

Thinking by Drawing: An Interview With Shelly Kagan

SHELLY KAGAN (Chicago, 1954) is Clark Professor of Philosophy at Yale University, where he has taught since 1995. Before coming to Yale, he taught at the University of Pittsburgh and the University of Illinois at Chicago. Kagan received a BA in Philosophy and Religion from Wesleyan University in 1976 and a PhD from Princeton University in 1981. His PhD thesis, on the limits of what morality can demand, was supervised by Professor Thomas Nagel.

Kagan's research is in normative ethics. He is the author of *The Limits of Morality* (1989), *Normative Ethics* (1998a), *Death* (2012a), *The Geometry of Desert* (2012b), and *How to Count Animals, more or less* (forthcoming). His work has appeared in various journals, including *Ethics*, *Philosophy & Public Affairs*, and *Proceedings of the Aristotelian Society*. Videos of Kagan's course on death have been very popular online. Kagan is a fellow of the American Academy of Arts and Sciences and a member of the advisory board of the Philosophical Gourmet Report.

The Erasmus Journal for Philosophy and Economics (EJPE) interviewed Kagan about his formative years (section I); his work on death (section II), the moral status of animals (section III), and desert (section IV); his views on convergence in philosophy (section V); and his advice for graduate students in moral philosophy (section VI).

I. Formative years

EJPE: Professor Kagan, what first drew you to philosophy?

SHELLY KAGAN: I was interested in Jewish religious thought from an early age. In highschool, I was reading Martin Buber, a great 20th century Jewish thinker. One of the essays I read was Buber's reply to Kierkegaard's *Fear and Trembling* (1983), which discusses the binding of Isaac in Genesis. I had not read any Kierkegaard and did not know

EJPE's NOTE: This interview was conducted by Huub Brouwer. Brouwer is a PhD candidate at Tilburg University, specializing in desertist and luck egalitarian theories of distributive justice. At the time of the interview, he was on a research visit to the Yale philosophy department. Brouwer thanks Christiaan Broekman, Willem van der Deijl, and Julien Kloeg for discussions on Kagan's work in preparation of the interview.

anything about him, and there was a lot in the Buber text that I found difficult to understand. I realized that I needed to get exposed to philosophy.

My father had a very good book collection. So I went down into the basement, where his library was, and started reading some Plato. I must have known enough about philosophy to know that I had to start there. After the Socratic dialogues, I read some of Aristotle's *Nicomachean Ethics*. I then read some Bertrand Russell. At that point, I was hooked. I had no intention of becoming a philosopher yet, but I realized that philosophy was something that I cared about deeply.

There is a wonderful quote by Naomi Scheman that I sometimes bring up when talking about my early interest in philosophy: 'Taking my first class in philosophy was like hearing my native tongue spoken for the very first time'. That sums up very well how I felt.

Did you have people you could talk to about philosophy during your childhood?

No, not really. I was a philosophically inclined child, but I had nobody who knew enough about philosophy to be discussing it with me. I actually think that is a fairly common situation. Most kids are natural philosophers and are bothered by questions similar to those that were bothering me. They get these things driven out of them by their parents and teachers.

I'll tell you an example of a philosophical question I was worried about from when I was nine or ten. In my English class, the teacher was trying to explain the difference between nouns and adjectives. I did not get it. The teacher said: 'Look, a brown dog; 'dog' tells you what it is, and 'brown' tells you something about what it is like'. And I said: 'Brown dog? 'Brown' tells you that it is a brown thing, and 'dog' tells you what kind of brown thing it is'. My teacher had no idea why I could not see the distinction. It wasn't until I had read Quine that I realized what I was worried about.

In highschool, I started having what I thought were philosophical discussions with some of my friends. It wasn't really until I went off to college, though, that I began studying philosophy in a systematic way.

You did your BA at Wesleyan University. Why did you apply there?

I wanted a small liberal arts college for my undergraduate experience. I had the sense that you would get more face time with faculty members

at such a college than you might at a research university. Many people at research universities primarily care about doing their research; teaching is something they do to pay the bills. Many people at liberal arts colleges, on the other hand, are especially interested in teaching. So I looked for a small liberal arts college with a really good academic reputation. I wanted to have good teachers and smart fellow students who would want to sit around and discuss ideas with me.

You majored in philosophy and in religion. Why both?

I actually went off to college to be a religion major, and so I was. I also took a bunch of philosophy classes, but did not declare my philosophy major until the second semester of my senior year, when I was looking at my transcript and realized that I had taken enough philosophy courses to declare a philosophy major. My philosophy major technically was an afterthought.

What did you want to do with the religion major?

I went to college planning to become a rabbi and expected that to be my career path until almost the end of my senior year.

What changed?

I applied to a particular rabbinical school—and they rejected me! I got the rejection letter during the spring of my senior year. What this meant was that, for the first time, I had the kind of introspective examination of the soul that many people have throughout college, asking myself: ‘What do I want to do with my life?’

One option, of course, was to reapply. The seminary I applied to had a reputation of regularly turning people down—perhaps to see how committed they really were. There actually is a tradition in Judaism that when a non-Jew comes to a rabbi, and says ‘I want to convert to Judaism’, the rabbi is supposed to turn them away. It is not an easy lot in life, being a Jew. If the person comes back a second time, the rabbi is supposed to turn them away again. It is not until they come back the third time that the rabbi is supposed to take the person into his study and explain what would be involved in converting. The admissions policy of the seminary was probably based on a similar thought: the lot of a rabbi is not an easy one, we want to make sure that you are dedicated.

Other options I thought about were applying for a secular PhD in Judaica, or, and this is something I started thinking about for the very first time, doing a PhD in philosophy. I discussed the idea with my philosophy teachers, and they reassured me that I was good enough, that I could get into some of the good places with their letters. In the end, I got into Princeton.

With hindsight, there were some clear indications that this had been brewing. Even before I had gotten the rejection notice from the seminary, I remember asking one of my philosophy professors: ‘How do you keep up with philosophy if you don’t become a professional philosopher?’ He looked at me and said: ‘You don’t’. I remember thinking that was sad; I would really miss doing philosophy.

You may remember the passage in Russell, I can’t remember where he says it exactly, where he says that people come to philosophy in the main from two different directions. From religion or the sciences.¹ I didn’t come from the sciences. I came to philosophy from thinking about the meaning of life and ethics. I think that is the explanation for why I went into moral philosophy. It may surprise you, though, that I actually enrolled at Princeton thinking that I was going to do metaphysics and epistemology. That’s what I said in my personal statement. A real case of lack of self-knowledge, I suppose.

That sure is surprising! I’d like to talk about your undergrad period a bit more. Which people and writings were particularly important for the development of your interests during your BA?

In my freshman year, I took a class where we read *Utilitarianism*. I remember thinking: this is the truth. What I don’t remember anymore is whether I already was, without realizing it, a utilitarian—or, whether Mill converted me. Somewhat later on, I read G. E. M. Anscombe’s “Modern Moral Philosophy” (1958). In her diatribe against a century of philosophy, she introduces the word ‘consequentialism’ to condemn everyone to the left of absolutist deontology—saying something along the lines of: ‘As if having good consequences could be enough to justify doing an action!’ I remember thinking: ‘Of course that is what justifies doing an action! What else would justify it?’

¹ Russell’s exact words are: “Philosophy, historically, is the intermediate between science and religion” (1955, 34).

I also read some G. E. Moore, and liked Moore's style. I know people find him plodding, but I found his drive to state with precision what was at stake very congenial. Long before I discovered that there was a thing called 'analytic philosophy', I was drawn to that kind of approach to philosophy. I remember reading Hegel during the fall of my sophomore year, and finding it completely opaque. I did not have the reaction: 'This is such great stuff, I have to master this language'. My reaction was much more: 'Why does this guy have to write so obscurely?' Temperamentally, I was driven to get clear about things.

Finding analytically trained philosophers, later on during my undergrad years, was finding my people. It is not that I bought all of the more dogmatic elements of analytic philosophy, though I went through phases. There was a period when I was arguing with my friends that when you have philosophical debates, all you are arguing about is the meaning of words. Now, I look back and think: 'What a dumb thing to say!'

And how about your graduate studies? Which people and writings were influential then?

Tom Nagel and Tim Scanlon were at Princeton when I arrived there as a grad student. Derek Parfit visited Princeton during my first or second year. I formed the opinion then, and nothing has led me to revise it since, that Nagel, Parfit, and Scanlon are just wonderful moral philosophers. They taught me not to be satisfied with a quick, dirty objection to a philosophical position; to really think about what makes a view promising and tempting—even if, at the end of the day, you want to reject it.

It easiest for me to identify the influence of Nagel and Parfit, because I went on to work closely with them. I got to know Nagel when I took a graduate seminar with him during my first semester at Princeton. He became my PhD supervisor later on. I think I sum up his influence well in the acknowledgements to my thesis, which was published a few years later as *The Limits of Morality* (1989). I wrote: "My debt to him will be obvious to all those who know his work, and even more obvious to those who know him personally. If I have any sense at all of philosophical debt, it is thanks to Tom; if I do not, it is not for lack of his trying to teach me" (1989, xiii).

Parfit was working on some early material for *Reasons and Persons* when he visited Princeton. I took a seminar with him in which we

discussed it. I found it fascinating, utterly fascinating. I had never read anything remotely like it. Parfit gave me the realization: ‘Oh, you can think about things by drawing boxes!’ A fair bit of my work has diagrams and formulas; not because I think that they get it all, but because sometimes they allow you to see more clearly what you are committing yourself to.

In Parfit’s case, it seems that the influence was mutual. He writes in the acknowledgements to *Reasons and Persons* (1984, viii) that you are the person he learnt most from.

That was an amazingly generous thing for him to say! I have often joked, when people quoted that back to me, that it is a pity that I didn’t die young. If I had died young, people remembering the Parfit quote would say, ‘Shelly had such promise’. But if I die now, people will say ‘Shelly had such promise, but he didn’t really follow through’.

How did you get to comment on *Reasons and Persons*?

Parfit would draft material and distribute it among people whose opinion he respected. He would then try to come up with responses to the objections these people raised, and send a revised draft to them. Parfit would go through this cycle over and over—until he was satisfied, or some deadline had been reached.

When I got out of graduate school, Parfit was sending around whatever the latest version of the manuscript of *Reasons and Persons* was. I sent him some comments on Part I. My intention was to turn my attention to revising my thesis for publication after that.

One day, some time after I sent the comments, Parfit called me and said ‘I love your comments, would you be willing to send me comments on the rest of the book?’ I said: ‘I’d very much like to, but I don’t really have the time. I have to revise my dissertation’. Then he said: ‘Well, I will make you a deal. If you send me comments on the rest of *Reasons and Persons*, I’ll read your dissertation and send you comments on it’.

I knew a good deal when I heard one. Parfit had a reputation as one of the most deep, trenchant critics in the profession of moral philosophy. If he was willing to do this for me, absolutely! I loved the *Reasons and Persons* stuff, and this gave me an excuse to not just read it, but study it—think through it paragraph by paragraph. I devoted a summer to doing that. It was a wonderful experience.

How were Parfit's comments on your dissertation?

His comments were great! They came at a good moment as well. After finishing my PhD, I started as an assistant professor at Pittsburgh. I prioritized my teaching, as I always have since. One day, David Gauthier, the department chair at the time, called me into his office and showed me a graph on his blackboard. On the x-axis were the years that I had been at Pittsburgh, and on the y-axis was the number of publications. I had not published anything!

David said: 'Look Shelly, we really like you. We think you are smart. We want you to stay. But you've got to publish some stuff!' I told him that I was planning to revise my dissertation. In my memory, it was that very night that Parfit calls me a second time. This is too good a story to be true, so I suspect I just conflated the two events. But, at least in my mind, that very night, he calls me and says: 'I have read your dissertation. I really like it, and I want to publish it with Oxford'.

I could not believe it. I said, 'I just want to get clear on what you are saying, Derek. You think my dissertation is pretty good and you are going to help me by giving comments. When it is revised, you'll see whether Oxford is interested in publishing it'. He said, 'No, no, no! I am doing a new series for Oxford, I am the editor for it, and I want to publish it in my series'. When I hung up the phone, I remember thinking: 'Well that changes the graph now, doesn't it?'

You remained in Pittsburgh until 1986. You then went to the University of Illinois at Chicago, why did you move?

The Pittsburgh philosophy department was wonderful. It was as if I could continue my graduate education there. I would sit in on seminars from my colleagues and have regular one-on-one discussions with David Gauthier and Kurt Baier. David Gauthier was working on *Morals by Agreement* (1987) at that time. We would have weekly meetings to discuss the manuscript. Kurt Baier read my thesis, and met with me every week to talk about it. He would help me to see things that I would need to revise and fix.

My then girlfriend, and now wife, Gina eventually joined me in Pittsburgh. For professional reasons, Pittsburgh did not really work out for her. That is why we left. We moved to Chicago, which had a big plus for me, because I grew up in a suburb of Chicago, called Skokie. I am one of five children and the only one who ever really left Chicago. The others went to school either in Illinois or a nearby state, and all came

back. My parents were still alive, and I have an extended family of cousins, almost all of whom had stayed in the Chicago area. So there was this huge network of family that was just wonderful.

You became a full professor in Chicago in 1994. In 1995, you went to Yale. What made you decide to do so?

I was very happy in Chicago. But my wife's family was out East. So we would talk about moving East at some point. The list of options was short though: There only was a small number of departments that I would have considered a step up from Chicago. Although the department wasn't top 10, it was always in the top 20. I had great colleagues there: Michael Friedman and Anil Gupta, among many others.

I am not sure how widely known this history is anymore, but the Yale Philosophy department was in bad shape at the time. There was a lot of infighting. Yale had hired Robert Adams during the early 1990s to rebuild the department. I had known Bob since I was a graduate student. He and his wife Marilyn had spent a year visiting Princeton when I was a graduate student there. I really respect Bob as a philosopher, and his taste in philosophy. I thought that if Bob was behind this, then maybe the department was going to be turned around.

One day, I was looking at 'Jobs for Philosophers', which, in those days, was a newsletter published by the American Philosophical Association. There was an ad for Yale, looking to hire somebody in moral philosophy willing to be involved in an interdisciplinary program. I remember seeing that ad and saying to Gina: 'This has my name on it'.

I applied to the job. Bob wrote back quickly, saying that he was going to take my application very seriously. Eventually, he made me an offer and I started to have a series of discussions with him about his vision for the Yale department and the kind of resources that Yale was putting behind the effort to rebuild the department. It was all completely reassuring. There was something about the chance to be a founder of a great department that tickled my fancy; perhaps it would be more accurate to say *refounder*, actually, because Yale used to have a great department. Anyway, that's how I ended up here at Yale.

II. Death

Many people know you because of your class on Death—which was recorded in 2007 and put online as a Yale Open Course. How did you get interested in death?

The story of the class actually goes back to my time in Chicago. When I joined the Chicago department, Jerry Dworkin, the department chair, said to me, ‘We used to have a philosopher in the department who taught a class on death. It was reasonably popular. Any chance you might be interested in teaching it?’ I told him that I would think about it. Jerry didn’t say what they covered in the class. When I started thinking about ways of setting it up, I was just blown away by the possibilities. Death allows you to talk about so much: metaphysics, epistemology, personal identity, value theory, the nature of emotions. I agreed to teach the class in Chicago, and continued teaching it here at Yale.

In the first chapter of your book *Death*, which came out of the class, you write that there are roughly two ways to write a philosophy book (2012a, 3).² One way is to lay out various alternatives, the pros and cons, without taking a position yourself. The other is to state what your own position is and defend it as well as you can. In the course and in the book, you clearly take the latter approach. Why so?

On many topics in philosophy, people do not really have views. When you teach a class on these topics, you have to lay out the possible views for them. That is not the case with death. Death is a topic where lots of people think: ‘Such and such is certainly the case’. I happen to disagree with many of the standard views about death. Many people here in the United States think that we have a soul. I don’t. Many people believe immortality would be a good thing. I don’t. Many people believe that we should be afraid of death. I don’t. Many people believe that suicide is always irrational. I don’t.

Given my disagreement with so many of the standard views about death, it seemed to me that it might be valuable to lay out my cards right at the start, and say: ‘Here is a bunch of things that many of you believe about death. I think it is all wrong!’ I probably wouldn’t have chosen this set-up if I had not rejected the standard view about death so

² All references in this section are to Kagan’s *Death* (2012a), unless otherwise indicated.

completely. It was easier, cleaner, and more upfront to do it this way. And it seemed to work. Although particular topics and arguments covered in the class changed over time, I was happy with the basic approach from the get go.

You say at the start of your book that you will not try to argue the reader out of their religious beliefs (5). But don't some of your arguments go against religious beliefs? To give just one example: you argue extensively against the existence of immaterial souls (chaps. 4–7). If there are no immaterial souls, then this seems to be at odds with religions that postulate their existence.

It is true that, on certain religious beliefs, there are immaterial souls. Anyone who is claiming there are no such things would be threatening this religious belief. The arguments I am giving, however, speak to the conclusion that we would be justified in reaching, religious authority aside. I claim that there is no good reason to believe in the soul on a non-religious basis, but that is perfectly consistent with believing in the existence of a soul on religious authority. The believer and I are simply making distinct claims. I am not directly criticizing any religious views. Many people have taken me to be hostile to religion because of the lectures. But my aim was not at all to argue against religion!

Part of your book is devoted to the question what it means for us to survive. To answer that question, you delve into the problem of personal identity. You write that, for much of your philosophical career, you have found yourself torn between two views: the personality view, according to which someone survives if their personality remains intact (127–139), and the body view, according to which someone survives if their body remains intact (118–127). For some time now, however, you have been “inclined to think that the key to personal identity is having the same body, as long as there’s no branching, as long as there is no splitting” (162).³ Why is that?

³ Kagan imposes the no-branching requirement to prevent fission cases—in which a being splits into two continuants. To illustrate the need for a no-branching requirement, he gives the example of him having a horrible accident over the weekend, which leaves his torso destroyed, but his brain intact (152–153; see also Parfit 1984, 254–255). There are two other men, Smith and Jones, who have also had an accident of sorts over the weekend, which has liquidified their brains, but left their bodies intact.

Assume, against our best medical science, that it is possible to continue living while an entire hemisphere of your brain has been destroyed. Given this, and because

It sometimes happens in philosophy that one just finds different arguments and considerations more plausible over time. Perhaps this is not the most admirable aspect of the discipline. Early on in my career, I was drawn to the personality view. That probably was under the influence of Derek Parfit, who, as I mentioned, visited Princeton while I was a grad student there.

It is not that there really was an argument that pushed me over to the body view. I started thinking of the continuity between people and other physical objects. You follow the body when you are following a lion around. You follow the body when you are following a tree around. Such considerations slowly moved me towards the body view. I am aware of the objections that have been raised against it. The body view just strikes me as, on balance, the most plausible view.

It is important, though, to be precise on what it means to have the same body. Not every part of the body needs to stay the same in order to still have the same body. When you get sunburned, for instance, and your skin peels off as a result, you still have the same body. When you lose weight, you still have the same body. We do not say that someone has ceased to exist when they have had a sunburn or lost some weight.

Even more drastic changes to the body seem compatible with having the same body. As I mention in the book (123), one of my favorite examples here is from *Star Wars*. In *The Empire Strikes Back* (1980), Darth Vader cuts off the hand of Luke Skywalker, just after he says ‘I am your father’. In the very next scene, Luke has an artificial hand. Darth Vader doesn’t think that his son no longer exists. And rightly so!

The question we should ask when deciding what it is to have the same body, is what, if any, part of Luke’s body Darth Vader should have chopped off in order for us to think that Luke has ceased to exist. My

brain transplantations have a high failure rate, doctors decide to split Kagan’s brain and transplant one half of it into the body of Smith, and the other half into the body of Jones. Miraculously, both brain transplantations are successful. Lefty, the person with the left half of Kagan’s brain, wakes up thinking he is Shelly Kagan. Righty, the person with the right half of Kagan’s brain also wakes up thinking he is Shelly Kagan. Who, if anyone, is Shelly Kagan after the transplantation?

Kagan argues that it would not make sense to claim that either Lefty or Righty alone is now Shelly Kagan. After all, both of them received exactly half of his brain! It would also be unattractive to hold that both of them are Kagan—because that would mean that he can now be at two places at the same time. The most sensible response to this scenario, he argues, is to hold that neither of the two are him, which is achieved by imposing the no-branching requirement.

answer is: the brain—or, at least, a significant part of it, because there is some redundancy in the brain.

Your discussion on personal identity is premised on the idea that there are criteria of identity over time. What do you think of anti-criterialism, the view that there are no such criteria (see, for instance, Merricks 1998)?

I don't often believe in brute facts. Suppose you would ask: 'What makes the table I am slapping now, the same table I slapped last week?' My meta-philosophical view is that there are answers to that question other than: 'It's just a brute fact'. That is not because I am sympathetic to the principle of sufficient reason. Unlike my colleague Michael Della Rocca, I do not think that the principle of sufficient reason holds across the board. That doesn't mean, though, that one should say that something is a brute fact whenever the going gets tough. I don't have a master argument up my sleeve about how you know when something is a brute fact. It just seems to me likely that there is an answer to the question of personal identity. The thought that there couldn't be an answer is probably itself based on meta-philosophical views about what an adequate criterion of personal identity would look like. I might challenge the anticriterialist on that.

At the end of your chapter on the choice between the personality and the body view of personal identity, you point out that the focus on what it takes to survive as a person may actually be misguided. What seems to matter more than surviving, is psychological continuity (162-169). Why do you separate the question which view of personal identity is the right one, from the question what really matters in survival? Aren't we also guided, in thinking about personal identity, by our views on what really matters?

I certainly have sympathy for the thought that you can let normative considerations guide you in the development of your metaphysical views. If a certain metaphysical view has normative implications that are very implausible, I think it is perfectly legitimate to change your metaphysical views. So, I hear where you are coming from if you say that you want to track something like personality in terms of what matters, and use that as an argument for the personality view of personal identity. I am just not as wed to using that particular normative insight as a ground for choosing between views of personal identity. There are

cases, for example, where it just seems to me, intuitively, that I *have* survived, but I don't have what I wanted out of survival.

Throughout your book, you discuss a number of fantastical thought experiments, particularly in the part on personal identity. To think about whether the personality or the body view of personal identity is correct, for instance, you ask the reader to imagine a mad scientist who has kidnapped two people, Linda and Shelly (132-139). The scientist has developed a machine with which he can download people's memories, beliefs, and desires, scrub their brains completely 'clean', and transfer memories, beliefs, and desires to the brain of another person. He uses this machine to transfer the contents of Linda's brain to Shelly's body, and vice versa. This raises the question: Who is Linda and who is Shelly after the transfer?

The method of using fantastical thought experiments is controversial. There is an oft-quoted passage from Kathleen Wilkes that sums up the criticism:

we cannot extract philosophically interesting conclusions from fantastical thought experiments [...] because we have the following choice: either (a) we picture them against the world as we know it, or (b) we picture them against some quite different background. If we choose the first, then we picture them against a background that deems them impossible [...]. If we choose (b), then we have the realm of fantasy, and fantasy is fine to read; but it does not allow for philosophical conclusions to be drawn, because in a world indeterminately different we do not know what we would want to say about anything (1994, 46).

What do you think about this type of critique of using fantastical thought experiments?

As it will hardly surprise you to learn, I am not sympathetic to it. It seems implausible to me that we don't know what we would say in certain fantastical cases. Superman is a mythical being, who obviously doesn't exist. The laws of physics would probably have to be very different for a creature to have the kind of powers that Superman has.

Now, Lex Luthor comes along with some kryptonite to kill Superman. We all have the reaction: 'That's wrong! Lex Luthor is doing something

immoral'. I think it is perfectly ludicrous to say that we do not know what we would say if Superman actually existed—assuming that he is as portrayed: stopping crime, helping people, and promoting justice.

There are certain moral intuitions that are quite robust, even about utterly fantastic cases. I am prepared to be more piecemeal about it. I am prepared to have my confidence shaken about the robustness of certain intuitions. I disagree, however, with the bare remark that we don't know what we should say in fantastical cases.

You argue that immortality would not be all that attractive. The “problem with immortality seems to be one of inevitable boredom. The problem is tedium” (243). Your argument for this claim seems to rely on the premise that any pleasure is ultimately exhaustive: after we have experienced a pleasure a (large) number of times, we grow tired of it. As there is a limited number of possible pleasures in the world, eternal life would, at some point, get eternally boring. Might there not be certain pleasures—such as those of food, friendship, love, or music—that are indefinitely repeatable (see, for instance, Fischer 1994)?

Certainly, if one grew hungry for all eternity, one would be glad to eat some food and have the hunger disappear. But suppose that you had to eat a peanut butter and jelly sandwich for all eternity. That would make the hunger disappear, but I suspect that after a certain period of time, maybe not a very long one, you would grow tired of it.

This response relies on there being one particular type of food. I vary my lunch every once in a while not to grow bored of it. Why wouldn't that be an option?

I don't think there is enough food not to get bored for all eternity. Suppose you go to a restaurant where there are five choices of food. Is that enough not to get tired? I suspect that after a month, or two months, you will say: 'O my God, the same food again!' Now, of course, the world offers us far more choices of cuisine. Perhaps it would take a thousand years, perhaps it would take a hundred thousand years, maybe it would even take a million years. But at some point, I think you would grow tired.

But the number of different foods that we can choose from is not that strictly limited. We can invent new foods as well.

Let me give you an example that I learned from Larry Temkin.⁴ Temkin is a huge art fan. He has travelled to all of the leading art museums in the world. Sometimes he comes across a nice art museum that he has not been to before and finds himself thinking: 'I haven't seen these particular paintings before, but this isn't really something new either'. It seems unlikely that you will find, for all of eternity, a new form of art, or a new form of food, thrilling.

There are, as you say, "billion, billion, billions" (221) of potential children who will never be born because they will never be conceived. You argue that we should not feel bad about these potential children never experiencing the goods of life, because "[s]omething can be bad for you only if you exist at some time or the other" (222, Kagan's emphasis). Does this existence requirement apply symmetrically? Is it also true that something can only be good for you if you exist at some point in time?

I ask this, because if the existence requirement does apply symmetrically, then it seems to allow for 'miserable child' cases. Suppose a prospective mother finds out through a genetic test that, if she would decide to procreate, she would almost certainly give birth to a child who would suffer excruciating pain and die at a young age. The prospective mother therefore decides not to procreate. Would it not be good for the baby to never come into existence, even though this baby does not exist?

I do think that any plausible version of the existence requirement would apply symmetrically. If something can't be bad for a being that never existed, then something can't be good for a being that never existed either. But, as you point out, that seems to have the implication that we do not have an explanation anymore as to why we shouldn't be having the miserable child.

Of course, one could dig one's heels in, as some people have, and say that the explanation for why you should refrain from having the miserable child has to do with the costs it would impose on the rest of society. I don't myself find that a particularly attractive answer. Although I do think it is relevant to bring in the costs to the rest of society, the central bit of the explanation would have to do with the child that would come into being. Strictly speaking, it is not true that it

⁴ Temkin provides this example in his "Is Living Longer Living Better?" (2008, 202-203).

is good *for* the merely potential person to not exist, nonetheless there are reasons, having to do with the welfare that that potential person would have had, that should be taken into account.

One might, of course, then push the asymmetry question and say: ‘Are there also reasons having to do with the welfare of a potential child for why it is *bad* that a potentially happy child will not be born?’ My answer is: ‘Yes, there are’. I embrace a wide person-affecting principle and count the welfare of never actualized individuals. So although, strictly speaking, I endorse the existence requirement, I don’t endorse the implications that it would be natural to think follow from it.

III. Counting animals

In your book How to Count Animals, more or less (forthcoming)⁵ you argue that moral status is hierarchical. Human beings typically have a higher moral status than horses, who, in turn, typically have a higher moral status than mice. It seems that many people find the hierarchy view of moral status intuitively plausible. Why did you feel it needed defending, given its widespread intuitive appeal?

Although many people agree that people commonly have a higher moral status than other animals, they disagree about the relative moral status of the ‘lower’ animals. My aim was to sketch the outlines of a framework with which we can determine the moral status of *all* animals.

Also, there are actually surprisingly few defenders of the hierarchy view in the contemporary animal ethics literature. A very prominent position nowadays is that all animals count in exactly the same way. As Peter Singer puts it, “pain is pain”—it should not matter, in our moral calculus, whether the pain is suffered by a dog or by a person (Singer 2009, 20). I call this view unitarianism.⁶ Unitarians do, of course, recognize that people and other animals are beings with different interests, which can make it morally appropriate to treat them in

⁵ All references in this section are to Kagan (forthcoming), unless otherwise indicated. All citations are accompanied by the chapter and section numbers in which the citation occurs.

⁶ As Kagan points out in the first of his Uehiro lectures (“Consequentialism for Cows”), on which the book is based, the view has nothing to do with Unitarianism, a religious view that denies the Christian view of God being a trinity. He would have preferred to describe the view as ‘egalitarianism’, but wanted to avoid confusion with egalitarian theories of distributive justice.

different ways. The point of unitarianism is that similar interests should be given the same weight, regardless of what sort of being we are talking about.

I should say that I have considerable misgivings about defending a hierarchical view of moral status. I run the risk that people will take me to be developing a defense of the way we currently treat animals. Far from it! The way we treat animals is a moral monstrosity, in my view. I am going to disappoint those who think that my book contains an argument for that conclusion, however. The framework I sketch would need to be worked out in much greater detail to draw any practical implications from it.

On your hierarchical view, moral status is grounded in psychological capacities. The relevant capacities include those necessary to have “deep relationships”, “more sophisticated and advanced knowledge”, “more significant achievements”, and an “ability to act out of moral conviction” (chap. 5.2.). I wondered why you focus exclusively on the psychological capacities of individual animals when determining their moral status. Shouldn’t bees be assigned a higher moral status, for instance, if it turns out that they play an important role in the food chain by pollinating crops?

The importance of an animal for the food chain is instrumental. I think it should enter our moral calculus, but not with regard to how much weight we give to the interests of particular animals. Then again, I do hold the view that the intrinsic value of objects sometimes depends, in part, on their instrumental value.⁷ I am not convinced, though, that importance in the food chain is the type of instrumental consideration that could determine intrinsic value.

Let me try again. You hold that beauty is intrinsically valuable, right?

Yes, I do. I think that if we have two worlds, one with only beautiful things in it, and the other with only ugly things in it, then the beautiful world is better than the ugly world—even if there are no people around in either world to observe the beauty and the ugliness.

I wonder whether we could say that beauty is, in part, constituted by diversity. If so, then we could construct an argument that diversity is

⁷ Kagan defends this view of intrinsic value in his “Rethinking Intrinsic Value” (1998b).

one of the intrinsic factors that determines moral status. Animals who are rare, and hence add more to diversity, would then have a higher moral status.

I do think it possible that contributing to diversity might make a given creature more intrinsically valuable (though with billions of bees, it is unclear how much a given individual bee does that). But even if so, it remains to be investigated whether it also enhances the moral *status* of the given individual.

At various points in the book, you talk about animals deserving well-being. As you mention in a footnote (chap. 4.1.), there does not seem to be any sustained discussion of the topic of animals and desert. How could animals be deserving of well-being?

Many people are attracted to the view that people deserve a certain baseline of well-being simply because they are a person, and that their virtue and vice might raise or lower what they deserve against this baseline. Furthermore, on that view, the fact that you are a person might mean that no matter how vicious you are, there are certain punishments that you couldn't possibly deserve, because they would get you to levels of well-being that you couldn't deserve to be at.

If you accept that how generally deserving you are is a function of both the kind of being you are and of your moral track record, then it seems that the door gets opened for thinking about animals and desert. It could be that simply in virtue of being an animal with a particular moral status, you deserve to have a certain level of well-being. We might or might not think that some animals could be moved off that baseline because of their behavior. I am somewhat sympathetic to the thought that a few of the higher animals might be. But I need more time to figure out what I think exactly on this topic.

A significant part of your book (chap. 7-9) is addressed to the deontologist. Why so, given that you are a consequentialist yourself?

The fact of the matter is that most people are deontologists! I have been teaching a Normative Ethics seminar here at Yale for many years. Every year, I ask my students to raise their hand if they think it is impermissible to kill one person to save five others. Every year, a very large majority of them do.

Many students only start lowering their hand when I increase the number of people that will be saved. These students are *moderate*

deontologists: they think that the right to life has moral weight independent of the goodness of outcomes, but it may be outweighed if enough is at stake. Occasionally, some students in the class keep up their hand even when I reach the whole of humanity. These students are *absolutist deontologists*: they think it is impermissible to harm an innocent person no matter how much is at stake.⁸

So one of the reasons that I spend some time addressing the deontologist in the book is that it is a position that many people endorse intuitively. That is not the only reason, though. It was interesting to me to discover that there are some fairly powerful arguments even within deontology for preferring the hierarchical approach to the unitarian one. An example I give in the book is that of Tom, who has been shipwrecked and ends up on an uninhabited island (chap. 7.2.). He cannot survive by eating the sparse vegetation on the island. Would it be morally permissible for him to occasionally catch and eat some fish in order to survive?

The unitarian absolutist deontologist has to say no. If fish have moral standing,⁹ then the right to life of the fish is every bit as important as the right to life of Tom. He would have to starve himself to death. I think that is an implication even many absolutist deontologists would be unwilling to accept. So if they want to remain absolutist deontologists, they need to abandon unitarianism.

The unitarian moderate deontologist does not fare much better in the Tom case. Suppose that we were to say that it was permissible to kill one person in order to save a thousand others. If that is true, then Tom would not be permitted to kill the fish: killing it would have to save a thousand lives to be permissible! So, again, if they want to remain moderate deontologists, they need to abandon unitarianism.

A final reason for me to discuss deontology is that many consequentialists think that, in ordinary decision-making, you want to guide yourself by principles that have a deontological cast. So it is worth thinking about to what extent these deontological elements, which

⁸ Kagan discusses moderate and absolutist deontology in greater detail in (1998a, chap. 3.1.). He also surveys various other constraints, besides doing harm, that the moderate deontologist might be attracted to—such as constraints against lying, breaking promises, and not fulfilling special obligations in (1998a, chap. 4).

⁹ Kagan (chap. 7.2.) notes that those who deny that fish have moral standing could instead imagine that Tom could catch and eat a wild rabbit.

would be mere heuristics to the consequentialist, should take a hierarchical form.

Could you say a bit more about how abandoning unitarianism might help the absolutist deontologist to avoid the implication that Tom cannot kill a fish in order to survive? Wouldn't there still be an absolute prohibition against killing the fish, as long as it has moral standing?

You are right that abandoning unitarianism on its own will not do the trick for the absolutist deontologist. I think that absolutist deontology extended to all animals with moral standing is very implausible, precisely for that reason. Those who are drawn to absolutist deontology would be better served, in my view, by being an absolutist deontologist about people, and a moderate deontologist about animals. On such a hierarchical deontological view, there are no thresholds on people's right not to be harmed, whereas there are thresholds on other animals' rights not to be harmed.

Suppose that we are moderate deontologists and accept that moral status is hierarchical. How are we to determine when, if ever, it is permissible for Tom to kill a fish in order to survive?

Intuitively, for a moderate deontologist, the size of the threshold—the amount of good that needs to be done in order to justify harming some individual—will depend on the amount of harm being imposed. If we embrace a hierarchical approach to deontology it will also depend on the moral status of the individual being harmed. There are various ways to work all of this out (some more complicated than others), but the result will be one or another view where it is easier to justify harming an animal than it would be to justify harming a person, and easier to justify harming some animals than others. One such view is portrayed graphically in Figure 1. To actually work all of this out would be a complicated undertaking; but in principle, at least, an approach like this could justify Tom's catching and eating the fish he needs to stay alive.

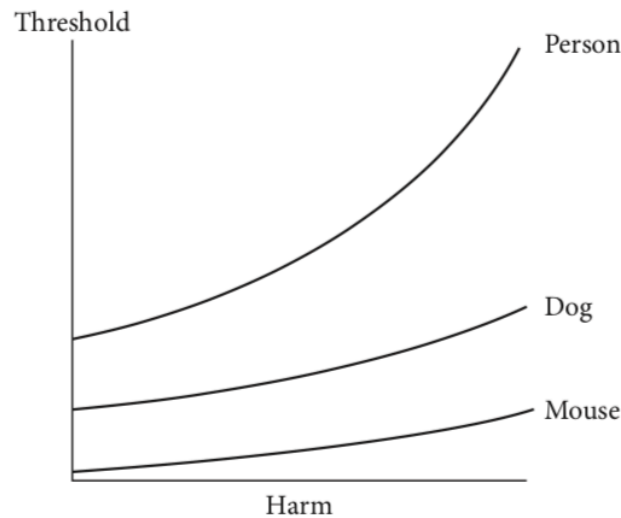


Figure 1: Thresholds with variable slope, exponential growth, and a positive Y intercept¹⁰

Would we be benefitting an animal by enhancing its psychological capacities—provided, of course, that it remains the same animal in the process?

If we could increase an animal's psychological capacities without changing its identity, then I think yes! Imagine that you can choose what type of animal you would like to be reincarnated as. The options are: a mouse, a dog, or a person. What would you choose? I assume most people would choose to be reincarnated as a person, or at least a dog rather than a mouse. That suggests it is better to *be* a person than a dog or mouse, and better to be a dog than a mouse. So if we can turn the dog into a person (or the mouse into something with the cognitive capacities of a dog), that might well be benefitting it. But, of course, it is questionable whether we could significantly enhance the psychological capacities of an animal without changing its brain so much that identity is lost.

An important objection to the hierarchy view of moral status is what you call the problem of 'marginal cases'—the possibility that severely impaired humans have a moral status closer to other animals than to humans (chap. 6.3.). Examples quickly get grotesque here, but might it be permissible, on your view, for an animal with high moral status,

¹⁰ Graph taken from (chap. 9.2.), in which Kagan discusses various other ways to determine thresholds.

such as a chimp, to eat a severely impaired human, if doing so were necessary for the chimp to survive?

That depends on the details of the case. I argue that, in determining moral status, we should not just look at a being's current psychological capacities. We should also look at its potential capacities. When determining the moral status of a human baby, for instance, it matters if the baby could become a person later. But not only that. I think that it also matters what capacities a being *could have had*, had the right conditions obtained. If a human has brain damage because of an accident she suffered when she was a baby, and, as a result, has never been and will never become a person—this human should still be assigned a higher moral status because she *could have been* a person.

So whether it is permissible for a chimp to eat a severely impaired human depends on whether the human still has the potential to develop higher psychological capacities and could have had higher capacities if the right conditions had obtained. If not, then I think it might be permissible for the chimp to eat the human in order to survive. But a lot of further details remain to be worked out to settle the case! (And, of course, this is all taking as given that it even makes sense to talk about moral permissibility when it comes to the actions of chimps.)

IV. Desert

You devoted 20 years to working on a book of 656 pages on moral desert (2012b).¹¹ Why desert?

A number of years after I published my dissertation, I started thinking more about what I wanted to include in my theory of the good. I wondered for a while about whether equality mattered, but I was pretty sure that it didn't. I also wondered whether desert mattered, and decided to explore the concept more fully.¹²

This was in some ways analogous to what I had done in the *Limits of Morality* (1989). What I had found most troubling about consequentialism is how demanding it was. I wanted to explore the common sense view that morality was less demanding and see whether

¹¹ All references in this section are to *The Geometry of Desert* (2012b), unless otherwise indicated.

¹² Kagan (1999) argues that equality's value is derivative of desert's. Olsaretti (2002) and Gordon-Solmon (2015) challenge his argument.

there were any arguments that persuaded me. As it happened, none of them did. What happened, in contrast, in the case of desert, is that I found the instrumentalist considerations for desert less and less persuasive. I spent some time thinking about the kinds of objections you might have against desert. None of these objections struck me as plausible.

Gradually, I found myself thinking: ‘I don’t really have any strong argument for desert being an intrinsically good-making characteristic, other than that it seems to me that it is’. Once I saw that this is just as good an argument as for other things that we are tempted to include in our theories of the good, I thought: ‘Alright, why should I be resisting it anymore?’ In this process of thinking more about the intrinsic value of desert, I realized that moral desert is a much more complex notion than I had previously thought and decided to explore these complexities.

Many philosophers are desert skeptics. Although, as you point out in the first chapter of the book (12), it is explicitly not the purpose of your book to convince the skeptic, I’d like to briefly talk with you about three common objections to desert: the free will objection, the distinctiveness objection, and the viciousness objection. Some of these have come up in responses to your book.¹³

Let’s start with the free will objection, according to which, as Derek Parfit puts it, we do not “have some kind of freedom that could make us responsible for our acts in some desert-implying way” (2011, 409–410). For Parfit this objection is so pressing, that he thinks it justifies leaving desert out of his project of reconciling normative ethics. You are a compatibilist. How do you think determinism and desert can be reconciled?

I am sympathetic to the family of views that sometimes get called reason-sensitivity accounts. To give a familiar analogy: you have freedom of thought if you are able to reason about what to believe—if your beliefs are not sticky, but will change in response to the evidence. I want to say the same thing, broadly speaking, about freedom of action and freedom of the will. If my volitions and my intentions respond, in familiar ways, to what there is reason to want or desire, then to that extent I am free. And if they do not, then to that extent I am not free.

¹³ For responses see, among others, Gordon-Solmon (2017), Hurka (2016), Lippert-Rasmussen (2016), Skow (2014), Smilansky (2013), and Tadros (2017). Kagan (2017) responds to Lippert-Rasmussen’s and Tadros’ critiques.

This kind of reason-sensitivity is, as far as I can see, perfectly compatible with determinism and desert.

Don't you think there is still trouble for reason-responsiveness accounts in the sense that what kind of reasons move you is determined by factors outside of your control?

I am not a fan of the control condition. And I am actually more convinced of the existence of desert than I am about any particular metaphysical thesis about free will and control. If you were to convince me that the most plausible account of free will requires control, and that we don't really have it, I would just say: 'Oh, in that case, desert does not require free will. You can deserve on the basis of things that you had no control over at all'. As it happens, I do think we have a fair amount of free will, but it won't trouble me if you isolate corners where we don't.

Would you also allow for people to differentially deserve things on the basis of factors they have no control over?

I myself think that we have some control over how virtuous and vicious we are, but if you were to convince me that we have no control over this and also convince me that, in virtue of that fact, nothing about our virtue and vice could alter how deserving we are, then there would be nothing left but the well-being baseline we talked about earlier.

You might think this would rob the theory of desert of any of its interest; we should just become egalitarians. But remember our discussion of animals and desert. The baseline component could help explain why it is not a problem that people are so much better off than animals. If you thought that desert can't differentiate any of us, you'd have to say: 'We have this tremendous egalitarian objection to the fact that people are so much better off than animals'. If, on the other hand, you have a more hierarchical desert theory that says: 'Dogs deserve something, but they don't deserve as much as people do', then you have an explanation for why our egalitarianism doesn't drive us to go into the world and focus all of our efforts on improving the lives of animals.

In the book you assume moral desert, but you do not defend it. Do you think moral desert is the only distinctive type of desert? Or do you think there are other distinctive types of desert as well?

We certainly talk about desert in a variety of contexts. We talk about the student who deserves an A because they turned in a really good paper. And even if the student is a horrible individual, morally speaking, they still deserve the A.

That is an example of an institutional desert claim.¹⁴ The student deserves to get an A because she is entitled to it on the basis of the institution of grading.

Indeed, and I think most of the desert claims we make are like this; they are institutional. But I think moral desert claims are actually non-institutional, or pre-institutional. Now, are there any other types of non-institutional desert besides moral desert? I am not clear whether the answer to that is yes or no. There is this example by Owen McLeod that the Grand Canyon deserves protection (2013). I see the force of saying that, and it certainly doesn't seem to be an institutional claim. So I am not at all wed to the thought that the only natural, non-institutional desert claims are moral ones. But, at the same time, it wouldn't trouble me either if that were the case.

This brings us to a second prominent objection to desert, which is that even moral desert is not distinctive. Some critics have raised an objection in this spirit against the factory accident case (23–24) you use to set-up your inquiry into desert (Lippert-Rasmussen 2016; Tadros 2017—Kagan 2017 responds). Could you describe the factory accident case and explain why you think desert does a good job of explaining our intuitions about it?

The case is this: there has been an explosion in a factory. Two workers, Amos and Boris, have been harmed. They have been harmed by the same amount and are now at the same, lower level of well-being. Unfortunately, you can only help one of them—for instance, because you only have a single dose of painkillers. Who should you help? Absent any additional information, it seems that the best response is that you should just flip a coin.

Suppose we learn that it was Boris' fault that the explosion took place. That changes things. If Boris is to blame for the accident and Amos is innocent, then it seems that they no longer have equal claims

¹⁴ Olsaretti (2003, chap. 1) discusses the distinction between institutional, pre-institutional, and pre-justicial desert claims—as well as various other features of desert.

on being helped. It is now better to aid innocent Amos than it is to aid culpable Boris: intuitively, fault forfeits first.¹⁵ I use this example to introduce and explore various elements in an overall theory of desert.

Now, some critics have pointed to other principles that they think are better able to adjudicate the distribution of who should be saved first. Typically, people who say that are not sympathetic to desert. They are taking the stuff that they like and say: ‘Since I don’t believe in desert, this can’t be a desert principle’. I, as somebody sympathetic to desert, want to say: ‘I don’t know why I shouldn’t classify this as part of my desert theory’.

Sure, if, at the end of the day, we have a unified theory of desert and the example of the explosion turns out not to be a case of desert after all, then I will say: ‘Oh, it turns out this wasn’t one of the best examples for making that point’. But it is worth bearing in mind that my example wasn’t intended to convince somebody that desert was at work in that particular case. Its purpose was, rather, to illustrate the different moving parts of the concept.

You adopt a whole life approach to desert in the book, according to which “we look at lives as a whole, to see what one deserves (overall), and whether one has received it (overall)” (11). Doesn’t it matter to you at all that you are getting what you deserve when you deserve it?

To modify an example of Fred Feldman’s for a similar, but not identical purpose, think of the Make-a-Wish Foundation (Feldman 1995, 70). That foundation tries to realize wishes of children who are going to be very ill. Are you troubled by the fact that, when we take them to Disneyland right now, before they become too ill to appreciate the visit, they are getting more than they deserve now? Do you think it would somehow be better if we could take them out of their hospital beds and take them to Disneyland then? I don’t think so! And suppose we know that a person is going to commit a crime tomorrow, but your very last opportunity to punish them is today. I am happy to say that if we really know for sure that the person is going to commit the crime tomorrow, then we should punish them now. Many people find this completely wrong. If so, adopt the local view, but know that it is very difficult to make it coherent. I remain committed to the whole lives view.

¹⁵ As Kagan notes (24), the Fault Forfeits First principle was elaborated by Joel Feinberg (1970).

A third common objection to desert, particularly to moral desert, is that it is too harsh. It seems to require that people suffer—even if such suffering is avoidable. You argue, in your book, that such harshness is not implied by desert (23–27). Why not?

A commitment to desert does not at all commit you to thinking that people can deserve to suffer. Sure, there are retributivists who think that people deserve to have lives that are not worth living. But that is, by no means, the only possible position. It is perfectly consistent to be a moderate, as I call the position in the book (26), and think that everyone always deserves to have a life worth living.

Your book is full of graphs: 203 of them, to be precise. I would like to ask you about them, but to be able to do that, we need to get the reader up to speed with the basics of your graphs, which will make this the longest question ever asked during an EJPE interview. Let's try, in three steps.

The first step is to understand what the axes of your desert graphs represent. The X-axis represents well-being, where negative values indicate lives not worth living, and positive values indicate lives worth living. The Y-axis shows goodness from the point of view of desert, where negative values indicate that a level of well-being is bad from the standpoint of desert and positive values indicate that a level of well-being is good from the point of view of desert. That results in the following basic picture:

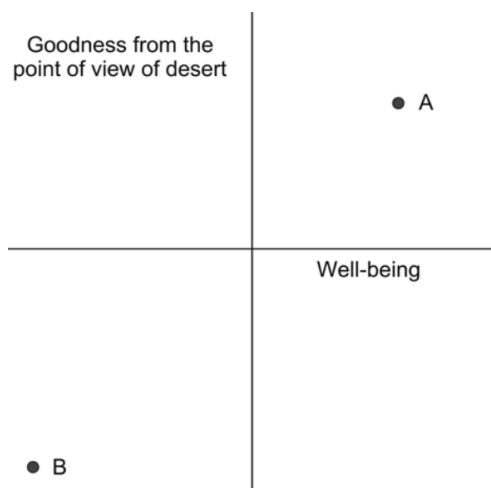


Figure 2: The basic graph¹⁶

¹⁶ Graph taken from (48).

The second step is to actually draw a desert graph. Now suppose we are retributivists and want to draw a graph for a vicious person, who deserves to have life not worth living. This is his desert graph:

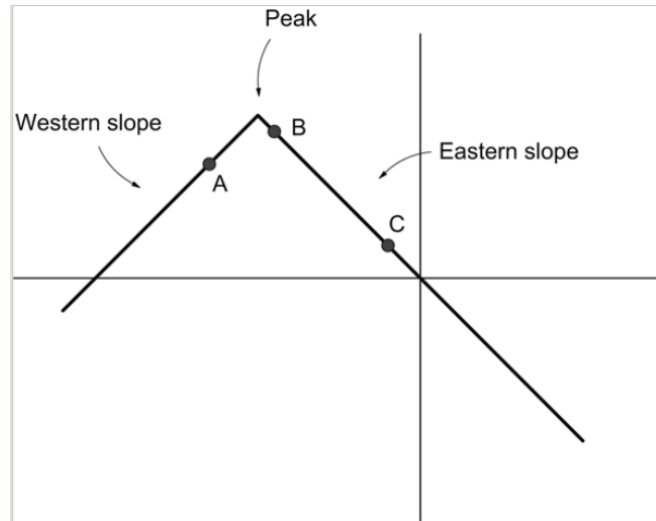


Figure 3: The basic desert graph¹⁷

There are two important underlying assumptions here. First, for every person, there is a particular level of well-being that has the most intrinsic value from the point of view of desert: that is the person's desert peak. Second, it is bad, from the point of view of desert, both if a person's well-being is below her desert peak (indicated by the western slope), and if her well-being is above it (indicated by the eastern slope).

The third step is to add some other persons into the picture. Let's suppose that these other people are all vicious, but differentially so. This means that they all have desert peaks in the upper-left quadrant of the graph, but at different points along the X-axis.

If we add the graphs of three of these differentially vicious people, we get the following:

¹⁷ Graph taken from (75).

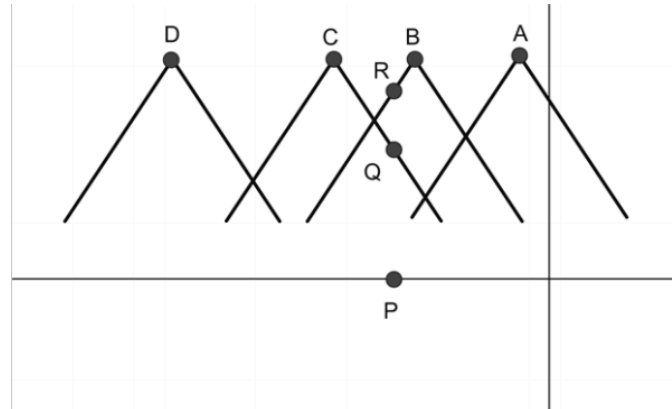


Figure 4: Multiple peaks¹⁸

Here, D deserves to be worse off than A, which is clear from the fact that her desert peak is more to the west than that of A.

Now, we can move on to questions about graphs. You argue in favor of something called ‘bell motion’. What is that?

I think that it is worse if a sinner gets too much well-being than if a saint gets too much. And I also think it is worse if a saint gets too little, than if a sinner gets too little. To capture this thought in the desert graphs, I vary the slopes. The more vicious a person is, the more the western and eastern slope of her desert line will rotate clockwise. Adding this type of slope rotation to Figure 4 above generates what I call bell motion.



Figure 5: Bell motion¹⁹

What do you think about the position that bell motion should go the other way—that it is worse, from the point of view of desert, if the sinner gets too little, than if the saint gets too little?²⁰

¹⁸ Graph taken from (79).

¹⁹ Graph taken from (105).

²⁰ See (98-107) for Kagan’s full defense of bell motion.

That's a crazy view! Remember the factory accident case we talked about earlier. If it is Boris' fault that the accident occurred, then virtually everyone agrees that he loses a claim to being helped first: fault forfeits first. That is exactly the idea that bell motion captures. Reverse bell motion would mean that it would be better if Amos were not helped first, even though it is Boris' fault that the accident occurred. We can reasonably disagree, I think, about the degree of bell motion—but not about the direction of bell motion.

You note that a complete theory of desert should contain comparative elements as well. Why is that?

It does not only matter, from the point of view of desert, whether you are getting what you absolutely deserve. It also matters how you are doing compared to me, in light of how absolutely deserving we are. Suppose that we are equally deserving, but I am getting more well-being than I deserve—I am beyond my desert peak—whereas you are getting exactly the well-being you deserve—you are exactly at your desert peak. Now, if we cannot change my well-being, but we can change yours, should we do it? Non-comparative desert says that we shouldn't. Moving you beyond your desert peak would make things worse from the point of view of non-comparative desert. So if we feel—as many of us do feel—that there is something to be said in favour of moving you beyond your peak, it is a *comparative* value. The theory of comparative desert explores that comparative value.

Your notion of comparative desert piggy-backs on non-comparative desert. There are philosophers who are skeptical about the possibility of non-comparative desert. What, if anything, do you have to offer to the non-comparative desert skeptic?

For me the guiding thought is that comparative desert is satisfied when the offence against non-comparative desert is equal for all individuals—or, in the limit case, when there is no offence against non-comparative desert at all. As you say, on that view, comparative desert piggy-backs on a theory of non-comparative desert. One could reject my own theory of non-comparative desert and still use the guiding thought. If you do give up on the notion of non-comparative desert altogether, however, then much of what I say in the part of the book devoted to comparative desert (part III)—which has many, many pages—would be of reduced interest to you. But not all of it! One of the famous, if not the most

famous, theories of comparative desert is the ratio view, which does not rely on a theory of non-comparative desert. I offer a variety of objections to that view. Those objections still need to be responded to by anybody who uses it. There might be other solely comparative theories that I didn't consider, and then I must see what to think of them.

You spend a whole chapter arguing against the ratio view (chap. 7). What do you think is the most important objection against it?

An important type of case in which the ratio view has problematic implications occurs when one person deserves to suffer and the other person deserves to be well off. According to the ratio view, comparative desert is satisfied when my level of well-being stands to your level of well-being as my level of virtue stands to your level of virtue.

Now, imagine that Amos deserves to suffer. His peak is at -10. Boris, on the other hand, deserves to be well-off. His peak is at +20. The relevant ratio here is -10 to +20. For every negative unit of well-being Amos has, Boris should have two positive units of positive well-being. Suppose that Amos is actually at -5 units of well-being and there is nothing that can be done about that. We can, however, change Boris' level of well-being. How much well-being should he have according to the ratio view? The answer of the ratio view is +10. After all, -5 to +10 is the same ratio as -10 to +20. But that is absurd! It would mean that even though Amos is above his desert peak, we should move Boris below his desert peak.

There are various responses that defenders of the ratio view might offer to this case, but I think that none of them are satisfactory.²¹ And there are a number of other types of cases in which the ratio view has problematic implications.

You end up defending what you call the Y gap view, according to which "comparative desert is satisfied only when the situation of each person is such as to involve a drop along the Y axis of exactly the same size" (395). What do you think of the criticism that the Y gap view has implausible implications when it is combined with bell

²¹ Kagan discusses this case, and various responses the defender of the ratio view might offer to it, in (357-358).

*motion?*²² Imagine that Amos is currently beyond his desert peak of 10 units of well-being, at a level of 20 units of well-being (A's actual location, in Figure 6). We cannot do anything to alter his level of well-being. Boris, who is near saintly, is absolutely deserving of 100 units of well-being, 10 times as much as Amos. The Y gap view, combined with bell motion, could require that we give him as much as 250 units of well-being (B_1)! Don't you think that's unfair?

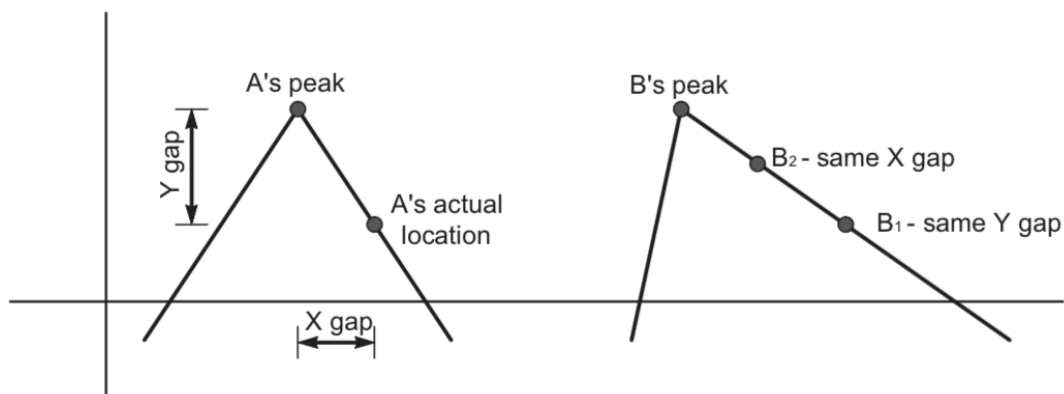


Figure 6: Y gap and bell motion²³

No, I don't! People who are bothered by this type of case are thinking about welfare differences. They have the intuition that it would be fair if Boris would get double the amount of well-being he deserves (B_2), just like Amos. What we should be concerned with, however, is whether Amos's and Boris's levels of well-being are equally offensive from the perspective of noncomparative desert. Well-being differences—that is, differences measured along the x-axis—*do not* capture this. Value from the standpoint of noncomparative desert—differences measured along the y-axis—does. I should stress, though, that I do not settle on a precise account of bell motion in the book. So it may be that the numbers you provide in the example are too extreme; the slopes of Amos's and Boris's desert lines may vary less than you assume.

²² Gordon-Solmon (2017) and Hurka (2016) raise criticisms along these lines.

²³ Graph taken from (407).

V. Convergence in philosophy

We already discussed that, over the course of your career, you changed your mind about which view of personal identity is correct. Another topic on which you changed your mind is kantianism. What happened?

It was clear to me early on that I am a consequentialist. Ever since I realized that, I thought I needed to get a better grip on the opposite view, on deontology. I needed to study Kant and kantianism. Like many people, I found Kant very obscure when I first read him. The opportunity to study his work more closely presented itself when I decided to teach a seminar on Kant's ethics, in particular the *Groundwork for the Metaphysics of Morals*. I set myself the goal of studying it paragraph by paragraph.

I had always assumed, and I think this was the standard view at the time, that if you rejected deontology, then you also have to reject Kant's view on the foundations of ethics. Much to my surprise, however, Kant's view on the foundations of ethics, or at least my reconstruction of it, not only struck me as interesting, it struck me as right. That raised the question: Is it possible to agree with Kant about the foundations of ethics, and still reject deontology?

I came to think that it is. I agree with Kant that his views on the foundations of ethics lead to the categorical imperative, but I disagree with him that the categorical imperative leads to deontology. That is where I get off the bus. I ended up becoming a kantian consequentialist.²⁴

My change of views on this point was surprising to some people at the time. I remember a lunch with Tom Nagel and Derek Parfit, when Parfit was visiting NYU. They asked me what I was working on, and I said: 'I am teaching a seminar on Kant's ethics, and I think that there really is something to it'. Derek turned to Tom and asked: 'Is this our Shelly?'

I should stress that I was not the only one thinking about the compatibility of kantianism and consequentialism then. Richard Hare had written an interesting paper titled "Could Kant Have Been a

²⁴ A lower case 'k' is used here, because Kagan is primarily interested in the kantian approach to ethics, rather than in Kant exegesis. For a more elaborate discussion of Kagan's views on the compatibility between kantianism and consequentialism, see his "Kantianism for Consequentialists" (2002).

Utilitarian?” (1993)—although I think that his reading of Kant was somewhat superficial. David Gauthier had written quite a nice book, titled *Kantian Consequentialism* (1996). But kantian consequentialism was really a small minority view at the time. Nowadays, in no small part because of Derek Parfit’s *On What Matters* (2011), it has become a view that people at least have some passing familiarity with.

Do you think there is a possibility of convergence on moral views, as Parfit argues in On What Matters (2011, 2017), or are you more pessimistic about this than he was?

I am agnostic on the question of convergence. I think we don’t have the slightest idea right now whether, at the end of research time, there will be convergence about philosophical views. As we have discussed, I certainly have changed my mind a number of times—even about philosophical views that I had for decades.

One important question we need to think more about is how much moral disagreement there actually is. It seems to me that a lot of moral disagreement is derivative. We have a number of fundamental moral beliefs, we have some empirical beliefs, and those combine into divergent derivative moral beliefs. Although it may seem that there is a lot of moral disagreement if we focus on the derivative level, I think this disagreement may be due, to a significant degree, to our disagreement about empirical beliefs.

Now, when it comes to fundamental moral claims, there clearly are two possibilities: we converge or we don’t converge. If we won’t converge, then why not? Could it be that we are just too stupid? I think that certainly is a possibility. It often takes a much lower level of intelligence to formulate a question and understand it, then it takes to formulate and understand the answer. Any four-year-old can ask: ‘Why do objects fall when you let go of them?’ It takes a Newton to answer that question. It is not just that the four-year-old hasn’t thought of Newtonian physics; a four-year-old cannot even grasp it.

Maybe we are the equivalent of four-year-olds when it comes to certain philosophical questions. We are smart enough to ask questions about, for instance, free will and determinism—but not to formulate plausible answers to them. Even if Martians landed and explained the answers to us, maybe we would just not be able to understand them.

Are there also reasons to think that we are converging?

Well, we do deepen our understanding of views. I think that is a form of progress and, possibly, convergence. A view gets stated, somebody raises an objection, there is a period of time in which this objection seem really damaging, and then, 10 or 20 years later, a clever philosopher raises a good response, after which the view has a revival. This happens all the time. I am also sympathetic to Bertrand Russell's view on progress in philosophy.²⁵ He points out that if you start with the beginnings of philosophy, then almost all disciplines were part of it. We didn't know how to pursue questions in them. When we did discover how to make progress, these things got spun off and became disciplines in their own right. That's what happened with physics. That's what happened with chemistry. That's what happened with psychology. That's what's still happening with cognitive science and linguistics. So, Russell's answer is: Philosophy is just our name for the set of questions that we have not yet learned how to make progress on. If we ever make progress, we simply stop calling it philosophy.

Parfit apparently couldn't believe that you were not converging with him when it came to desert. Amia Srinivasan describes, in a blog commemorating Parfit (2017), that she visited Oxford while she was an undergraduate at Yale. When she met Parfit during the visit, he asked her: 'Does Shelly Kagan still believe in moral retribution?'

There is a passage in Sidgwick where he says, in trying to decide which of your intuitions you should trust, you've got to look for intuitions for which there is a kind of consensus (1874, 338). Because if not, if the other person is just as smart as you are, and has reflected on the same arguments as you have, then why would you think that there is something magical about your ability to think things through? Sidgwick was a kind of conciliationist in terms of the modern discussion of the epistemology of disagreement. Derek was the only philosopher whom I have ever known who actually internalized and practiced this. That's why he would do all of these revisions of his work. That's why it would bother him when people that he respected didn't share his views and he couldn't persuade them. There are very moving passages in *On What Matters* about Bernard Williams, in which Parfit describes how much he

²⁵ Kagan is discussing Russell's *The Problems of Philosophy* (1912) here.

regrets that Williams died before he had the chance to get him to change his views.

VI. Advice for graduate students in moral philosophy

If you had to name three philosophical works that any grad student in moral philosophy should read, which would those be?

I would always recommend Mill's *Utilitarianism*. Although it does not tell the whole story about consequentialism, it is still the easiest, most accessible version of a consequentialist theory. I would also recommend Kant's *Groundwork*. I am, as we just discussed, sympathetic to a view that takes kantian foundations, but squeezes consequentialism out of them. I think the truth lies in a mixture—some, no doubt, would say an incoherent mixture—of kantianism and consequentialism. Going on to more recent work, I am a big fan of Nagel's *The Possibility of Altruism* (1970). Even though I think some of the details of the argument don't go right, the fundamental insight that the immoralist is making a metaphysical mistake in the practical realm strikes me as exactly on target.

That already gets us to three recommendations, but let me just mention two more. I think, as will not surprise you, that Parfit's *Reasons and Persons* (1984) is just a wonderful work of moral philosophy. It is incredibly ingenious and stimulating—raising all sorts of questions that hadn't been asked before and making new moves in questions that had already been asked. I'd also like to put Nozick's *Anarchy, State, and Utopia* (1974) on my list of recommendations. He is wrong, but brilliantly so. A lot of moral philosophy is stodgily written. Nozick's work is just filled with these challenging, exciting, fun examples. Trying to figure out how you are going to get off the bus is a great exercise; not just for a moral philosopher, but really anybody who wants better training in philosophy—especially if you don't agree with him.

What further advice would you give to graduate students aiming to pursue an academic career in moral philosophy?

I would congratulate them on choosing the right area of philosophy! It is not only the most important area of philosophy, it is also the area where there are the most jobs. Teaching in moral philosophy is more fun as well. Nobody comes into a class on philosophy of language already

having a view about whether natural kind terms are rigid designators. In ethics, everybody comes into the class having intuitions about, for instance, whether life in the experience machine is worth living,²⁶ and whether we should chop up one innocent person to save five others.²⁷ People also have a much easier time seeing why ethics matters—so you are going to have an easier time getting students.

You are also known for your writing advice.²⁸

Yes, the way many students start thinking about a philosophical problem is by reading much of what there is to read on a subject. I think that is a bad way to start! What I learned from my advisor, Tom Nagel, is to start by thinking about the the problem on your own, and see what aspects of it really get a grip on you. Try to find out what the promising views are and what their difficulties are. Only go off and see what other people have said on the subject *after* you have done that. If you read other people first, you will get sucked into thinking in terms of the categories, the distinctions, the claims, and the favorite positions that the literature has already produced. The sad fact of the matter is that most published philosophy is not great. So you might get sucked into thinking about your subject with ideas that do not provide the most illuminating ways for thinking about it.

What are you most grateful for in your career as a philosopher?

That I have this career! Most humans, for most of human history, did not have the luxury to think about philosophical questions in a systematic and sustained way. Even most humans alive today do not have that luxury. I have a job that allows me to sit around and think about philosophical questions. Not only that—I teach at Yale, where I work with some of the smartest people on the planet: undergraduates, graduate students, and colleagues. My job gives me the gift of being able to think about philosophical questions with these people. I am forever grateful that I have been given this chance.

²⁶ Nozick (1974, 42–45) puts forth the experience machine example as a challenge to mental state theories of well-being.

²⁷ Kagan discusses the organ transplant case as a challenge to consequentialism in his *Normative Ethics* (1998a, 70–74).

²⁸ For more advice from Kagan on academic writing, see his “How To Write a Philosophy Paper” (2007).

REFERENCES

- Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy* 33 (124): 1-19.
- Aristotle. 1999. *Nicomachean Ethics*. Translated by Terence Irwin. 2nd ed. Indianapolis: Hackett Publishing Company.
- Brennan, Samantha. 1995. "Thresholds for Rights." *The Southern Journal of Philosophy* 33 (2): 143-168.
- Cummiskey, David. 1996. *Kantian Consequentialism*. New York: Oxford University Press.
- Feinberg, Joel. 1970. "Sua Culpa." In *Doing and Deserving*, 187-221. Princeton: Princeton University Press.
- Feldman, Fred. 1995. "Desert: Reconsideration of Some Received Wisdom." *Mind* 104 (413): 63-77.
- Fischer, John M. 1994. "Why Immortality is Not So Bad." *International Journal of Philosophical Studies* 2 (2): 257-270.
- Gauthier, David. 1987. *Morals by Agreement*. Oxford: Clarendon Press.
- Gordon-Solmon, Kerah. 2015. "Can Comparative Desert do Without Equality?" *Philosophical Papers* 44 (2): 189-205.
- Gordon-Solmon, Kerah. 2017. "Comparative Desert vs. Fairness." *Law and Philosophy* 36 (4): 367-387.
- Hare, R. M. 1993. "Could Kant Have Been a Utilitarian?" *Utilitas* 5 (1): 1-16.
- Hurka, Thomas. 2016. *The Geometry of Desert*, by Shelly Kagan. *Mind* 125 (498): 598-604.
- Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Oxford University Press.
- Kagan, Shelly. 1998a. *Normative Ethics*. Boulder: Westview Press.
- Kagan, Shelly. 1998b. "Rethinking Intrinsic Value." *The Journal of Ethics* 2 (4): 277-297.
- Kagan, Shelly. 1999. "Equality and Desert." In *What Do We Deserve? A Reader on Justice and Desert*, edited by Louis P. Pojman and Owen McLeod, 298-314. New York: Oxford University Press.
- Kagan, Shelly. 2002. "Kantianism for Consequentialists." In *Immanuel Kant, Groundwork for the Metaphysics of Morals*, edited by Allen W. Wood, 111-156. New Haven: Yale University Press.
- Kagan, Shelly. 2007. "How To Write a Philosophy Paper." Accessed November 18, 2018. <https://oyc.yale.edu/sites/default/files/philosophy-paper.pdf>.
- Kagan, Shelly. 2012a. *Death*. New Haven: Yale University Press.
- Kagan, Shelly. 2012b. *The Geometry of Desert*. New York: Oxford University Press.
- Kagan, Shelly. 2017. "Exploring Moral Desert." *Criminal Law and Philosophy* 11 (2): 407-426.
- Kagan, Shelly. Forthcoming. *How to Count Animals, more or less*. Oxford: Oxford University Press.
- Kierkegaard, Søren. 1983. *Fear and Trembling/Repetition*. Edited and translated by Howard V. Hong and Edna H. Hong. Princeton: Princeton University Press.
- Lippert-Rasmussen, Kasper. 2016. "Desert, Bell Motion, and Fairness." *Criminal Law and Philosophy* 10 (3): 639-655.
- McLeod, Owen. 2013. "Desert." In *The Stanford Encyclopedia of Philosophy*, Winter 2013 edition, edited by Edward N. Zalta. Accessed November 18, 2018. <https://plato.stanford.edu/archives/win2013/entries/desert/>.
- Merricks, Trenton. 1998. "There Are No Criteria of Identity Over Time." *Noûs* 32 (1): 106-124.
- Mill, John Stuart. 1979. *Utilitarianism*. Indianapolis: Hackett Publishing Company.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Clarendon Press.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Olsaretti, Serena. 2002. "Unmasking Equality? Kagan on Equality and Desert." *Utilitas* 14 (3): 387-400.
- Olsaretti, Serena. 2003. *Desert and Justice*. Oxford: Oxford University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

- Parfit, Derek. 2011. *On What Matters*, vol. 1–2. Oxford: Oxford University Press.
- Parfit, Derek. 2017. *On What Matters*, vol. 3. Oxford: Oxford University Press.
- Russell, Bertrand. 1912. *The Problems of Philosophy*. London: Williams and Norgate.
- Russell, Bertrand. 1955. *The Art of Philosophizing and Other Essays*. Lanham: Rowman & Littlefield Publishers.
- Sidgwick, Henry. 2011 [1874]. *The Methods of Ethics*. 7th ed. London: Macmillan.
- Singer, Peter. 2009 [1975]. *Animal Liberation: The Definitive Classic of the Animal Movement*. Updated ed. New York: HarperCollins.
- Skow, Bradford. 2014. Reviewed work: *The Geometry of Desert* by Kagan, Shelly. *Ethics* 124 (2): 417–426.
- Smilansky, Saul. 2013. Review of *The Geometry of Desert* by Shelly Kagan. *Notre Dame Philosophical Reviews*. Accessed November 18, 2018. <https://ndpr.nd.edu/news/the-geometry-of-desert/>.
- Srinivasan, Amia. 2017. “Remembering Derek Parfit.” *London Review of Books*. Accessed November 18, 2018. <https://www.lrb.co.uk/blog/2017/01/06/amia-srinivasan/remembering-derek-parfit/>.
- Tadros, Victor. 2017. “Moving Mountains: Variations on a Theme by Shelly Kagan.” *Criminal Law and Philosophy* 11 (2): 393–405.
- Temkin, Larry S. 2008. “Is Living Longer Living Better?” *Journal of Applied Philosophy* 25 (3): 193–210.
- Wilkes, Kathleen V. 1994. *Real People: Personal Identity Without Thought Experiments*. Oxford: Clarendon Press.

Shelly Kagan’s website:

<<http://campuspress.yale.edu/shellykagan/>>

Kagan’s Open Yale course on death:

<<https://oyc.yale.edu/death/phil-176>>

Kagan’s Uehiro lectures on counting animals:

<<https://www.practicaethics.ox.ac.uk/uehiro-lectures-2016>>