

## ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Perspectives on the Self***Who am I? Beyond “I think, therefore I am”**

Alex Voorhoeve, Elie During, David Jopling, Timothy Wilson, and Frances Kamm

Can we ever truly answer the question, “Who am I?” Moderated by Alex Voorhoeve (London School of Economics), neuro-philosopher Elie During (University of Paris, Ouest Nanterre), cognitive scientist David Jopling (York University, Canada), social psychologist Timothy Wilson (University of Virginia), and ethicist Frances Kamm (Harvard University) examine the difficulty of achieving genuine self-knowledge and how the pursuit of self-knowledge plays a role in shaping the self.

**Voorhoeve:** Many thanks to both the Nour Foundation and the New York Academy of Sciences for organizing this event. In a moment, I’ll introduce the speakers to you.

But first, I’d like to reflect on today’s theme, “Beyond ‘I think, therefore I am.’” Of course, it comes from René Descartes. Descartes has become a bit of a whipping boy in discussions of the self, for two reasons. The first is his claim that the mind and the body are two completely distinct substances, which of course leads us to wonder how the two could ever interact. And Descartes was frank about the fact that he hadn’t really figured that one out. We won’t belabor this problem with his substance dualism tonight.

The second error attributed to Descartes, which is closer to the heart of what we’re interested in today, is mentioned by Tim Wilson at the beginning of his book, *Strangers to Ourselves*. The purported error is that our thinking and our sensing are always *transparent* to us. That is to say: we are conscious of what we think and sense. We know what’s going on in ourselves, insofar as we’re thinking. And you might think that seems pretty plausible. In a moment, we’ll get to the bottom of why Tim thinks this is nonetheless mistaken. But first, I want to stress that it doesn’t follow that Descartes thought that we were *entirely* transparent to ourselves. Indeed, he believed some of our motives were unconscious.

We know this, because Queen Christina of Sweden wrote to Descartes to ask what “causes. . .often incite us to love one person rather than another before we know their merit.” Descartes replied that when we experience a strong sensation, this causes the brain to crease like a piece of paper. And when the stimulus stops, the brain uncreases, but it stays ready to be creased again in the same way. And when a similar stimulus is presented, then we get the same response, because the brain is ready to crease again.

And what did he mean by all this? Well, he gave an example. He said that all his life he had had a fetish for cross-eyed women. Whenever he came across a cross-eyed woman, desire would enflame him. And he figured out, he said, after introspection, that this was because his brain had been strongly creased by his first childhood love, who was cross-eyed. Now what’s interesting there is that he says there was this unconscious desire moving him quite strongly, of which he became

conscious through introspection. And he adds that once he became conscious of it, he stopped immediately falling in love with cross-eyed women.<sup>a</sup>

There are three interesting things in this example, I think. First, we have unconscious desires. To some extent, according to Descartes, we are strangers to ourselves. Second, we can, at least in this case, through introspection become aware of these desires. Third, by bringing them to awareness we can extinguish these desires, insofar as we don't want to be moved by them. The example is a story of success in that respect. Now, tonight, we're going to look at challenges to the latter two ideas: that we can come to know our unconscious desires through introspection and can eliminate them when they are undesirable. In addition, we'll look into the aforementioned question of whether our thoughts, at least, are transparent to ourselves because we are conscious of everything we *think*.

And some of these challenges involve the following questions. First, what is the nature of the unconscious? Is it like the Freudian unconscious that we've all heard about, or is it something different? Second, how can we come to know ourselves, our hidden motives, our hidden thoughts? Is introspection, the thing that Descartes engages in this snippet of a conversation with Queen Christina, a good, reliable way? (Many of our panelists have challenged introspection.) And the third question is, if we try to gain this knowledge of our unconscious self, what is it good for? In Descartes' case it was useful, but it might not be so good in every case. Maybe, as Jack Nicholson's character says at the end of *A Few Good Men*, when he is pressed by a lawyer to confess the whole truth: "You can't handle the truth!" After all, we all can think of aspects of ourselves that we'd rather not have brought to our consciousness.

Let me briefly introduce the people we have here to discuss this. First, Timothy Wilson, professor of psychology at the University of Virginia, author of the book *Strangers to Ourselves*, which focuses on the three themes I just outlined. This is a fascinating book, with the subtitle *Discovering the Adaptive Unconscious*. I'll be asking him some questions about that.

Then we have a philosopher who's deeply interested in psychology, David Jopling, professor of philosophy at York University in Canada. He has written a beautiful book, *Self-knowledge and the Self*, in which he explores, among other things, how we come to know ourselves. He argues that introspection and the therapeutic sessions you've all been paying so much for are of pretty dubious value, and that instead, to know ourselves, we should engage in a different form of dialogue. So I'll be asking him some questions about that.

And then there's Elie During, associate professor at the University of Paris-Ouest Nanterre. In one word, Elie is an omnivore in philosophy, because he's written on questions in the philosophy of science, on questions of self-knowledge, and on ethics. One of the key things I want to question

---

<sup>a</sup>Here is the relevant passage:

. . . The objects that touch our senses move, by the intermediary of the nerves, certain parts of our brain, and make there, as it were, certain creases that undo themselves when the object ceases to act; however, the part where they were made remains afterward disposed to being creased again in the same manner by another object resembling the first in some aspect. . . . For example, when I was a child, I loved a girl of my age who had a slight squint. Because of that, the impression made upon my brain by the sight of her crossed eyes so closely joined itself to that made there to rouse in me the passion of love, that for a long time afterward, in seeing cross-eyed persons, I felt myself more inclined to love them than others, only because they had this defect. Nevertheless, I did not know that was the reason. On the contrary, since I have reflected upon it, and recognized it was a defect, I have no longer been affected by it.

From *Descartes: His Moral Philosophy and Psychology*, translated by John J. Blom, New York University Press, 1978. Quoted on <http://sites.google.com/site/psychiatryfootnotes/case-histories-from-the-history-of-psychiatry/descartes-self-analysis>

him on tonight is his claim that "ethics is a self-transformative process." That's to say, you come to know yourself in order to make something of yourself, to change yourself.

Our fourth panelist is Frances Kamm, professor of philosophy at Harvard and one of the world's leading ethicists. [Jokingly:] I would *almost* say I could recommend—but you have to have the stomach for it!—her two-volume *Morality, Mortality*, and following that, *Intricate Ethics*, and, most recently, *Ethics for Enemies: Terror, Torture, and War*. Frances is known for a distinctive method of doing ethics, which is, if I could sum it up rather roughly, introspective. She imagines herself in a particular situation, a moral case, and asks herself, "How would I respond in this case?" and then tries to uncover the reasons why she would respond to it in that way. Now, that's an interesting method, but it seems that some of the other speakers here tonight challenge it in their work—which I hope will make for interesting debate.

But let us return to that second purported error of Descartes I mentioned at the outset. Tim, maybe you could start us off by helping us to understand what unconscious *thought* is?

**Wilson:** Well, as with most things, it depends how we define it. If we mean by thought the self-talk that we're all used to doing and engaging in, then almost by definition that is conscious. But research psychologists over the past few decades have expanded their views of the extent to which our minds operate unconsciously. The kinds of cognitive operations we can perform outside of awareness, such as learning new material, detecting patterns, filtering information as it comes in, is vast. Even such higher-order processes as setting goals for ourselves and combining information to reach decisions can occur unconsciously. There is gathering evidence that all of these things, which we used to think were performed only by our conscious minds, can occur unconsciously as well.

**Voorhoeve:** And you conceptualize this unconscious as the *adaptive unconscious*. What does that mean?

**Wilson:** Well, partly it's to contrast with the Freudian unconscious. I'm not denying that there may well be a Freudian unconscious, a repository of instinctual urges that we do our best to keep hidden in the basement of our minds. The point I want to make is that there may be a much different kind of unconscious that probably evolved earlier in our development as a species than consciousness did; an example of what I have in mind is the ability to transform information and to think in ways that further our survival—hence the term *adaptive*. I think this kind of quick sizing up of the world, interpreting information, and deciding how to act can happen very quickly and outside of conscious view; it is not something that's buried because we don't want to know it. It's part of the architecture of the brain that is unknowable.

**Voorhoeve:** So, unknowable. That seems to say that what Descartes was engaged in, trying to figure out the reason why he had this fetish for cross-eyed girls, is an impossible exercise.

**Wilson:** There are various metaphors for introspection and I think the Descartes example is an interesting one. One metaphor for introspection is that it's like a flashlight that we're shining in the dark; and we flash it in the corner and discover that crease in the brain that we just hadn't bothered to see before.

Another metaphor that Freud was fond of is *introspection as archeology*. We're digging up things that were buried deep down in our minds. If we dig deeply enough we'll find them.

The metaphor I prefer is introspection as *narrative building*. We're constructing stories about ourselves based on some access to desires, but not as much. This process is similar to how we would construct a narrative about somebody else, such as observing what they do and bringing our vast cultural knowledge to bear. I'm not so sure that isn't what Descartes did. He deduced almost as another person might have: "Ah, this is the tenth cross-eyed girl I'm attracted to and I

remember that first one, so maybe there is a pattern here. . ." So it may not have been by shining a flashlight or digging deeply, rather he may have simply inferred his liking for cross-eyed women based on self-observation.

**Voorhoeve:** So you're saying we figure out things about our unconscious by imagining how someone else would look at us?

**Wilson:** In part. I think that's actually not a bad way to try to do it, to see ourselves through the eyes of others. That's a difficult thing to do, to get outside of our own heads. But I think that is one interesting path to self-knowledge.

**Voorhoeve:** One of the things that struck me when you said that what we're trying to do is create a narrative about ourselves is that a narrative brings to mind the thought that we're just storytelling. And you mention some cases to illustrate this narrative capacity. For example, psychologists hypnotize people and then get them to do absolutely crazy things—I think, in one case, to put a lamp shade on the head of someone else who is in the room and kneel in front of this person. (Quite interesting, by the way, why one would choose that to suggest to one's experimental subject!) But when the subjects in this situation were asked, "Why did you just put the lamp shade on the head of the other guy in the room?" they would come up with a story, such as, "Oh, I just thought we should lighten the mood here a little bit." But that's *confabulation*.

**Wilson:** Yes. There are very striking examples of brain-damaged patients and others who can't know exactly why they did what they just did but are very fluent in inventing reasons that they seem to truly believe are the reasons they did it. It seems to be very important to us to have a narrative of why we are doing what we are doing. And we're very skilled at constructing such narratives.

**Voorhoeve:** David, I thought I might bring you in on this point. You claim that this construction of an unreliable narrative is what happens in a lot of therapeutic sessions, right? But you don't seem to be so sanguine about the idea that such a narrative will do.

**Jopling:** No. I have some reservations about the so-called *talking cures* or the exploratory psychotherapies, where the goal is insight or self-knowledge. If you look closely at the interpersonal dynamics in a psychotherapeutic or clinical psychological situation, there's a huge amount of pressure on patients, or clients, to produce the sorts of insights that are expected of them by their therapists or psychiatrists. There is also a very strong expectation for them to comply with the doctrine. It's called *doctrinal compliance*: comply with the doctrine or the theoretical orientation of the therapist. If they don't, they're not good patients. There are a lot of subtle cues and pressures, leading questions, and suggestions to get patients to see things as their therapists want them to see them.

And certain turning points in the talking cures—such as psychodynamic psychotherapy, Freudian analysis, and narrative psychotherapy, for which insight is one of the highest goals—the therapists offer interpretations to the clients about what their problems are, or about what their Freudian unconscious looks like, or about the various pathologies from which they're suffering. And these interpretations are generated from hours and hours of clinical material that have been collected by the therapist. But they're very powerful frameworks for patients to use to think about themselves and their problems. It is expected, then, that the patients develop insights that more or less conform to these and confirm these interpretations. The insights that clients thus acquire are vivid, they are intense, and the clients often have very high levels of conviction about them. But my question is, are they true? I don't know.

**Voorhoeve:** Does it matter?

**Jopling:** Sometimes it does not matter because these insights trigger therapeutic improvement. They generate more clinical material, they can be adaptive, very helpful; clients feel good about them. And yet in my interpretation, some of these insights are placebos. They are the psychological equivalent of a sugar pill. They're missing some vital ingredient; that's the nature of placebos. They're missing active, medicinal ingredients; they're empty pharmacologically. The psychological equivalent is what is missing. And I would venture that in some cases, not all cases, these so-called insights are placebos because they're not truth-tracking. They're not true. They're false but adaptive; they work and they bring about some degree of therapeutic improvement.

**Wilson:** If I could just jump in. I agree with most of what you said except perhaps the idea that truth is the ultimate goal. There are many versions of truth about one's life and many narratives that might make sense to a person. In fact, the data on psychotherapy of which I'm aware show that psychotherapy does work for many problems. What's interesting and not inconsistent with what you're saying is the theoretical discipline doesn't matter much; very different kinds of psychological therapies work, and the key predictor of whether they work is whether the client buys into the belief system of that therapy.

The only place we may differ is whether that means that therapy is somehow false or not helpful or placebo. Instead, perhaps we all crave a narrative that we believe in and can help us make sense of our lives and move beyond our problems. I tend to think the latter.

**Voorhoeve:** Elie, you and I were discussing this before and you said, "if it works, it must be true in some sense," no?

**During:** Well, the intuition is that if it works somehow then it means that you've hit upon something.

But to come back to the general question, the one you raised earlier, which is why should we care at all about knowing ourselves?

My first answer would be, it's not very clear what we mean by *knowing*. The self is one thing, and although we decided not to raise the general question of what is the self, I'd like to raise the question, What do we mean by *knowing*, by knowledge? And from the various examples we just had, it's clear that it can mean a wide variety of different practices, actions, and operations, from locating in your memory one particular event—Descartes and the cross-eyed girl and kind of weirdly relating that to a crease in your brain, or today we would speak of a synaptic path rather than crease—to the therapeutic relationship. And it also could mean the kind of self-modeling process, unconscious mostly, that's going on in any learning activity. These are very different ways of relating to oneself and very different ways of trying to know something about oneself. And I don't see how you could be interested in knowing yourself in general, if self-knowledge covers all these very different activities. I don't see what it means really.

So what I'm interested in, when I grapple with the problem of self-knowledge, is what do I get personally? And my intuition is that if we are interested, in the first place, in knowing ourselves, it's because we want to change something about ourselves—that there's something in ourselves that is not very satisfying—something that must be transformed. In my view, the proper reference frame or context we need to make sense of self-knowledge is *ethics*. That would be my answer.

**Voorhoeve:** I can imagine wanting to know myself not merely in order to change myself, but in order to make a big life decision. To take an imaginary case, I might ask, "Should I stay with my girlfriend, and do I really love her?" That's not necessarily about my wanting to change myself, but just really wanting to know, is this the person for me?

And what struck me in some of Tim's and David's work is that they claim that sitting in a room thinking deeply about whether this is the person for me is probably one of the *worst* ways I can go about it. Why is that?

**Wilson:** Well, I think there's only so much that gazing at our navels can accomplish, and I think being acute observers of our behavior can be very helpful. There is increasing evidence in psychology that our friends often pick up on things about us that we don't see. But I think, as in the Descartes example, we can be very good observers of how we're reacting and what the circumstances are, of what seems to trigger joy and love and what doesn't.

**Kamm:** But surely you ask yourself, besides how do I feel about this, you look at certain objective factors, such as do we enjoy being together, do I feel satisfied after having spoken with this person, do we share the same interests? And these are more objective things than simply how do I feel. Will there be conflicts because I want one thing and this other person wants another thing? Surely. But if you're not looking at those things, you're looking at the wrong navel. I mean, there are lots of holes in the body—right? There are these more objective factors we refer to that will tell us or help us decide whether we're in an appropriate relationship or doing the right thing in our own lives.

**Voorhoeve:** That's true. And the example might move us to some of your cases in ethics. One of the things that Tim claims is that there are a lot of things that I *want*, *think*, and *believe* that I can't articulate. No matter how hard I think about it and try to introspect, I can't figure them out. One way of figuring them out is to place myself imaginatively in a situation and imagine how I would respond. And that's a way of gaining indirect access to my subconscious desires and thoughts.

In my imagined case, the question might be, "Does my heart leap up if I imagine spending the next ten years with this person?" But here's how I thought it linked to some of your cases, Frances. For those of you who aren't yet familiar with much of Frances' work, her methodology involves placing yourself in a given case, seeing what your intuitive response would be when you really imagine yourself in that case, and then trying to figure out, through introspection, what the reasons are that drive this response—I see you shaking your head! Am I mischaracterizing?

**Kamm:** That's right.

**Voorhoeve:** Well, how about you take over then.

**Kamm:** Well, I was reading the writings you helpfully gave me of Tim Wilson's things, and I thought, there's a lot here that I agree with. So, let me tell you the sort of cases that I imagine, that I'm discussing. They're thought experiments. They're like scientific experiments done in a laboratory, where you can hold all factors constant and alter a variable, just one variable at a time, to see the effect. Scientists still do that, I assume, right? However, of course, in thought experiments, you're just imagining.

Philosophers of my type, analytic moral philosophers, spend a lot of time discussing this one case that has been made famous called the trolley problem. There's a trolley headed toward five people that's going to kill them, and a bystander could turn the trolley onto another track away from the five. But unfortunately there's one person on the other track who will get hit and be killed. And the question is, is it permissible to do that, to turn the trolley? So that's an imaginary case. It's not like I'm there, standing there.

There are thousands of variations of such imaginary cases and each one is a little thought experiment. This is what Alex means by saying that I think of myself in a case.

So, the trolley's headed toward the five in another variant: there's someone standing on a bridge over the trolley tracks. If you push that one person in front of the trolley, it will stop the trolley

and save the five people from it; may you do that? Now, many people say no, though I haven't done surveys. And that is another part of the method: you think about something rather than do surveys.

So, I think it's permissible to turn the trolley away from the five people, though it will hit the one person. But I do *not* think it's permissible to push the one person over a bridge in front of the trolley even though, again, one person will die and five people will be saved.

So what's the difference? Why do I have these intuitive judgments? I make a judgment about right and wrong. And this is supposed to be an objectively true judgment; it's the sort of judgment I think you all should agree with. It's not something that's just expressing my own personal point of view any more than Professor Wilson's views that you've come to hear are just a way of reaching into his soul and understanding himself. He claims to be talking about the truth about the adaptive unconscious.

**Voorhoeve:** He's not just giving us a narrative—he's aiming at the truth.

**Kamm:** That's the thing. We're not just interested in ourselves—you know, what am I thinking? We want to know whether we're latching on to the truth about the subject matter we're thinking about. And that's the same in ethics.

Where I agree with Professor Wilson is that I think, first of all, that a lot of people tell themselves stories. He calls it confabulation; I would call it conjectures that are not correct. For example, some psychologists have thought that the only distinction between these two trolley problems I've introduced you to is that in one, I'm up close and personal to the person I push over the bridge, but not to the person that I turn the trolley to. The way you deal with a hypothesis like that is you think, suppose instead of pushing the person over the bridge, I had a machine that I could use to press a button from a distance, and this would push the person over the bridge. Do I think *that* makes it permissible to push him over the bridge? Intuitively, my judgment is no. So when I remove the factor that is presented as the crucial variable between permissible and impermissible, I do not get a change in my judgment about impermissibility, which suggests that this factor is not accounting for my judgment. This is a way of teasing out whether a particular factor or conjecture is correct or incorrect.

Similarly, you could do it by taking this factor that people say makes the action impermissible, construct another case where it is present, and find it doesn't make the action impermissible, which I've done also, which is why I have so many cases.

I think that this is like the method that Tim Wilson says gives access to the adaptive unconscious, which is called *inference*. It doesn't mean that I have privileged access to myself that I couldn't have with someone else. Because I'm constantly trying out and testing hypotheses about why I have made that judgment. And Professor Wilson grants that you can have access to yourself of that sort, inferential. That's one of the reasons I agree with something else he says, namely that I could have as good knowledge about why someone else makes a certain judgment as about why I do. Because when I see all the people who say, "Yes, you may turn the trolley this way but you may not push the man over," if I have explained by this method of inference, considering all different cases, what the crucial factor is, I would understand that they would respond to that factor. That's why I think that there is some degree of agreement here. But it doesn't mean that this adaptive unconscious is completely inaccessible. That isn't what he's claiming.

The interesting other thing that I found was that very often the so-called confabulations that people give, they tend to be quite simple, like this one: well, you're up close and personal, you're pushing the guy over. My explanation of the difference among all these cases, and not only these two, but many others—you've got to get a theory that accounts for all of them—it can be quite complex or at least unexpected. One's response can be something like, "What? I would never have thought of that."

I was very struck in Professor Wilson's book by what he called *implicit learning*. Experimenters flashed certain lights or "x's" in various quadrants of a screen or grid according to a very complicated formula. Implicitly, subjects learned to adapt to that flashing of light, so that eventually they came to predict where the next "x" would be. If you asked them to verbalize this complicated formula, what made them do that, they said other things, simple-minded things that didn't correlate at all with the way they behaved.

The actual truth was complicated, and so when the theory that explains people's intuitions is discovered, if it's complicated or not, that really may be what's going on, that's really what is causing people to behave in certain ways or respond in certain ways. I take that as consistent with something that Professor Wilson had said about implicit learnings.

I want to emphasize that we're not just looking into our navels when we do this as philosophers. When I claim that this is a factor that accounts for why it's permissible to do something in one case and impermissible to do it in another, I'm claiming that that's the *moral truth*. This is a step beyond investigating grounds for intuitions. And that means that it's supposed to be *universalizable*. That is, everybody, no matter how much they differ from me in their likes or dislikes, their upbringing, their interests in life, should agree that this is the way they ought to behave—that it's permissible for them to do this and impermissible to do the other thing. So there is this claim to objectivity and universality about permissibility. And that's not just self-knowledge. And my sense is that's what's *really* important—to find that out.

**During:** It's definitely important, and I find it highly interesting to, as you do, experiment on our intuitions to dig out the invariant structures of moral judgment. In what sense is this self-knowledge at all in the sense of how myself, this particular self that I am, is concerned by this? If it's a universal structure of moral judgment, I would say the real question, the ethical question, is how is it that I'm not always ready to agree with this universal truth that you unearth?

**Kamm:** Agree?

**During:** How is it that I don't always act upon this kind of intuition? What prevents me from seeing the moral truth behind these very sophisticated scenarios? And at this point I think we're touching on an actual problem of self-knowledge. It's why me, as this particular individual that I am, why am I having an issue with these general questions?

**Kamm:** I agree with you. You come back to the question of the self when you see, for example, that some people may disagree about the principle. A member of the Nazi party may say, "There's nothing wrong with burning these people; those people it's impermissible to burn, but *these* people it's right to". And then you would want to see by considering their views whether you could find fault in their reasoning, and who has the correct moral view. But you're right, they may say, "I don't see your point of view." And then you may just say, "Well, that doesn't mean it's not correct; you may have reasoned incorrectly." But you will have people who will say, "I see that it's wrong to take the drugs, but I want them so much, I'm going to give in."

So we have personal qualities, weakness of will or habitual behavior, that can interfere with doing the objectively right thing. And that's the point at which we have to consider what philosopher Harry Frankfurt has considered, the meta-judgmental self. This self may stand and see that you have this desire to ignore the distinction between permissible and impermissible: you want the money, you're going to take it, you know it's wrong. But if there is this meta-judgment, some people will have a part of themselves that says, "I don't want to be this way; I do not endorse my desires; I do not endorse my predilections." And then the question is, can they be put in control? Can that part of the self that judges the other part be put in control? The trouble is that this process may go on to the next level. Because when you're thinking about truth, you may say, "Is

the perspective from which I do or do not endorse my desires, is that the correct perspective?" You've got to check on that first meta-judgment, whether that's true as well.

**Voorhoeve:** The thing that you bring up there, Frances, this question of our endorsing our desires, I think, is of crucial importance. If I were to answer the question you put to us earlier, namely, "Why do we care about figuring out ourselves, including our subconscious selves and desires?," then my answer would be: to gain self-control, to become more autonomous. Because I think one of the most unsettling things is when we don't know why we do what we do, or we don't endorse the desires on which we act. And one of the reasons we turn in on ourselves, or try to figure out by inferential methods what we are like, is to get past the point of lacking this control over ourselves. Either because we don't know what's driving us, or because we *do* know what's driving us but we don't want that to lead us. And Elie, at some point in our correspondence, you quoted Bergson to the effect that one *becomes* a person only with great effort, rather than, as we ordinarily would say, simply becomes a person once one has reached the age of reason.

**During:** Yeah. That's if you're a psychologist. If you're a philosopher, you'd say you are a person when you're a moral agent and you're responsible.

Bergson had this theory about the fact that it's a strenuous effort to be a person, to maintain this continuous attention; attending to oneself is something that costs a lot. But I was thinking also of another thinker whom Richard referred to, Ostad Elahi, who had this very sophisticated notion of the *imperious self*, which echoes what you were just saying. I think the issue of self-knowledge really becomes ethically significant and important when you start differentiating within yourself between different layers or levels of the self. A deep level—I wouldn't say authentic, because I don't know what that means—but a deep level and a superficial level, which is the superficial ego and the one that projects a social image to other, this social self that William James referred to. And as I said, there is a spiritual self or deep self.

And Ostad Elahi had this notion of the spiritual self, or self modeling itself, self modeling its personality, as attention between an imperious self, which is a bundle of drives or urges or desires that basically come down to animal nature, and a higher self, call it the second-order self, which is present not so much to establish control (well, yes, of course, the ultimate goal is to maintain some control over this and transform yourself accordingly) but to get something from the knowledge of this imperious self. It's not about crushing your desires or just getting rid of these very annoying urges. It's about getting something from them and building virtues upon them. We use these very old-fashioned terms ("virtues") that Aristotle reflected upon.

Virtues are basically about transforming your passions and desires into something that's constructive and will make you flourish as a spiritual self. But I think it's very important to have this self-model and this dynamic tension between two levels of ourselves. And the notion of imperious self developed by Ostad Elahi is a case in point. It's a very interesting way of self-modeling without having any perspective of gaining an absolute objective knowledge of yourself, as an embodied person or as an unconscious psyche. These are all elements that can be useful in the ethical process of self transformations, knowing about unconscious drives or aspirations. But the ultimate goal is not to map your personality in every detail. That doesn't make sense. Again, the idea is to come up with an effective model of your ethical self. And this involves, at some point, a tension between two levels, call them first-order and second-order selves, or a superficial self and something that you identify with.

**Voorhoeve:** Tim's been shifting uncomfortably . . . Would you like to come in on this point?

**Wilson:** To go back to something that Frances was saying. I actually resonated a lot with the idea of thought experiments. It's something we psychologists do a lot too. But I think there are somewhat

different goals. We do sit in our offices doing these thought experiments, but then we go and do the actual experiments. Because the goal, for us, is to predict what people actually do. Your goal is, I think, somewhat different: to specify what people ought to do.

And the trolley example was very interesting, because I've always wished I could do a real experiment with that. Obviously I can't for ethical reasons (*participants laughing*). But I'm not sure that any of the analyses of people's intuitions about it would match what they would actually do in that situation, faced with a split-second decision of "I can save five lives or one" and "I can do it this way or that way." I'm not sure that the kind of hypotheses we come up with through thought experiments would do a great job of predicting people's actual behavior.

My friend and colleague Dan Gilbert and I have done a lot of research on what we call affective forecasting, which is people's ability to predict their future emotional reactions to future events. And although we're not terrible at it, we do make systematic errors in knowing exactly how we will feel, for example, whether it really would be horrible to push one person to save five. These are predictions that aren't always correct.

**Kamm:** But suppose you could put someone in this situation, as you say, if it were ethical to do so, and they pushed the person over the bridge. The question, as you said, is not resolved of whether they ought to have done that. They may have done it, and they may have done the wrong thing. They may have not done the right thing in the split second. So as you say, your question is different from mine. Similarly, I know about some of these experiments, but please correct me if I'm wrong, about predicting your future affective state.

Many people say that it would be horrible for them to be paralyzed. They think about the future. They wouldn't want that—very bad. But if you ask people who have actually become paralyzed they say, "I have a fine life." What we have to be careful about is thinking that all that is at issue is your feeling of well-being—and you said, "Would I feel horrible if I pushed the man over the bridge?" That's not the important issue. The question is whether you've done the wrong thing, and that's not always a matter of how you feel. Similarly, the paralyzed person cannot walk. The fact that he adapts in a certain way or tells himself certain stories about how it's not important to walk—you can do lots of other things, you don't need to walk—does not show that it isn't bad not to be able to walk.

The fact that we can perhaps delude ourselves in a certain situation is not an indication that, for example, if people come into the emergency room and some people have severe headaches and another person will have spinal damage, that I should say, "Well, treat the person with the headache because he's really in pain. The other guy, he'll be paralyzed but he won't be in pain, and psychologically he'll adapt very quickly." I think that is absolutely the wrong decision to make, because there are more objective criteria. There's something else besides how you *feel* that is relevant to whether you're living a good life.

**Wilson:** Well, let me give what I think might be a fairer example. Suppose you're the physician who has to decide how much to treat someone who's going to become paralyzed or someone with another severe disorder that's more difficult to adapt to. I think certain digestive problems or urinary problems that have to be dealt with daily are much more difficult to adapt to than other disorders. And I think that's relevant information. And although there may be some self-deception and delusion in a paralyzed person, I think that what we can't predict is how much we will adapt and come to terms with something. I'm not sure I would use the term *delusion* as much as *adaptation*. But I think that's precisely what we find hard to imagine in advance.

**Kamm:** Okay.

**Voorhoeve:** David, it seemed that you had some objections to Elie's earlier remarks.

**Jopling:** No, I just wanted to follow up on something that Elie was saying about this, these different levels of self. Talking about the self this way goes back to William James, who talked about many different kinds of selves, and asked whether there was a self of selves that he could know, and he had a lot of trouble; all he came up with were some movements in his throat.

But this notion of reflective self-evaluation, which is directed to answering the question who am I, what am I like, who do I want to be? This is a theme that Frankfurt picks up. It's a very interesting process: you look at the desires that you actually have and you ask yourself, "Do I want these desires, do I want these desires to motivate me and to constitute my will?" And you can do this with a number of de facto desires and beliefs that you find in yourself. Persons are not simply given static entities, but they have an ability to shape themselves to a certain extent. But I wonder about who was there at the second-order level asking do I want these desires? And I wonder if it gets you into a kind of regress and into a third level and a fourth level and so on.

**Kamm:** Yes, that's what I mentioned before, that the meta-judgment can go on to the next level. You can have an evaluation of the values that you have right now.

**Jopling:** Where does it stop and where is the self that you finally want to stick with and say, "That's me. This is the self I want to be or I am"?

**During:** There's a very simple solution practically: desires never muddle themselves. There are particular instances where you frame yourself and picture yourself as having the kind of personality traits that you want to get rid of at some point. I don't know where that comes from, but that's what I'm interested in. But would that be a problem if this wasn't the monitoring superself that philosophers imagined—if it were different, if it were a bundle of different powers or instances within myself, that would be OK.

The important thing is that we have this tension between two levels.

**Kamm:** One level has authority. It's not just a conflict between different desires. The whole idea is there's an authority that one level has that the other one doesn't. The question is, where does that come from?

**Jopling:** And where does it go? How far up does it go?

**Kamm:** Yes, right, right.

**Voorhoeve:** No, I don't think this necessarily involves an infinite regress. Just take questions like: "Do I love my children?" and "Do I *want* to love my children?" Now, "Do I love them?" That's the first-order question. And the answer is: "Yes." The second-order question is: "Do I want to be the type of person who loves my children?" If I answer "Yes", then that's an affirmation of my first-order desire. Now, David's saying, "Well, you could go on forever, asking questions of an even higher order, such as "Do I want to want to be the type of person who loves my children?" But it seems to me that when I am satisfied with myself, with my constellation of desires, I can realize that no matter how much I would think about it, and how many orders I add to the question, I will always answer in the affirmative.

**Kamm:** But someone like Peter Singer, who thinks we should take a more impersonal point of view on ourselves versus others, will say, "I do now think that it's right to care for my children, but should I care for them more than for the other children in the world, given that, objectively, their status is equal?" And he thinks—I don't want to attribute to this to Peter Singer exactly so let's just call this a Peter Singer-like fantasy person—all right, someone who takes the impersonal point of view will say "You ought *not* to want to be the person who loves your child more than

someone else." Now, that may be the wrong moral conclusion, but at that point you can't just say, "I find myself wanting to be this sort of a person." Instead, you've got to give yourself a *reason* for wanting to be one sort of person rather than another. So you've got to start reasoning about these things.

**Voorhoeve:** I'm going to bring in the audience here, because I think you all deserved a chance to ask questions of the speakers. Here's the first question: "Is knowledge of the self an introspective process, or more of what other people say we are? If so, then should we weigh what people say against what we believe about ourselves?" Do you want to start, Tim?

**Wilson:** Well, as I mentioned before, I think there is increasing evidence that if you ask people close to us to rate our personalities, they have a somewhat different view than we do of ourselves. There's overlap of course. But both of those views are good predictors of our behavior. So it seems that there's some element of truth in how we view ourselves and some element of truth in how others view us. And so I think paying attention to how others think of us is important. But I don't think it's terribly easy. We tend to think others see us the way we see ourselves, for example.

I've often wondered if there should be a new Hallmark holiday called Friendly Feedback Day, (*participants laughing*) when our friends send us cards with helpful hints of things that we don't seem to know about ourselves. There might be some benefit to that. You heard it here first (*participants laughing*).

**Jopling:** Could I just add to that?

**Voorhoeve:** Yes, please.

**Jopling:** I think we can learn a lot from Socrates, who mastered the art of dialogue. Getting people to learn about themselves by coming up against Socrates' way of questioning—he was brusque and arrogant and tough-minded and really put people through the wringer—is one particularly difficult form of self-examination with another person. And I think we've inherited a tradition, going at least back to Descartes, possibly to Augustine, where self-knowledge is thought of as an activity that's pursued by the self, for the self, and, ultimately, with the self's own resources. And other people, their biases, their influences, they're kept at the side, because they are contaminating influences. So the best way to know about yourself is to retreat, or to suspend the influence of society and other people.

I think there are a lot of problems with that. I would say you learn about yourself in dialogue with other people; at least this is one of the ways to learn about yourself. Not every kind of dialogue is conducive to self-knowing, but there are some dialogs where the other person is a presence that evokes honesty or truthfulness from you that you might not otherwise have been amenable to. The Other is possibly a moral witness, to your accounting, to your owning up. You're owning up to an Other and learning about yourself this way. So I think there's a role for dialogue, and not just interjecting what other people say about you, but other people serving as foils and challenges to your explorations.

**Kamm:** I think you might have to be careful, though, especially if you are a member of a group that has been subordinated or for whom there has been a stereotype for certain behavior. That what you may get back from others is the expectation of a continuation of that position. So you should always question. And again, think about the reasons. Ask the other person, "Well, why do you think I should change in that direction? Give me a reason." Because otherwise you open yourself up to the possibility of self-diminishment.

**Voorhoeve:** That's the point about the nature of the ideal partner in dialogue. Socrates wasn't a friend to the people he was investigating, right? Tim was talking about friends, but you don't put your friends to death for being really annoying, which is what the Athenians did to Socrates. Then again, telling your friends what you really think of them might end rather badly (*participants laughing*).

David, if I could ask, how is the type of dialogue you propose different from other types? It's not the dialogue of therapist and patient, according to you; it's not necessarily a friendly dialogue. It's not a dialogue with someone who has preconceptions about you. So what type of dialogue is it?

**Jopling:** Well, I think perhaps there are psychotherapeutic or psychiatric dialogues that strive toward this, but they get trapped up in technique, technology, manipulation, and goals that are foreign to the person. I think some dialogues—probably quite rare—are spontaneous, unpredictable and from which you have no idea where it's going to end up and how you're going to be changed, but you are changed. A blind spot that has been preventing you from seeing yourself—some really basic dimension about yourself—has been removed by the challenge of the presence of the other person.

**Voorhoeve:** So, in part, it's about the dialogue being open ended?

**Jopling:** Yes. Very much an open-ended dialogue, and it's being open to another person.

**During:** Another point is that most of the insights we get from the feedback of other people come through problems of ethics, such as "you're behaving in a certain way, it infringes upon my rights." Most often they are not worded like this, but it comes down to "what you're doing to me or to us is a problem." So it's really ethical feedback. It's not an insight or inner psyche. You have a privileged access to whatever happens in your mind, I'm convinced of this, and the feedback you can expect is directly moral or ethical feedback. So it directly contributes to this ethical self-modeling. It's interesting in that particular way. I'm not expecting very deep, insightful feedback on my inner personality from people. I'm really mapping out my relationship with others in terms of rights and what am I doing wrong to you.

**Kamm:** I wanted to ask the psychologists here. There are certain techniques that are used, mostly in the Eastern traditions, for example, mantra meditation, which has been studied by Western scientists; sometimes it's called the relaxation response, sometimes it's more sophisticated. And people who practice these techniques find that a lot of the inner chatter, the conflict about "what should I do, should I go this way or that way," tends to go away and there is a confident sense of what one really wants to do or what one really should do. Not as a result of dialogue, not as a result of introspection or inference, but simply as a result of using a very mechanical technique that seems to affect a certain part of the brain. Do you think that if someone were to have a difficulty in life figuring out what he should do, if the reports of these effects of the mantra, for example, are true, that this would be a way to discover what he really wanted to do?

**Wilson:** Very interesting question, and while I'm no expert on the research on this, my reading from somewhat of a distance is that meditation of these sorts can be very useful for turning off the chatter and for achieving at least temporary peace of mind. Whether it actually gets us more in touch with our so-called true desires, I'm not so sure. If it's turning off the bad introspection that's confusing us and misleading us, then perhaps. But I don't think it gives us a direct pipeline into something that's true.

**Kamm:** Okay. Thank you.

**Voorhoeve:** I have another question here from the audience. "What are the principal obstacles to gaining self-knowledge?" Who wants to tackle that one?

**During:** Jack Nicholson answered this question. . . (*participants laughing*).

**Voorhoeve:** You mean Nicholson saying, "You can't handle the truth"? Now we've quoted the great Jack Nicholson in the same breath as Descartes (*participants laughing*)! Can you elaborate on Jack Nicholson?

**During:** Not really. The interesting paradox is that, to come back to this notion of imperious self, the more we gain knowledge of this personality within us, which we want to get rid of, the more difficult it becomes, in a way. Because this personality in us is also what prevents us from facing the truth. It's all the self-justifications, the rationalization that we give ourselves, in order to avoid framing problems in ethical terms. But as the process unfolds and we get to know ourselves more, this impediment, this obstacle, fades out. It can become a very interesting process. You can really get some taste of self-knowledge as a very interesting, passionate activity at some point. That would be the natural solution: that the obstacles just, at some point, disappear.

**Voorhoeve:** Is the obstacle that we don't desire the truth, or is it that some part of us knows it would be frustrated if it were made conscious?

**During:** Yes, this egoistic surface level of the self, which is driven by self-interest, has no particular interest in knowing what it is acting for. But the problem is solved in a natural way as soon as you start practicing, that is, modeling yourself in those terms and experimenting on what you're able to transform—the parameters you can alter, which is the process of ethical self-transformation. There's no principal *a priori* obstacle to self-knowledge. I think it's merely a matter of getting the process.

**Wilson:** I will mention one that Frances alluded to and agree with, which is that our culture gives us lots of expectations and stereotypes about who we ought to be, and it can be difficult to see through that smoke screen sometimes. It can be hard to see that that's not who we are, or that we have certain traits or desires that don't conform to those cultural expectations. That can be very difficult to sort out sometimes. And it takes a lot of good self-observation to do so.

**Jopling:** Here's another obstacle, a very simple observation; it's as close to a truism as I can get: reality is richer and more complex than we can ever figure out, and that includes ourselves. I think the self is enormously complex—its causal history, its developmental history, various components of the self and their interrelations with one another—it's really complex, and it's hard to figure out. There is one other obstacle that comes to mind, which is this: in some respects, it's easier to remain self-ignorant or self-deceived.

Tim you know this research on creative self-deception and depressive realism. The research seems to be showing that people who maintain creative self-deceptions about their abilities, their looks, their talents, their future, and also positive illusions about themselves, tend to be more well-adjusted, have better health, and be better off than people who have sober, realistic, truth-tracking accurate self-knowledge. I don't know how the research stands up methodologically, but it's a very interesting idea. And so the authors of this work, Shelley Taylor being one of the main ones, actually defend the virtues of creative self-deception over accurate self-knowledge (*participants laughing*). So, that's yet another obstacle: it's perhaps just easier and more conducive to well-being to be self-deceived!

**Kamm:** What I wanted to say was somewhat in the same line, though a little bit different. If you are succeeding at the things that you value and you are interacting successfully with others, it

will seem that that is an obstacle to your investigating yourself. I don't think it's a bad thing. Because ultimately, your investigation of yourself—if you're not just a scientist who wants to give an adequate diagnosis of yourself—is for another purpose; it's for succeeding in what's important and successfully dealing with other people and living up to values. And if you're doing that, then you're doing the most important thing. And it seems that you needn't investigate this other part, this other aspect.

But I want to bring it back to the investigation that the philosopher of my sort uses in dealing with these hypothetical cases, in finding out why exactly do I have that intuitive judgment—forget about the truth of it, the objective side of it—there is the rush to judgment. People are always thinking, "Oh, it's this factor and I finally found it." And tomorrow a counterexample comes up. So one of the obstacles is that it takes a long time and it can be hard, and the tendency is to want to say, "I found it."

**Voorhoeve:** You and others mentioned that self-knowledge is really an effortful enterprise. We have to make inferences; we have to try to see ourselves the way others see us; we have to engage in a difficult dialogue with an unpleasant Socrates. So why do it? Well, you all answer: "Do it only if it'll make you more successful at your projects or will help you figure out the right projects to pursue." But one audience member has an opposing view, writing: "Knowing yourself in order to make a change or decision seems less important to me than doing so simply in order to fully know yourself for the sake of doing so." How would you respond?

**During:** There are so many things we can know about the universe.

**Kamm:** Yeah. I'd rather know the cure to cancer than know everything about myself, I think. So, you have a choice to make about what you know.

**Voorhoeve:** That's a very instrumentalist view about the pursuit of self-knowledge! I'm shocked (*participants laughing*).

**Kamm:** The important thing is to be a good self. And if someone were a really good self—like one of Dostoevsky's characters in *The Idiot*—he might not realize he's good or be very self-reflective, but act from the best motives and respond to others correctly. To *be* a certain sort of person rather than to *know* that you are that sort of person is important. Bernard Williams has this phrase "One thought too many" that may apply here. To *be* a Tristan could be more important than to *know* that one is a Tristan. Too much reflecting on what one is may be unnecessary.

But someone may be intrigued by the whole idea of human personality. That we can reflect on ourselves; that we have consciences. Then it's not so much an interest in oneself but in the possibility of self-knowledge. So when Professor Wilson said he thought self-knowledge was such an important topic, he didn't immediately consider himself, he just thought about the topic, you see? And those are different issues.

**Wilson:** Yes, I would second that, and I think there's some recent research of which I'm fond that shows that if the goal is to be happy, then helping others and doing volunteer work is a better path than sitting around navel-gazing. So, it depends what your goal is. But if it's to be happy, then an outward focus is often a better path.

**Voorhoeve:** Well, on this surprisingly instrumentalist note about knowledge of the self, I think it's a good time to end, especially since we've just been instructed not to navel-gaze too much. Before ending, I want to thank all our panelists for their contributions.