring prominently in the sitcom *The Good Place*) whereby morality is a matter of rules that reasonable people would agree to be constrained by if others did likewise. This is a "minimal" view of morality which has been criticized as giving too weak a reason against torturing babies for fun, a charge Hieronymi rejects. When she is not taking on consensus on the big questions in normative philosophy, or giving TV producers crash courses in philosophy, she spends time with her cat, despite her view that dogs are more likely capable of morality.

# Pamela Hieronymi

## a philosophical profile

**SC: According to a former interview with Richard Marshall of 3:AM magazine, you first became interested in philosophy when your dad gave you Descartes' Meditations in the fifth or sixth grade. What was it that gripped you about it?**

PH: The idea that we could be living in an illusion.

**So, that was sort of a shock to you, like a "red pill" moment?**

Well, I mean it's the combination of the kind of metaphysical shock but also the realization that there was not an intellectual way out of it; that you can't think your way through this, or that all options seemed closed off in terms of establishing what seemed like the most obvious things.

**Right. I remember having that experience with Hume when we were made to read him for the first time. Now, in the same interview, it said that you took two years working in a non-profit in DC after your undergraduate career. Would you care to say what it was and why was it that you did that instead of grad school? And why was it that you stopped doing that and then went to grad school?**

So, the non-profit was called the Ethics Resource Center. I don't think it still exists, but it was an interesting combination. Its bread and butter was corporate consulting for an ethical environment. It had almost no philosophical interests, but I think it did a lot of good work. It was more or less straight up corporate consulting but looking at what they called values and vulnerabilities. You have a manufacturing company; productivity and safety are always going to be conflicting values. So, they would identify those points of conflict and advise management on how to avoid ethical [problems] - ending up on the front page, or what have you. So, that group made some money. I worked with a group that spent some money making videos, K- 12 classroom videos, about being a good kid. So, I did that. I wasn't interested in going to graduate school.

**Why not?**

I'm not sure actually. I was definitely encouraged to go to graduate school.

**Was it that you felt like you wanted to make a difference?**

I definitely felt that I wanted to make a difference. It wasn't though that I was going to do that in any large way doing this non-profit thing that I did... I was placed there by a fellowship. There was an organization of alumni that would network graduates with non-profits and so I found my way there through that network. But, yeah, I could ask my teachers what I said to them when they asked me to go to graduate school. But, then as I was doing the work for the non-profit, I was put in charge of producing the print

materials and doing some research on childhood development and the development of moral personality. And I realized that what I enjoyed the most, was that work. Going to the library, writing things. And that's what made me realize, oh, actually probably I should go back to graduate school.

**So, was it that experience that directed you in grad school? I mean did you go to grad school and say, "Ok, this is what I want to do. I want to talk about moral psychology," because that is an area you've ended up in?**

No. I mean my interests in the very things that I am still writing about today pre-dates that. So, as an undergraduate, I needed to write a senior thesis and there's more or less a straight line from the things I was asking then to the things I'm still doing however many decades later. So, I think it was just good luck that the research they asked me to do at this non-profit had some relation to those issues.

**And, when you got to grad school, did you quickly fall in with a particular professor or did you have a particular class that really sparked your interest and sort of said, "Ok that settled what my dissertation is going to be on?" Or, did you sort of bounce around a little bit?**

I bounced around a little bit. In my second year, I realized, there were terrific classes being offered in epistemology and metaphysics, but in my second year, I realized that I really should stick with ethics. That's where my interests were and that that's where I had motivation to work hard. And, so I did. And then one of my undergraduate teachers, Richard Moran, at that point moved to Harvard so then I had somebody who I was familiar with there. But, I kicked around a little bit, but fairly quickly settled into taking every course that Chris Korsgaard offered. I just missed getting the class with John Rawls which was a disappointment.

**One of your earliest articles is on forgiveness. Now, was this sort of a one-off thing that you thought maybe this is a little problem that I can isolate and settle absent of other concerns or did you already see it as sort of part of a grand theory or a web of projects.**

It turned out to be a smaller problem. So, it turned out to be a good first paper for that reason, but that's not why I went there. I was at that point thinking that I was going to write a dissertation on things you don't do for reasons. Things for which there aren't reasons - you know, faith, hope and love, something like that. I had thought that forgiveness might be one of those and then as I worked on it, I realized it's not one of those, but I had created this paper that I was encouraged to send off to a journal. And then I'm often asked if I would write more and don't feel I have more to say on that topic, but I am surprised when I go back to it again, how much of the subsequent work I've been doing about attitudes and the wrong kind of reason for attitudes is actually in that paper just without all of the machinery to highlight it.

**So, at the moment you're working on a book. How's that going by the way?**

> **"** I was at that point thinking that I was going to write a dissertation on things for which there aren't reasons - you know, faith, hope and love, something like that. I had thought that forgiveness might be one of those and then as I worked on it, I realized it's not one of those **"**

I'm working on two, actually. So, I have a little book on [P.F.] Strawson's "Freedom and Resentment," a close read of that, that is nearly ready to go to press.

**A much neglected paper.**

Yeah. I'm excited about this manuscript. I'm convinced that the main argument of the paper has not been understood by the ethicists at least, in kind of a big way. So, I spend nearly a hundred pages, well I spend nearly eighty pages, walking through his few pages very, very slowly and drawing on other parts of his corpus to try to say exactly what's on his mind, which [is that] I think he has an underlying metaphysics of morals that to most ethicists will look like a non-starter. Then, at the end, I try to defend that to make it seem like it's not a non-starter, which I think it isn't, to show that that's a contender. So, I'm excited about that book: it should be out in 2020. And then I have what I think of as my big book, Minds That Matter, on the free will problem - moral responsibility and free will. That is going very well: I have a full draft of it. It kind of falls into two halves, and especially the back half, a lot of it's still pretty new so I think I'm going to sit on that for a little bit, present pieces of it around, do the usual thing of allowing it to mature in response to feedback. So I think that's still a couple years off.

**From looking at the introduction, which you've made available, it does seem like a lot of, at least the first half, is drawing on papers that you've already written, other articles you've been doing, so it seems like it's now bringing them all together. Because one of the problems that no doubt you face, is that every time you write an article, you have to say, "here's all the other stuff I bring to bear to this article that I can't really talk about now, but go read these other articles," and you have to try to make them self-contained, whereas the beauty of a book is you can just say, "well, as we've just been talking about in the previous chapter..."**

Exactly.

**About the Strawson book: I read Strawson as an undergraduate, but it was all his stuff on philosophy of language. Then he writes this paper, and he admits "this is the only paper I'm going to write on ethics," and he writes it very quickly. [It was initially well-received] and then it sort of fades into the background and then all of sudden, wham, it's like the most important article of the 20th Century according to this huge group of thinkers who work on this topic. Do you have a theory about why that is? I mean, he does say at one point, what is it, "the moral sentiments are an unjustly neglected area" or something like that. So, you think that people have suddenly become interested in that because of him or just become interested in that and then discovered him or why is it suddenly the bee's knees?**

I think it's complicated. So, the main thing that people mean when they talk about Strawson or the Strawson article or Strawsonians has to do with the reactive attitudes and the contrast between the reactive attitudes and a more objective attitude and that itself is a kind of fascinating and fecund distinction or idea.

> **❝** I'm convinced that the main argument of Strawson's "Freedom and Resentment" has not been understood by the ethicists at least, in kind of a big way. **❞**

**In the article he just says, "there are these things, the reactive attitudes that I'm going to use to say why compatibilism seems to work [or rather] why we don't have to worry about the panicky metaphysics of the incompatibilists." So, it's almost like, "oh you know here's this thing that people don't really talk about that helps us talk about the free will problem," which was a big deal back then as now. Whereas people have since said, "oh my God, it's like you discovered gold and then you ignored it."**

I think that it's complicated. One reason has to do with an interest coming from not just Scanlon and his contractualism, but then a similar thing has been picked up by [Stephen] Darwall with his second personal standpoint. [R. Jay] Wallace is now talking about the moral nexus, so these are all accounts of morality that are focusing on interpersonal relations. The Strawson article is pointed to as a kind of beginning or as a touchstone for that kind of approach. And, that's not incorrect. I'd like to be able to say, "but it has nothing to do with the argument he gives;" that's not true. It does, but the argument he gives, as I understand it, relies on the claim that we adopt the objective attitude in outlier cases. And, it's impossible for everyone to be an outlier. And that's why determinism isn't relevant [to free will].

**Just in case anyone is unfamiliar with this article, the "objective attitude" is when you treat someone as if from a God's eye point of view. Instead of interacting with them, it's as if you step back and say, "oh yes, this person is clearly crazy, so we don't have to interact with them as a fellow person." We can just treat them as we would a mad dog or something like that and to take this attitude is something we can't even bring ourselves to do in most cases and is a terribly objectifying attitude to take toward someone. Right?**

I think that's the most extreme version of it, but he allows that it's also what we do, it's to step away from the engaged attitudes of resentment or indignation or gratitude and to treat someone like an issue or a problem.

**So you can still talk to them and interact with them. It's just an "I'm not gonna let their bullshit get to me" kind of thing.**

Right, and he thinks we do that sometimes just because somebody's having a bad day so they're being mean today. But because they're having a bad day, I'm not going to react to it. They're under extreme stress, he says. Or children, he thinks we have a kind of partially objective way or relating to them because of the immaturity. And then, in the most extreme cases, yes, it would be the kind of thing you described.

**Obviously Strawson is going to come up again, but I'm going to read the first sentence of your introduction to _Minds That Matter_. Incidentally, that title makes it sound a little bit like it's a philosophy of mind work, which of course it isn't primarily. Are you going to have a subtitle?**

I probably should have a subtitle. It [does have] some pretty strong implications in philosophy of mind.

> 66 The argument [Strawson] gives, as I understand it, relies on the claim that we adopt the objective attitude in outlier cases. And, it's impossible for everyone to be an outlier. And that's why determinism isn't relevant [to free will] 99

**"The immodest ambition of this book is to unwind the traditional problem of free will and moral responsibility." Now again, for the uninitiated, could you lay out the "traditional problem."**

So, the first chapter works on that. The first two chapters, actually, are sketching the various problems, plural, that go under that head, but most basically, the problem is, if our science can explain everything that happens, at least in the macrophysical world, how is it that we're free or responsible? If what I do can be explained by facts about either nature or nurture and my environment tracing back ultimately to before I was born, then it seems that what I do isn't up to me and so it can seem that I'm not responsible. It can in fact seem that nothing really matters, so, in a nutshell, that's the problem.

**Now, Strawson, of course, was tackling this problem and also another famous mid-century response to this is from Harry Frankfurt. You don't think that Frankfurt's response works. Could you say a little bit about your take on his response because it has been so influential and caused such debate and what you think is lacking?**

So Frankfurt's first paper on this topic, "Ultimate Possibilities and Moral Responsibility," presents a counterexample, and I just think the counterexample doesn't work. I suspect he has to think the counterexample doesn't work because it prompts a massive literature to which he contributes not one letter, and instead I think that example revealed an intuition that was important that he pursued. The counterexample is a counterexample to what gets called the Principle of Alternate Possibilities, which is that if I'm morally responsible then I could have done otherwise. The intuition he has is well, suppose I wouldn't have done otherwise even if I could have. He gives an example [the famous "Mr. Black" example]. If I wouldn't have done otherwise even if I could have, why should the fact that I couldn't have, if you take away that possibility, why should that matter for moral responsibility?

**Which is a point that Locke had already made.**

Yes, it is Locke's point. So, you could say the insight is that what matters for moral responsibility is in some way whether what I did was what I really wanted to do, but then of course all the work is in the "really." And his later work is trying to work out what it would be to really want to do something in a way that's sufficient for moral responsibility. At that point, he starts talking about hierarchies of desires. So, what it is to really want to do something, is to want to want it and to want to, want to, want it and to have no conflict [among one's hierarchies of desires]. So, that's interesting and fascinating and worth a long hard think about. In the end, I don't think it's what's required for responsibility because you can be responsible even when you're conflicted. So, his view and Gary Watson's similar view have a hard time with ordinary weakness of will, giving in to temptation. But I do think, that looking at the sophisticated way in which we relate to our own mind, is the right direction to go to. So, what happens in the first part of the book, a lot of which has not been published before, surprisingly to me - I thought I was going to be doing a lot more repackaging, but I ended up having to do a lot of new work - one thing that happens is to go through some of the attempts

> 66 In the end, I don't think [Frankfurt's view is] what's required for responsibility because you can be responsible even when you're conflicted. So, his view and Gary Watson's similar view have a hard time with ordinary weakness of will, giving in to temptation 99

by Frankfurt, by Watson, by Charles Taylor, and then to suggest a simpler way in which we are in control of our own wills. So, to suggest that just by being able to answer questions, that puts us in a kind of control over ourselves.

**So, what you're getting to is your notion of evaluative control. Now, to introduce that, you contrast that with the ordinary notion of control. So, could you make that contrast?**

Sure, so the ordinary notion of control is the kind of control we enjoy over our own actions and over ordinary objects. So, if I am controlling my piece of chalk or my coffee cup, I have in mind a way I want that to be and I control it to the extent that I can bring it to be as I would have it to be. So, that ordinary notion has what I call a two-part structure. There's the thing controlled, the coffee cup or my future behavior, and there's the activities by which I bring that to be to match my idea of how I would have it to be. I impose my will on it. If the only notion of control you have is that one that has this two-part structure of controlling activity and object of control, you'll then have a worry, you might wonder how it is that the controlling activity is in your control. We're sophisticated, reflective creatures, so one thing we can do is think about our own activity and make it into a higher order object of control. I'm now returning a little bit to what's the weakness in Frankfurt. So, I think that kind of reflective idea, that we can stand back, think about ourselves, make decisions about how we can be, that looks satisfying, or seems tempting because it's recreating some of that ordinary control.

**His notion of a "wanton" seems to be an intuitive one, where he says we know that there are these people who do not reflect. Essentially, a wanton is someone who is just driven by their first order desires and do not have desires about their first order desires. So, that is a powerfully intuition by which he drives his account, but you'd agree with that right? That a wanton is somehow lacking in control?**

A wanton does not exercise ordinary control or what I sometimes call managerial control with respect to their own desires. The story I'm going to end up telling about evaluative control, the wanton may well exercise evaluative control.

**As your previous interviewer said, a lot of the things that you like to work on are important basic ideas, like blame and forgiveness, things that are a part of the common-sense folk language. But, at the same time, you give stipulative definitions of them that seem to make them more narrow. For example, control. It seems to me that it's perfectly natural to say that "the robot is controlling it" or "the microchip is controlling it," but you would say "that's not the kind of control I'm interested in. That lacks something. That's not real control." And, your notion of control seems to be very intellectual.**

I don't want to rule out ordinary control. But my quarry is the free will problem. So, one way to think about that problem is to say, well, "the thermostat controls the temperature in the room." And if we were content with that as a form of control, there'd be no question about whether we control things. But if a compatibilist were to try to say "don't worry about determinism, don't worry about these kinds of things because,

look, we control things," then someone like Kant would say, "no that's the freedom of a turnspit." Someone could easily say, "that's the control of a thermostat." That's not real control. So, I take my task to be in part to say what it is about humans that we control our behavior in some way that is more than the control of a thermostat. So, I don't want to deny that there's a perfectly good use of the idea of control such that the thermostat controls the temperature of the room.

**But you also want to say that that's not enough and you want to say why it's not enough. It's not enough because it lacks what we care about.**

Right. And I want to do it in two stages. So, I want to say that it lacks self-determination. The thermostat is in control of the temperature, but how it controls the temperature is not up to the thermostat. Whereas I want to say how we control our environment, is up to us, in a way. That's the idea of evaluative control. We bring things to be as we would have them to be by answering questions for ourselves. But that idea of being self-determining or answering questions for yourself is going to be broader. There are going to be creatures that are capable of that that are not yet morally responsible. So, it's going to be a two-stage process because in addition to being self-determining, you have to be capable of showing respect or disrespect. You have to be capable of standing in certain relations - so we're coming back to the relational aspect - standing in certain relations with others such that each of you can hold the other to expectations that amount to standing in relations of possible respect or disrespect.

**So, the incompatibilist says, "Ok, you compatibilists, you've got control but you don't have self-control." You agree that by our ordinary notion of control, we do not have self-control. That would require an infinite regress and that's the problem Frankfurt seems to face. You don't just need second order, you need third order and so on. So, your solution is to say we need to expand our notion of control to include what you call evaluative control and according to that, we do have the control over our self that the incompatibilist says we can only get by sort of severing ourselves from the causal web of the universe or by entering the noumenal realm or whatever. You say, no, we don't need to get involved in that panicky metaphysics, we just need to broaden our concept of control. Or maybe you think that it's not broadening it, you're just bringing out what we already thought and that's evaluative control. Could you explain that?**

Yes. It won't get what the incompatibilist wants. The regress is real. But, maybe the most intuitive way into this is to think about the active/passive distinction. There are things that we are clearly active with respect to - raising your right hand, going for a jog, running for office. Those are voluntary actions. On the other extreme, things that just happen to you, having a headache, aging, being subject to gravity, those are purely passive. But I think that a very large and important class of things about us don't fit well in either category, and the easiest example is belief. So, I can't believe at will, I can't believe voluntarily. It's not up to me what I believe. I just have to believe what's true, what's obvious. I just open my eyes and I end up with beliefs, as though they were passive. On the other hand, believing that the butler did it is not like having a headache.

> **"** I take my task to be in part to say what it is about humans that we control our behavior in some way that is more than the control of a thermostat. And I want to do it in two stages. So, I want to say that it lacks self-determination [and] you have to be capable of showing respect or disrespect **"**

It's not something that just befalls you. In fact, even perceptual beliefs, as opposed to perceptions, I think, are up to you in a way. If you think that what you're seeing is an illusion, you're not going to believe your eyes. That's a thing we can do. And, the mark of this is that it's apt for people to ask you why you believe. When I'm asked "why do you believe the butler did it?" my reply - and this is a point made by Richard Moran - is going to refer to things separate from me in space and time. I'm going to talk about the butler and his access to the home and his motives for revenge, and I'm going to give a case that supports the conclusion that the butler did it. Which is very different than if you ask me why I have a headache. There, I'm going to tell you about things to do with my brain that explain this pain. So, belief, it seems to me, is not voluntary: you can't just believe anything you want to. On the other hand, you're not just passive with respect to believing. It's up to you in some way. So, I think that anything about us that reveals our take on what's good or true or important or worthwhile or lovely or irritating, will have just the same in-between, or alternative, relation to activity and passivity.

**Ok. I'm going to push back, as I'm sure you get an awful lot of pushback on this point. For one thing, you sound like a philosopher, which is only right and just, but if you ask some people why they believe what they believe, that really pisses them off. They are going to get defensive and they are going to get annoyed with you and they're going to shut down because people preface things by saying, "as a Christian, I believe such and such." What they mean is, "I have adopted this identity and it comes with a set of beliefs and my spade is turned at that point. I do not have to justify this to you." Now, philosophers are of course appalled by this. But, on the other hand, if you ask philosophers why you believe this, you would say, "Well, I am compelled to by the facts of the matter." In both cases you're kind of compelled. Now, you say, believing is something we do rather than as you say a headache, but that seems, again, if I'm pushing back, I could argue that's kind of an artifact of language and we don't really have the kind of control over beliefs that is what I want. Suppose, I'm greedy. I don't want to be an incompatibilist because that never made sense to me, but I want more than that because I want to say the only way in which I can refuse to believe what my eyes are telling me is if I have other reasons for doing so, like having been given Descartes' Meditations at a formative age, something like that. So, what do you say in response to that? That your evaluative control, that's not really control, that's broadening the term too far.**

I want to accept the idea that if you think the word control should not go to anywhere that you don't have a kind of discretion or voluntariness, then I agree, that we won't want to call it control. So, I have to do work to explain why I think you should let me call it control. But, I don't want to get hung up on the verbal dispute. What's important is that it's up to you in a way that it reflects your take on things. What I need is a division between facts about you like your bone structure or your birth place that you're stuck with, and facts about you like your beliefs, your resentments, your fears, your cares, your concerns, that are what we might call essentially contestable. That are things that you can be asked "why?" Things that you can be criticized for. That set of things, they're not voluntary. It's not that you can just wheel back and decide, "oh, I want to stop caring about academic freedom, it's really getting in the way of my life," which you might

> **❝** Belief, it seems to me, is not voluntary: you can't just believe anything you want to. On the other hand, you're not just passive with respect to believing. It's up to you in some way. So, I think that anything about us that reveals our take on what's good or true or important or worthwhile or lovely or irritating, will have just the same in-between, or alternative, relation to activity and passivity **❞**

think, right? "Let's stop caring about democracy, it's really bumming me out." So, we can manage ourselves out of cares, and we can manage ourselves in and out of beliefs, that's something we can do, but if we had that kind of discretion, that direct voluntary control, over a state of mind, that state of mind couldn't have the significance of a belief or a care or a resentment.

**There are two senses of "belief," right? There's a belief in the sense of a mental state of mine, that I can be aware of. Whereas there's also the sense of "belief" where I can say, "You don't really believe that, and the reason I can say you don't really believe that is because I happen to know that you have these other beliefs that are inconsistent with that." So we can disagree your beliefs in that sense of "belief." That's a belief that is not really a mental state of yours, it's an abstract object. Is what you're talking about the mental state?**

I don't think we can make that distinction.

**So, the first kind of belief is if I could crack open your skull and go, "there it is – there's your belief!" Whereas the other is something that I attribute to you, and that if I'm Daniel Dennett I can attribute to a thermostat maybe. It's not something that is not necessarily realized in your neuro-chemistry. It's just like a proposition or something like that.**

Now, we're getting into why it is that this view does have implications for philosophy of mind. It seems to me that belief and intention are the purest case of this, but that a wide variety of our folk-philosophical, ordinary mental concepts - well, let me stick with belief and intention. It seems to me that they have their first home in our social lives. That any non-solipsistic language-using, rational (whatever that means) set of creatures, sharing a world of limited resources, will have need and use for a way of picking out what other creatures represent as so and what other creatures plan to do. That those concepts will come freighted with expectations that the other will make sense. Those expectations will be both predictive and demand-like. So, I want to be able to predict what you will do and if you don't make sense, I'll give you a hard time about it. So, that's kind of a heavy socialized concept of what belief and intention are. If you have a stubborn representation as of something being so, that won't enter into this network of expectation and demand, I don't think that deserves to be called a belief. That's a delusion. It's a fixation. It's a hallucination. It's something other than a belief. For it to be a belief, it has to be the kind of thing that interacts in this way.

**So, that the Clinton's murdered Jeffrey Epstein. What would that be? If I claimed to believe that?**

You could say, "I don't know. I just keep having this thought. It's making me crazy." That might not be a belief. But, if you come to me and say "look, we have to do something because the Clinton's murdered Epstein," that's a belief. It doesn't have to be well-grounded. I do think an awful lot of philosophy, especially in the areas that I work in, gets off on the wrong foot by thinking that reasons are good reasons, or beliefs are

> ❝ A wide variety of our folk-philosophical, ordinary mental concepts - well, let me stick with belief and intention. It seems to me that they have their first home in our social lives... If you have a stubborn representation as of something being so, that won't enter into this network of expectation and demand, I don't think that deserves to be called a belief. That's a delusion ❞

justified beliefs, or something like that. There can be a lot of going wrong and it still can be a belief.

**I keep side-tracking you. Let's go on to your discussion of responsibility. Start with what you say is the nearest thing we have to a sort of folk-concept of responsibility, which you call the merited consequence notion. Could you explain that?**

Yes. So, there's a massive philosophical literature on moral responsibility and most of it focuses on praise and blame. Most of that conceives of praise and blame as consequences deserved for good or bad choices, or good or bad behavior, or good or bad character.

**When you say "consequences," the consequences might simply be me feeling a negative emotion toward you. I don't have to hit you or put you in jail.**

That's right, but it does seem to me that it's thought of as the sort of thin end of that wedge. So, when you're a kid and you misbehave, you get sent to your room or what have you. If you're a criminal, you get some sort of criminal consequences, punishment. If you're a grown-up and you make a big mess, you have to clean it up. Then, the interpersonal reactions we have, again Strawson's reactive attitudes of resentment or distrust, end up being conceptualized as "another one of those." So, it may be that I can't make you clean up the mess, but I do feel badly toward you over what you've done. Then, the question becomes, in the literature, when is that appropriate? When is it okay to have these negative reactions to people in response to their violations or failings? So, that's what I think of as the merited consequences conception. It's thinking of all of the negative reactions we have to moral failing as a kind of maybe very thin rarified penalty that's imposed in response to violation or fail.

**And what is wrong with this view?**

That's not how grown-ups interact. This connects back to the fact that we need this category that neither voluntary nor passive. If you offend me and I resent it, if I'm thinking of my resentment as a penalty or sanction that I'm imposing on you, then I'm engaged in what I would call guilt-tripping. And guilt-tripping is a thing we do, but just to resent what you've done to me isn't necessarily to send you on a guilt trip.

**"Guilt-tripping," I think of as: in order to guilt-trip you, I have to produce the response in you. It's not simply me resenting you – it's you feeling bad. I can guilt-trip you without actually caring. I can think, "oh, you know, she did that thing to me. I know that that's the kind of thing that she might feel guilty about, even though it really didn't bother me, and I'm going to use that."**

Yes, so instead, I think that these reactions are natural reactions to the quality of will [a term introduced by Strawson] that others have to us. To resent what you said to me, say, is the way that the fact that you matter shows up in me in this relationship what we're in. In a recent piece I wrote, I said, to find yourself on the wrong end of blame, and I think "blame" is a terrible word to use for what philosophers mean by this…

> **❝** There's a massive philosophical literature on moral responsibility and most of it focuses on praise and blame. Most of that conceives of praise and blame as consequences deserved for good or bad choices, or good or bad behavior, or good or bad character… [But] that's not how grown-ups interact **❞**

**Yeah, because we talk about "I blame Trump for everything," or "I blame the Santa Ana winds for the fact that my kids were fighting…"**

Yeah. I think blame is really interesting, so we could talk about that. But, if you find yourself on the wrong end of resentment, I want to say it's more like having a hangover than it is being sent to your room. So, having a hangover is a negative consequence of drinking too much, but it's not a sanction that anyone's imposing on you. If you resent what I've done to you, when I've wronged you, that's a negative consequence of my poor behavior. It's crucially unlike a hangover in that you're active on your end of it. It's not just a passive happening like a hangover is. But, you're not necessarily imposing a negative consequence on me.

**I liked your analogy of distrust.**

That might be better. If someone is unreliable. [Cat meows.] If they regularly forget to be available for appointments, for example, then you might distrust them, but the question of whether the distrust is fair isn't a question…

**This is where you were getting into the discussion of blame. For example, your problem with the regular way that philosophers treat blame is they say sometimes blame is unfair and you say because of their misconception of blame, this is why they think blame is unfair. But, when you understand blame in the way that you think it should be understood it becomes like distrust. I think you can say it's unfair of you to distrust me, but that's only if you're mistaken [about what I'm like]. It's not to do with [how I got to be the way I am, for example] whether or not I had a bad upbringing. I can distrust somebody and you say, "Well, come on, it's because of fetal alcohol syndrome," or "he was abused as a child." And, I can say, "Those are very good reasons to distrust the person." So, you're saying blame should be like that. That it's only unfair if we attach a penalty to it and the correct understanding of blame [doesn't do that]. Now, how is blame different from resentment on your view?**

Blame in the ordinary sense or in the philosopher's sense?

**Let's go with philosopher's sense.**

For philosophers, resentment's just one possible way to blame. I could be indignant. I could give you the cold shoulder. There's any number of ways in which your failing could show up in our interpersonal relationship that would get called blame and resentment's just one piece of that.

**So, blame is a more blanket term of which resentment is an example. Now, it's also the case that resentment is kind of a cognitive notion for you. So, for example, I think you have a cat?**

Yes, she just walked through.

**It's not something your cat could do to you. I mean we could talk about my cat. My cat resents me because I forgot to feed her. Or, for example, my wife when she first came to graduate school, had to leave her cat behind because the apartment she moved into did not allow cats. When she came back to get it a year later, it stiff-armed her for a good hour. And I say it resents her. Well, you would say "no, that would be misusing the term: cats can't resent in the way that I mean it"?**

Probably not, but it's not to do with how cognitive it is. It's that to resent is to feel offended against, or wronged, or disrespected. It's to take oneself to have a certain kind of standing and to have that standing violated. Cats are certainly not engaged in developed kinds of relationships like that…

**Maybe dogs.**

Yeah. I think you get closer to it with dogs. I think there's a closer kind of system of expectations that can be mutually understood where the violation can be mutually important, not just for the fact that you're not getting your food, but that your place in this relationship hasn't been recognized. That's the piece that you need.

**From your analysis of resentment, somebody could resent the fact that you don't resent them. Could you explain that, for example?**

So, in the paper I wrote long ago on forgiveness, one puzzle is how can you forego resentment in a way that counts as forgiving. And one way of foregoing resentment that doesn't count as forgiving is just thinking the person is not worth your time. That's holding in contempt or discounting. So, if somebody just cares so little about what I think or do that my offending them is no occasion for them to have any reactive attitude, then that's a way of not recognizing my status as part of the moral community.

**Now, this is unrelated, but it's going to bug me. Do you think that is a belief that you can bring about just by itself? Do you think that I can decide to de-rank you? For example, you are my boss. You diss me, you show me disrespect, and I resent you and for my own personal benefit, I say: "This resentment is killing me. It's taking years off my life. The way I'm going to deal with it is hold you in contempt and remove you from the category of people whose respect matters." Can you just make that call? And if so, is that a belief or what is that that you're doing there?**

Right. Strawson talks about "the objective attitude" that we were talking about earlier. He thinks there are certain conditions in which we naturally tend to adopt it, such as children and people who are suffering certain pathologies, but then he says we have it available to use as a resource that we can sometimes use. And he says one reason is to avoid "the strains of involvement." So, the thought is that you can, it seems, to some extent in limited ways, sort of step away from the reactive attitudes and adopt a more objective attitude. I think we can do that to a limited degree for a limited period of time. That's not the same as believing that someone is not part of the moral community. So, to go get yourself to that belief, I don't think you can do that by deciding or at will. You can do various things to try to manage yourself into it.

> **"** To resent is to feel offended against, or wronged, or disrespected. It's to take oneself to have a certain kind of standing and to have that standing violated. Cats are certainly not engaged in developed kinds of relationships like that… I think you get closer to it with dogs **"**

**What you call managerial control.**

Yes: you can try to reframe the situation. You can try to think about it from different points of view and we do that all the time if something's really bothering you…

**Go interview their school friends and find out if they were jerks in school or something like that.**

Or try to convince yourself to focus on this aspect of the situation instead of that aspect. The reframing and reinterpreting that we do as a way of managing our beliefs and our reactions. It doesn't make them voluntary. The fact that we have to go through all of that reframing and reinterpreting sort of shows that they're not voluntary.

**It does seem that one thing we do have voluntary control over is something to do with attention. With where we put our focus. With what beliefs we're going to bring to the fore and what beliefs we're going to background, to a certain extent. So, you can put aside red flags, for example. Later on, when it turns out that they were betrayed by their lover or something, people say, "oh yes. I was aware of all those red flags, but I ignored them." So, even if you can't control your beliefs in the voluntary way that I want, you can control your focus.**

To some extent, absolutely.

**Ok. So, your slogan and I like this slogan, is "mattering not meriting." Would you explain that please?**

It was what we were just trying to talk about where I got sort of stuck. So, the thought is that the right way to think about moral responsibility, what it is to be responsible, is not to rightly merit or deserve certain sanctions or penalties or consequences. But it's instead to be such that the quality of your will, your take on the world, matters to other people. So, for me to regard you as responsible, is for me to care about how I figure into your world. For you to regard me as responsible is for you to care about how you figure into my world. And then what it is I think to be in moral relations is for us to live among one another not just caring about how we figure into one another's world, but having some shared understanding of the expectations, the rules, that amount to showing one another respect or disrespect.

**Right. Now, do you think that this view drops out of contractualism because you are a fan of Scanlon's contractualism and I imagine it's because of you that this book [What We Owe To Each Other] was seen on our TV screens in The Good Place.**

Yes.

**So, do you think that you have to be a contractualist to have this view, or that having this view is the right view to have and it leads to a contractualist view?**

It certainly sits nicely with contractualism. And, I think one good way into that book you just held up is through its sixth chapter, which is the chapter on responsibility. Scanlon

> **❝** The right way to think about moral responsibility, what it is to be responsible, is not to rightly merit or deserve certain sanctions or penalties or consequences. But it's instead to be such that the quality of your will, your take on the world, matters to other people. So, for me to regard you as responsible, is for me to care about how I figure into your world **❞**

from his initial article through to that book, has been focused on "what's the big deal about moral wrongdoing? What's the significance of moral failing?" And his answer is, it's the significance of violating a certain kind of relationship of mutual regard or mutual respect. And so there's a way to read the contractualism as starting with that core thought and then working from there to an understanding of what the standards of mutual regard would look like. So it certainly fits well with it. I'm not sure it either falls out of it or requires it. Philosophers are really clever at fitting things together with other things.

**I'm using a jargon term, "contractualism." Could you give a little precis of why it's called contractualism and what the core idea is?**

Sure. So, the core idea is that the moral principles, or the moral rules or standards are those that we would agree to under certain circumstances - it's like a contract, that's why it's called contractualism. What are the circumstances? Well, they differ for different contractualists. For Scanlon, they're the principles or rules or standards that we would agree to if we were what he calls "reasonable." He doesn't fully specify what reasonability is, but among other things, it requires being committed to finding some agreement with other people who are also committed to finding that agreement. So, to put it more intuitively, what it is to be moral, or the moral rules that we would agree to, if I were committed to and willing to constrain my pursuits for you, so long as you were also committed to and willing to constrain your pursuits for me. So, what you are looking for is kind of the freedom of each consistent with the freedom of others. Or, another way of thinking about it that Scanlon doesn't focus on is: it's the rules we would come up with if we were all equally powerful and determined to come up with rules. Then, what it is to be moral is to continue to abide by those rules, even though we're not equally powerful.

**As you put it, it's a minimal conception of morality. Could you say in what way and why that appeals to you?**

It's minimal so it stands in stark contrast with ideas of morality that have to do with a good human life. So, Aristotle, in thinking about ethics, wants to think about what's a good human life and what would fulfill human nature. That's not what contractualism is interested in. And there's other ideals you could have. You could have an ideal of rationality or an ideal of maximizing pleasure and minimizing pain. Contractualism as a moral theory is a progeny of social contract theory as a political theory. Social contract theory as a political theory was born of the need of people who disagree about God and the meaning of life to find a way to co-exist peacefully. So contractualism as a moral theory has that same minimalism to it. It's meant to enable people who may disagree about their conceptions of the good - may have different and competing ideals of human life - nonetheless to find some way to live respectfully together. Strawson wrote one other thing on ethics and it's this little piece called "Social Morality and Individual Ideal" [Philosophy, Vol 36, No. 12 (1961)], but this is just the thing that he sets up. If we want people with competing and even conflicting ideals for the good human life to be able to coexist in a single society, we need a minimal morality. We need a way of adjusting

> ❝ The moral rules or standards are those that we would agree to under certain circumstances... it's the rules we would come up with if we were all equally powerful and determined to come up with rules. Then, what it is to be moral is to continue to abide by those rules, even though we're not equally powerful ❞

those ideals as little as possible, but in a way that makes room for other ideals to be adjusted as little as possible.

**You wrote a fairly long piece ["Of metaethics and motivation: The appeal of contractualism"] in a collection of works on Scanlon [*Reasons and Recognition*, eds. Wallace, Kumar & Freeman (Oxford, 2011)] defending his contractualism and one of the criticisms that you take on is Judith Jarvis Thomson's "torturing babies" objection. So, perhaps, you could give that and explain why Scanlon's theory survives it.**

Right. So, Scanlon says an action is wrong just in case its performance under the circumstance would be disallowed by principles that no one can reasonably reject for the general regulation of behavior. Something like that. That's his formula. And, he says that's what it is for an action to be wrong. For it to be disallowed by the set of principles we could all agree to. And, Thomson says this is insane. Torturing babies for fun would be disallowed by those principles because it's wrong. And Scanlon gets this back to front.

**So, it's the Euthyphro problem.**

He says it's wrong because it's disallowed. So, I think that's a very good objection to raise against Scanlon in part because it brings out the distance between his ambition and aim and the ambition and aim of other moral theorists. So, he accepts the criticism. That is to say, he says, no, he stands by his theory. He says no, what it is for torturing babies to be wrong is for it to be disallowed by these principles. So, the explanatory priority goes in the direction that he said it went in, but what's motivating Thomson's example is that torturing babies seems so obviously wrong and it seems horrible. So, there are two things to be said. One is the fact that something is horrible is not the same as the fact that it's wrong on anybody's theory. So, everyone's theory needs you to add something to the horribleness to get the wrongness. So, there's a way in which the criticism is unfair. But then the other piece that it brings out about Scanlon's theory is that it needn't be the case that the wrongness of something is the most important thing about it. So, when Scanlon talks about wrongness he wants his theory to capture the kind of importance morality has for those of us who think it has importance. He wants to capture that, and he wants to show why it normally takes priority in cases of conflict. Neither of those is to say why it's the most important thing in any given situation. He just wants to say the kind of importance it has and then why it should take priority in cases of conflict. But he can have that whilst acknowledging that the horribleness might be the most important thing in any given context. Not the wrongness.

**I wrote a note to myself and perhaps you can say if there's something wrong with this, that Scanlon's view explains why we give the label "moral" to some demands, but it doesn't explain the oomph behind them. So, for example the oomph behind our abhorrence of torturing babies is not because it's immoral. But, Scanlon's theory can tell us why "don't torture babies" is a moral demand.**

That's right.

> ❝ What it is for torturing babies to be wrong is for it to be disallowed by these principles... [But] it needn't be the case that the wrongness of something is the most important thing about it ❞

**So, for him, moral demands are a category that doesn't reduce to the various reasons we have for doing them, or the most basic reasons. In other words, he's a non-reductionist about what morality is, but his theory explains what all moral demands have in common and what makes them part of the moral group rather than say the taboo group or whatever.**

Right. It's a two-level theory. So, the moral principles are the ones we would all agree to, but the basis for that agreement looks to the reasons for rejecting principles which will have things to do with pain and pleasure and convenience and ambition. The wrongness of something depends on the way it affects individual interests but doesn't reduce to the way it depends on individual interests. So, there's a way in which it's similar to rule utilitarianism. Rule utilitarianism is also a two-level theory. It gets criticized because people don't see why they should follow the rules, once we notice that it's two levels. Scanlon has an answer to that, namely, the importance of standing in certain relationships. Sorry, that might have all gone too fast.

**[According to a famous argument by J.J.C. Smart] rule utilitarianism collapses into act utilitarianism and you're saying [Scanlon] forestalls [an analogous] collapse. So, for Smart, for any rule, if it is not the act utilitarianism maximizing choice, then we have a more fundamental reason for not abiding by the rule than we have to abide by it. You're saying that doesn't happen with Scanlon because respect is serious enough that's that always going to be a reason.**

So, you rule out free riding for example, even though no one would really be harmed if I free ride on this occasion, but it's still a serious wrong. Far more serious than what's actually at stake if I cut in line and make everybody wait another three seconds, because it's communicating that I'm not part of this network of people who are constraining their pursuits in light of the pursuits of others.

**This reminds me again of my problem with the status of beliefs: whether beliefs are things in individuals' minds or abstract entities. Related question about respecting: can I disrespect someone without them ever knowing it?**

Yes.

**So, in other words, I can disrespect someone by free riding even if nobody knows.**

Yes.

**So, in that case, it's starting to sound a little bit Kantian in that I'm violating a rule that sort of offends against rationality or something, rather than I'm having a bad effect on people.**

In the full-on Scanlon contractualism, the rules are themselves grounded in the fact that we're each according the other symmetric standing to determine what the rules will be. So, by violating the rules, I have thereby, whether you know it or not, acted in a way that does not accord you symmetric standing with me to determine how I will act.

To determine the minimal terms of our cohabitation. So, respect is still relational even though it doesn't require your knowledge.

**Ok. Now this all ties back to the title of your larger book, Minds That Matter, because the mattering is related to the respect and the disrespect.**

That's right.

**You want to tie together a bunch of threads from the various works that you've done. There are these works on blame, there are works on forgiveness. Now, those seem basically intersubjective. That there has to be awareness. Right? Whereas there's also this aspect that it doesn't matter if people are aware. So, do you think those are compatible? That you can have both things?**

I'm not a fan of awareness… I hadn't thought about it until you put it that way… In terms of control, it's a crucial feature of ordinary control that you have some sort of awareness about what you're doing - [Elizabeth] Anscombe pointed that out - and it's a feature of evaluative control that it does not require that same form of awareness.

**Right, so as an example, I can change my beliefs almost as a side effect without even noticing it. By investigating things, it just so happens that I no longer believe that the butler did it because I see exculpatory evidence. So, my belief results from other stuff that I've done, but you want to say I have again in some sense got control over that belief even though my change in belief has happened to me, one might say.**

I want to say the fact that you didn't notice it is no reason to think it didn't happen to you. You also might realize that you've decided to attend your own graduation, let's say. Somewhere in the background, you made that decision without having to pull it into the foreground, so to speak, and think about it. So, I argue against the importance of awareness of control in the first half of the book. When it comes to forgiveness and blame and that sort of thing, certainly the normal case is the case where people are aware of one another's forgiveness or blame or disrespect or what have you, but I'm not sure that that's essential to it. I mean, so a standard thing is, can you forgive the dead? Can you wrong the dead? Can you resent things done by somebody who you're not in communication with? All of that seems right to me. That all those things can happen, and so I wouldn't want to insist on awareness there either. But, they're still intersubjective. There's still the relation between subjects.

**Yeah, because otherwise why is it important that they be minds. You could say stones "matter" to each other without ever being aware of it. That the top stone in a bridge and the bottom stone in a bridge matters to each other. But your notion of mattering is tied to awareness or at least consciousness in some sense.**

No.

**At least the capacity?**

> I'm not a fan of awareness… I want to say the fact that you didn't notice it is no reason to think it didn't happen to you

Yeah. So, I have tried to stay far away from consciousness as a topic. It does seem to me that people think of consciousness and control and volition and awareness as very tightly held together. I don't see it. So stones don't have minds, and I do think it's minds that matter to one another, and maybe minds and consciousness go together. I don't know about that. I'm very wary of consciousness.

**There's only so much you can take on. If you're taking on free will, you can't also be asked to do the other biggie.**

I mean it is the case that, maybe I can just say a little bit why awareness seems to me misleading. So, if I'm thinking about the butler and trying to decide whether he did it, my focus, my awareness, my consciousness, in some sense, is on the butler and things at a distance from me. If I come to the conclusion that yeah, I think he did it, I'm coming to a conclusion about him, but I'm making something true right here at home. But the thing I'm making true is not the thing I'm thinking about. The thing I'm thinking about is out there. The thing I'm making true is here. So, there's an indirection to that particular exercise of agency. That's why I think that I can decide to attend graduation without noticing. It doesn't have to be "on screen" to be my doing.

**It's funny. All of this sounds rather Wittgensteinian, and it's funny that you cite Anscombe and you cite Strawson who was doing his work around then and I just think of a passage I remember reading from Wittgenstein about expectation, where he describes Smith going around, you know, lighting candles, emptying the ashtrays, and it says that he is expecting Jones, (it's always Smith and Jones), and expectation is realized not between Smith's ears, but expectation is something that you can say occurred because of all of this evidence that is out in the world. It's almost as if you can talk in terms of something that most people would consider a mental state in non-consciousness terms. It sounds like something like that is what you would find appealing.**

That's stronger than what I want. But, it's in the ballpark.

**[Before we wrap this up], there are a couple of things. You used this phrase and you don't mean by it what Catholics mean by it: "original sin." What is it and why does the fact of original sin help inform your view of ethics?**

What I mean by original sin is standing in contrast to a widely held view that gets put in the slogan that "ought implies can." So, people widely think that if you can't do something then you can't be required to do it. Kant used that to argue for a very strong form of freedom. So, he's worried about Newton and the deterministic world, and then he says we know that we're under obligation, and so we know we're free. We can reason from obligation to freedom.

**When you say you can't be "required" to, you mean morally, not literally forced to.**

That's right. You cannot be morally required to. You can't be blamed if you don't do it, that's one claim people make. Now, people are less sanguine than Kant about our

capacity to always do what's right, and so now people think that moral demand or moral obligation or certainly at least blame, should shrink to fit ability. I think that's not so. If it were the case that moral demand shrinks to fit, it would make it unlike almost any other kind of expectation that we hold one another to. So, the demands of parenting, for example, are grounded in the needs of the child. The demands of policing are grounded in the needs of the community. The demands of being Dean are grounded in the needs of the institution and if a particular parent or police officer or Dean is such that they can't satisfy those demands - they're too selfish, they have too much of a temper, for some reason they can't pull it off - the demands don't shrink to fit their ability. So, if you think, like a contractualist does, that the moral demands are the requirements of sharing a world together with others who are equally real or equally important, then, like parenting or policing, the moral requirements are grounded in other people's standing and won't shrink. But that means, given the hazards and contingencies of moral development and given that it's difficult to become somebody who can put aside their own screaming needs and allow other people to exist, that sometimes people will arrive at adulthood unable to satisfy demands that are nonetheless rightly placed on them. And that's what I call original sin: the idea that you can find yourself in situations in which there's a legitimate moral requirement that you just cannot satisfy.

**So, basically, you're saying that being immoral is like being an asshole. It might not be your fault that you're an asshole, but you're still an asshole. It might not be your fault that you're being immoral, but you are still immoral. So that's why it is connected with your notion of blame. You can blame someone even if it might be unfair to punish them. But because your notion of blame is not the merited consequence one – it's to do with holding certain attitudes – then you're still okay to blame someone.**

Exactly. And recently it may be that [philosophers] actually have different conceptions, actually different attitudes. So, I've now started talking about resentment–plus and resentment–minus, where when I talk about resentment, I mean resentment-minus. It's a reaction to offense. So, in the forgiveness paper, I talk about it as a form of protest to a wrong done. It affirms my standing in the face of disrespect. I think some people think of resentment as what I now call resentment-plus, which says not only "you can't treat me like that," which is what resentment-minus says, but "you can't treat me like that and if you would pay attention and tried harder, you could have avoided this." So, if you think that resentment just has that packed into it, that that's part of what it is to resent, then when you learn that somebody is subject to original sin and is just too much of an asshole to avoid it, for example, then you're going to think that resentment is now out of place. But, one way around that is to think that that's not packed into resentment. That I can perfectly well hold together resenting the chauvinism I've been shown while also believing of this person that there's not a path for this person to avoid that.

**Again, I'm struck that your view seems, this is an inexact word, but intellectual. In the sense that, would you say that it is true that you could discover, "oh, I guess I resent that person!"**

Oh sure. I think that happens all the time.

> **❝** If it were the case that moral demand shrinks to fit ability, it would make it unlike almost any other kind of expectation that we hold one another to... Sometimes people will arrive at adulthood unable to satisfy demands that are nonetheless rightly placed on them. And that's what I call original sin: the idea that you can find yourself in situations in which there's a legitimate moral requirement that you just cannot satisfy. **❞**

**Whereas I think my gut level version of resentment is it eats me alive. It is a gnawing feeling. That is the most important thing about resentment: "Oh, my resentment of him is just driving me crazy." Whereas you're saying, "oh yes, I see that I believe that they've disrespected me. I see, you know, so and so, therefore, I resent that person." It seems like the affective component is more important than you give it credit, at least in the sketch that you've given.**

I see what you're saying. It's also the case that we might come to understand that this gnawing unhappiness you've been experiencing, is actually resentment.

**So: what I've come to understand is how to conceptualize my anguish or my affect? So you're not committed to the view that you can have that without the affect?**

I'm not committed to that view. I suspect you can, but I'm not committed to it. I suspect you can because I want to say things like, "oh, I've resented him for years, but I haven't had this affect for years." But, my reason for treading lightly on affect is just that my methodological tools are not those of a psychologist. My methodological tools have to do with relations of justification. So, I can talk about reasons, I can talk about agency, I can talk about relations of support between considerations and conclusions. So, I think that does give me a handle on some of our emotional lives, but it's just one little handle on it. And this thing is much bigger, but I feel like I'm not in a position to talk about an awful lot of it.

**I wonder if a contractualist counselor could get someone to deal with their resentment saying, "I can explain to you why resentment is illogical and let's see if the affect vanishes."**

It's a thing that's right there in the beginning of that forgiveness paper that there are conclusions you can come to that should just extinguish resentment. For example, when you learn that it was a mistake, that there was no ill will shown. It seems like that revises - I think that is the word I used there - you then revise your resentment. That's very different than the sort of process of trying to reframe, where you're trying to manage your way out. That's different again from inducing amnesia about the relevant stretch of history and so, you know, getting rid of the resentment.

**I think children are very problematic cases, of course, because I think when you're raising children, like toddlers and so on, they can do things that seem extraordinarily insensitive and you can genuinely resent them even though you know this is stupid. They are not capable of intending what I'm taking them to, but they seem like they could be, given the way they behaved. So, I believe that you can have the attendant feelings even while you're telling yourself "this is stupid." I think that's also a problem, for people like autistic people, who can say what seem like extraordinarily insensitive things or can appear to disrespect somebody without intending it at all.**

Yes, but the fact that there's that tension between the thought that it seems that they've disrespected me but they're not capable of that kind of delay of gratification or what

have you, that there's a tension between those thoughts, I think, says something about its cognitive aspect.

**So, suppose my toddler does something that I take is very disrespectful or showing ingratitude, and I feel this feeling, and then I work it through and say they're not disrespecting me, but the feeling doesn't go away. Is it no longer resentment because my beliefs have changed, it's just the residue of resentment? Or is it always a mistake to talk about the feeling as resentment?**

An even clearer case is the person who is afraid of getting on an airplane but believes it's safe. On the one hand, they're definitely afraid. On the other hand, they definitely keep telling you that they know that it's safe. The fact that we feel a tension between these… You know, it's not like having indigestion and believing the airplane is safe.

**See, here I would say again, there's the two kinds of belief. Because in the one sense they can say, "I honestly don't believe that it's going to crash," but then you would say, "Yes, you do. You do believe it's going to crash, because you're afraid."**

I want to say both of those things are true.

**So, you both believe and don't believe.**

People are complicated. We can find ourselves in lots of conflicted situations. So, you both are committed to [the belief] and committed to its negation. One motivation for talking about questions and their answers is to move away from what was called a while ago, judgmentalism about emotion. So, people used to think emotions have beliefs built into them. If you think of belief as a state of mind with its own functional footprint, so to speak, you might not want to insist that emotions have beliefs built into them. But you might still want to think that they have a kind of representational commitment to them that is something less than belief.

**Well, your work covers so many areas. This book is going to be huge.**

Right now, it's about 300 pages. I would like it not to get any bigger than that.

**Ok, so I'm going to wrap this up, otherwise, I will keep you talking for hours. What I normally ask is: What do you think philosophy's role in the public life is? Now, for you, this has a particular connection because you have been a consultant for a TV show [The Good Place]. Do you think that something good for philosophy has come out of that?**

Yes. I do. I think it's wild and wonderful that contemporary philosophy has gotten a little bit of product placement on television.

**It's not just one, there's two sitcoms. There's another one called AP Bio where the main character is not just a philosophy professor, but obviously they have a consultant because he drops in real philosophy once in a while.**

> 66 People used to think emotions have beliefs built into them. If you think of belief as a state of mind with its own functional footprint, so to speak, you might not want to insist that emotions have beliefs built into them. But you might still want to think that they have a kind of representational commitment to them that is something less than belief 99

Yeah and there was another [Here and Now]. I think it went on and off the air with Tim Robbins as a philosophy professor.

**I missed that one.**

I think it's good and do I think good things have come out of it? Certainly, there's some interesting philosophy on the part of students and people that I've encountered. But I think there's a much bigger role that philosophy can and should be having in public life. And, I think the kind of political and cultural upheaval that we're in the midst of, it would really terrific if one aspect of that was that philosophers became a little less insular and a little more engaged in the wider dialogue. Which I think is actually happening in certain ways.

**In England, more than here, there are pop philosophy books. There is a guy called Stephen Law, who is at [The University of London] now, who wrote a very good philosophy book forabout a 12-year-old. It was called *The Philosophy Files*, because The X-Files were popular at the time. Do you think that's what we need? Because philosophy is not studied in high school and there is a little bit of a danger that someone could see *The Good Place* and say, "Chidi [a well-meaning but multiply-flawed character] represents philosophy and that's just, you know, not necessarily one of the good options."**

Right. I guess I want to say all of the above and more. I'm happy to have Chidi. I'm happy to have The Philosophy Files. A man I went to graduate school with, Aaron James, has a book that I've actually used in my classes - *Assholes: A Theory* [Anchor, 2014].

**In it it gives as a paradigm example, "real estate magnate Donald Trump."**

Yes. Well before the election. There's a follow-up of course - Assholes: A Theory of Donald Trump.

**I haven't seen that one.**

Yeah, the publisher did not miss the opportunity to repackage. But I think he's done, it's more or less contractualism. So, when you said, "Oh, being immoral is like being an asshole", you might think, no, it is being an asshole! That's kind of what that book's about, but it's available, it's readable for a general audience. There's, The New York Times has The Stone, the column that seems to me a good development. Kate Manne has her book on misogyny [*Down Girl*, Oxford, 2017].

**So, YouTube has been taken over more by the Alt-Right, or it was. And in response to that, we have a lot of more left-wing people who are employing philosophical tools, but in a very poppy way. I mean, I wonder if philosophers need media training. Your training in your non-profit perhaps could be put to use. Why do you think they came to you about The Good Place? Mike Schur, the show runner, was he genuinely**

**working through philosophical problems in his head? He wanted help in thinking them through in this way?**

He genuinely was. It was remarkable. So, I'm not exactly sure, but he had this idea, it was before he had assembled his team, he was still developing it, and somehow, I don't know how, he came across my web page and on my web page I had a description of a paper that I still haven't gotten out for public consumption. The paper at one point was titled, "Why You Can't Be Good By Trying." The idea of the paper is, Aristotle says roughly, "Fake it 'til you make it", but if practice makes perfect, why wouldn't you become really great at faking it? There's got to be some transition from faking it to the genuine article and why would practice help with that? So, that was the topic of the thing and I think he was interested in that. So, I think that's why he contacted me to have a conversation about his idea. I actually think we need a lot more collaborations between philosophers and non-philosophers who are either people writing narratives or journalists. So, since my involvement in that show, I have had a number of interviews by journalists asking me about the show and a couple of them, I think have done an amazingly good job of translating the things I said to them into popular journalism, in a way that I don't think I could. I mean, I'm trained to do this other thing and it gets too involved. And my collaboration with Mike Schur, part of it that was extremely gratifying was I walked away realizing not only do I know a lot of stuff, but if you put me in a conversation with a smart person who has a project like this, I can package it for him. He doesn't have to do all the reading and work through all the things that I've spent my lifetime doing, but I can deliver to him, pieces of it.

**You deliver to him and he delivers to the general public.**

Yeah, and a bunch gets lost in the way and in the process. So, I do think philosophy needs to be more engaged, but I think one way for us to do that is to find partners in collaboration whose expertise and training is in communicating in ways that are not the ways we communicate.

> **"** I actually think we need a lot more collaborations between philosophers and non-philosophers who are either people writing narratives or journalists. So, since my involvement [with *The Good Place*], I have had a number of interviews by journalists asking me about the show and a couple of them, I think have done an amazingly good job of translating the things I said to them into popular journalism, in a way that I don't think I could **"**