

Inner World as a Useful Fiction: An Interview with Adam Toon

Bruno Mölder

Department of Philosophy, University of Tartu

Adam Toon is an Associate Professor at the University of Exeter and Director of Egenis, Centre for the Study of Life Sciences. He did his PhD in the Department of History and Philosophy of Science at the University of Cambridge. He works in philosophy of science and philosophy of mind, mainly investigating the idea of representations as fictions. Toon is the author of *Mind as Metaphor: A Defence of Mental Fictionalism* (Oxford University Press, 2023) and *Models as Make-Believe: Imagination, Fiction and Scientific Representation* (Palgrave Macmillan, 2012). He is also co-editor (with Tamás Demeter and Ted Parent) of *Mental Fictionalism: Philosophical Explorations* (Routledge, 2022). Adam Toon gave the Gottlob Frege Lectures in Theoretical Philosophy at the University of Tartu on December 5–7, 2023, under the title “The Story of the Mind: Cartesianism, Behaviourism, and Fictionalism”. The interview took place in Tartu on December 7, 2023.

Keywords: Toon, mental fictionalism

How did it all begin? What made you choose philosophy as a profession?

It is a strange choice as a profession, isn’t it? I studied maths and theoretical physics as an undergraduate. After a while, I felt like there were broader questions that I wanted to be able to ask about the sciences. I was very lucky that I was allowed to study history and philosophy of science in my final year at Cambridge. That was a wonderful experience, and gave me a sense of the history of science in all its complexity and its social context. At the same time, it presented an opportunity to think about questions that often you simply have to take for granted in the sciences—like what scientific method is, or what causation is, or what laws of nature are, and so on. I was very fortunate to go to Peter Lipton’s lectures on those topics. I felt like I found

Corresponding author’s address: Bruno Mölder, email: bruno.moelder@ut.ee.

an intellectual home and just carried on as long as I could get away with it! I don't remember making a particular decision about choosing it as a profession, but was just surprised and delighted to be able to keep going and get a scholarship to do my MA and PhD.



Adam Toon in Tartu in 2023. Photo by Bruno Mölder.

Although your latest work focuses on the metaphilosophy of the mind, your background is actually in the philosophy of science and you have written a book titled *Models as Make-Believe*. Could you tell us what was it about? Are you saying that models in science are basically like dolls?

I am, as strange as that might sound! The starting point is that there are many philosophers of science who have been interested in models, and there are many authors who think that whenever we want to apply theories to the world, we have to use models, by which they mean representations that are simplified or idealised in many ways. My first book was about the idea that we could understand modelling in science as something rather like what happens in fiction, in art and in children's games, in particular. I tried to show that scientists imagine, say, the plastic ball and sticks of a chemical model to be molecules, in something like the way in which a child might imagine a lump of plastic to be a baby. They also imagine the actions that they carry out on the model as being actions carried out on the molecule. In the case of theoretical models, a better parallel is fiction. A work of historical fiction like the novel *I, Claudius* by Robert Graves is a representation

of events in ancient Rome. Some of what it asks us to imagine about those events is true, but some of it is just made-up for the story. Similarly, the central method of representation in the sciences is that scientists ask us to imagine parts of the world in ways that they know they are not: they ask us to imagine, say, a pendulum in a grandfather clock to be a point mass on a massless spring. We know that those claims are false but, rather in the same way as a historical novel can still allow you to learn interesting things about ancient Rome, those models can still allow us to learn interesting things about grandfather clocks and other parts of the world. So, the main focus was on that kind of *as if* thinking that we have in modelling in the sciences, but arguably in lots of other domains too.

If models are fictions, then what are theories?

Nowadays a very common view is that theories are just collections of models. A theory like Newtonian mechanics is, roughly speaking, a set of generalisations. What it does is characterise a particular set of models like the simple harmonic oscillator or object in free fall that we can then use to represent particular parts of the world.

So science is all about models?

There are different views on that. From my point of view, the distinctive feature of modelling is the *as if* thinking that I mentioned: the distinctive feature of modelling is that it involves simplifying or idealising the world in certain ways. It involves taking your eye off the complexity of the real world for a moment and imagining the world to be simpler and more idealised than it is. There may be modes of reasoning in the sciences that do not do that and so do not count as models, but certainly the consensus seems to be that an awful lot of scientific theorising does involve modelling.

Your Frege lectures at the University of Tartu were centred around a view called “mental fictionalism”. Could you please explain what is the content of the view.

The starting point is the thought that we often think about the mind as a kind of inner world, and there are many views that take that inner world seriously in one way or another. In the lectures, I described those views as *Cartesianism*. In the past, we might have taken that inner world to be something non-physical, like a spirit or soul. Nowadays, many people will take that inner world to be something physical, like the brain. What I tried to argue is that the notion of an inner world is a useful fiction. We talk about people *as if* they had those inner worlds in order to make sense of their

behaviour. This is very useful, but it is a fiction because they do not really have them.

That inner world is purely conceptual and does not include any colours or feelings. It is an inner world of attitudes, right?

Yes. I have focused on thoughts and, in particular, on what philosophers called "propositional attitudes", like beliefs, desires, intentions, decisions, judgments, and so on. Within the broad view that I call Cartesianism, the standard way of understanding thoughts is as representations within the inner world. I have tried to argue that this way of talking about these inner representations is a fiction. Or, to put it slightly differently, it is a metaphor, based on external representations, especially language, but also pictures, diagrams and so on. So, you are quite right to say that I have focused on just one aspect of the mind, that is, thoughts. Of course, there are other aspects of the mind too.

Some philosophers of physics argue that the everyday world is a fiction because, in reality, it consists of quarks, elementary particles, vast empty spaces between electrons, and so on. The chairs and tables that we see are just fiction.

I think that is probably the wrong reading of the relationship between the world of fundamental particles and tables, so I would not say that tables are fictions for, example.

So what are tables?

What they are ultimately, I suppose, is collections of particles, but that does not mean that tables are not perfectly real as well. To focus on the fiction idea, the thought here is rather different from the table example. Quite often we operate with claims that we know are false, but we adopt them for certain purposes. That is really the basis of what I wanted to say about the mind: we talk as if people have contentful inner states, inner representations that capture their thoughts, but through philosophical analysis we are brought to see that this is not something that we take entirely seriously. We would probably take the existence of tables very seriously.

Someone might look at you with an incredulous stare: "But I can know my own mental states very well. I hear my thoughts in my mind, I feel my desires and emotions. How could they be just useful fictions?" How would you reply to such a person?

I think that there are aspects of the mind for which we do have a kind of privileged access, as philosophers sometimes put it. Of sensations like a grumbling in the stomach or a feeling of pain in the leg, it is quite right to say that we feel those sensations ourselves and that in a certain sense we cannot be mistaken about them. So, if I was telling people that they did not feel pain, the incredulous stare would be appropriate. The mistake, I think, we make too readily is to move from the fact that we feel those aspects of minds, like pains, to thinking of all aspects of the mind as being similar kind of inner states. So, for example, in the case of beliefs and desires, sometimes our sensations are an indication that we have a certain belief or desire. If we feel a certain sensation like surprise, when we walk out to the car park and see a gap in the space, that is a fairly good sign that we believe we parked our car there. But there is still a gap between those sensations and our thoughts. In other words, if we can access sensations fairly immediately and directly, the same is not true of our beliefs and desires. We have to, in a sense, interpret our own sensations and behaviour in order to attribute thoughts to ourselves.

For mental states that are a mixture of both sensations and thoughts, such as emotions, would you say that they also involve interpreting one's feelings or sensations?

I did not talk directly about emotions in the lectures, but I think that a roughly similar story can be told about them. One issue here is that the category of emotion, and this is a point that is familiar from Gilbert Ryle's discussion of emotion already, is a sort of mixed category. Sometimes when we talk about emotions, we mean feelings like a pain in the stomach or a pang of indigestion (to use one of Ryle's examples), whereas sometimes we mean something like happiness, for instance. In something like the way in which an interpretive gap exists between our sensations and thoughts, there is a gap between immediate sensations and an emotion like happiness. We can be happy for a few days, a week, or perhaps even longer, and during that time we might feel all sorts of different sensations. One might burn one's hand on the stove or feel hungry. So happiness looks like something longer lived than a sensation like a pang in the stomach. Nevertheless, it is associated with feeling certain feelings. Roughly speaking, a mood like happiness

depends upon a longer-lived disposition to feel certain things, to behave in certain ways.

We can be wrong in describing our own emotions. We can misinterpret the immediate sensations that we are having as part of that emotion—misinterpreting a pang of indigestion for a pang of anxiety, for instance—or we can misinterpret behaviour we are engaging in.

What about cases involving emotion terms that are absent in some languages, like *Schadenfreude*, for instance? Would it be useful to describe someone whose language lacks that term as experiencing *Schadenfreude*, even if they would not interpret themselves in those terms?

Good question. It is part of the view that I have tried to argue for that we should expect variation in folk psychology, in the language we use to describe the mind. I think that we can apply our own terms to describe the behaviour of people in other cultures. There is no impassable barrier. But I think your question points to the intricate feedback loops that can happen when we have a certain term for a particular kind of emotion that then comes to colour our own experience and behaviour. So it may well be that without that concept, we do not have quite that experience. I would want to resist the idea that that experience is just a bare sensation, like pain. I would imagine there is more involved in having the experience of *Schadenfreude* than just a particular kind of sensation, for instance.

You said that some mental states are useful fictions. But how could something be fictional and useful at the same time? On the pragmatist picture, very roughly, if a claim works in practice, it can be regarded as true. But fictional truths are not really true, since they are true merely in fiction. How could fiction be useful?

I think you are right to say that a fictionalist cannot accept a pragmatic theory of truth, at least in the straightforward way that you put it, because at the heart of the fictionalist view is the idea that certain claims can be false, but useful. And if you think that, roughly speaking, usefulness *is* truth, then it looks hard to make that distinction. If we go back to the case of scientific modelling, where typically a scientist will be perfectly well aware that some aspects of a model depart from reality and other parts do not, then it is part of what it is to come up with a good model that you are aware of that distinction. So we are aware that the claims we are making are false, but nevertheless they are useful to adopt in the sense that they allow us to make predictions. I am not sure that that requires us to adopt a particular theory

of truth, but it does seem to require us to think that we can tease apart the notion of truth and the notion of usefulness.

What are the advantages of mental fictionalism over other approaches to the mind?

I will mention two advantages. One is that despite how things might appear at first glance, if we go back to your point about the incredulous stare, I think that fictionalism fits surprisingly well with the way that we ordinarily talk about the mind. One way to make fictionalism seem more plausible is to start to ask yourself questions that take that notion of the inner world or inner representations literally, and you quite quickly see that some of those questions simply do not make sense. Now, some of those questions are familiar from similar motivations for dualism. So if we ask questions like where are beliefs, that seems to be a silly question. One explanation for that is that the mind is a non-physical thing. My own preferred explanation would be that those questions do not make sense because the notion of the inner world is metaphorical. To ask whether a belief is on the left or right side of my skull seems to me a silly question—much like saying someone has a chip on their shoulder and then asking whether it is on the left or right shoulder. One advantage of fictionalism is thus that it allows us to make sense of ordinary talk about the mind.

Another advantage is that it allows us to avoid a lot of the pseudo-problems that an alternative view, particularly Cartesianism, gets us into. Once you think of the mind as an inner world, then many of the classic problems of Western philosophy follow quite quickly: the problem of skepticism about the external world and the problem of skepticism about other minds, for instance. If you treat inner world as a useful fiction, you can see that we are not sealed off from the world outside and we are not sealed off from other people in the way that Cartesianism seems to suggest. So, avoiding solipsism may be one advantage of being a fictionalist!

Yes, but one could say that interpretivism, for instance, can also avoid Cartesianism and is compatible with the way the folk speaks.

Yes, there are certainly other anti-Cartesian views of the mind that have that advantage—and you would be in a much better place to talk about interpretivism, of course! One of the things that I have tried to say in the lectures is that some other anti-Cartesian views of the mind—behaviourism is a classic example—try to do away with the idea of an inner world entirely. Those views are often seen as being unworkable for various reasons. I have tried to argue that fictionalism fares better than those accounts, because it recog-

nises that the idea of an inner world is doing useful work, even if it is not true. One way to put this is to say that if Cartesianism makes the mistake of taking the inner world too literally, many anti-Cartesian views, such as behaviorism and arguably instrumentalism, make the mistake of not taking the inner world seriously enough—by which I mean not recognising the important role it plays in our language.

What are the downsides of mental fictionalism, if it has any?

I am probably not the person to ask that! I think there is, of course, the incredulous stare worry that it is not always a comfortable position to be in, to be telling your audience that they do not have minds (although I try not to put the view quite that boldly)! On a more theoretical note, fictionalism, as I understand it, does rely on some substantive claims about other areas of philosophy, especially philosophy of language, which are certainly controversial. In my own view, fictionalism relies upon being able to adopt a particular view of meaning as use within certain norm-governed social practices. I think that is a plausible theory of meaning, but fictionalism itself does not offer reasons to accept that view; it relies upon it. So, if there is one disadvantage of fictionalism that I feel most strongly, it is that it does rely upon some substantive positions elsewhere in order to work.

You say that inner representations are fiction but what is your main objection to the assumption that there are internal representations in our heads?

I just think it is not an idea that makes much sense and want to shift the burden of proof to opponents to show that it does. What I mean by this is that, at least at first glance, it seems that our ordinary notion of representation applies to marks written on paper, sounds in the air, public objects that are subject to particular norms or social conventions. The theory of meaning that I would favour stresses that aspect of intentionality. On the face of it, it is very hard to see how you could have states of the brain that gain meaning in that way. Now, of course, I am very well aware that there is a long tradition within philosophy of mind to develop a psychosemantics—in other words, to show how states of the brain could gain content in some other way, and I certainly am not able from the armchair to say that this attempt has failed or will fail. But I have the general impression that many people feel that this project has somewhat run aground lately. My main point would be that the psychosemantics project does need to be carried out in order for the Cartesian representationalist view to succeed. In other words, we cannot simply appeal to the success of talk about inner representations to demonstrate their

existence. We need to be able to show that this idea is coherent, that we can make sense of some notion of meaning or intentionality for inner states, and I am not sure that we have seen that yet.

You mentioned armchair already, so this leads to the next question: how would you describe your method of doing philosophy?

I don't have an armchair in my office, but I should get one! I was trained in what you might call old-fashioned conceptual analysis—that is, the idea that what we are trying to do in philosophy is, roughly speaking, to offer definitions, ideally a set of necessary and sufficient conditions to capture the content of our concepts. I suppose that, even if that is my training, my preferred philosophical method nowadays would probably be a combination of historical method and ordinary language philosophy. In other words, my general view is that concepts are woven into the pattern of human life and that they change over time. To borrow a conception of philosophy from the later Wittgenstein (or Ryle), many of our philosophical problems arise from mistaken uses of concepts and solving those problems involves untangling the concepts. So, for instance, Cartesianism is part and parcel of our ordinary talk about the mind. It is woven into our everyday life and practices. But if misunderstood and taken too literally, it leads us to all sorts of philosophical problems, like scepticism about the external world and other minds that have force for many of us. You could not make a film like *The Matrix*, for instance, if these problems did not strike many people as at least worth considering. To some extent, addressing those problems involves understanding how ordinary talk about the mind works and recognising how to dissolve them.

There is the philosophy of mind, and there are the sciences of the mind, such as cognitive science and neuroscience. What is your view on the relationship between them? Do you think there is anything distinctive about philosophy?

I have just outlined briefly a conception of philosophy that is one associated with thinkers like Wittgenstein and Ryle. That mode of philosophy is often linked with the idea that there is quite a sharp distinction between philosophical inquiry and the sciences. I am not sure I would go that far. If I go back to the example of *as if* thinking that I talked about earlier, it seems to me that is a style of thinking that you find in the sciences. But it is also a style of thinking that you find in everyday life, in metaphors, in children's games, and in many other areas. So, I do not think there is any sharp break

between the conceptions of mind that you find in everyday life and those in the sciences.

I also think there is a role for philosophy in questioning some of the metaphysical assumptions that are made in the sciences of the mind. For example, much of cognitive science adopts representationalism and there is quite a close relationship between many areas of philosophy of mind and cognitive science. Part of what I am saying here would have implications for the way that we ought to treat talk about representations within cognitive science, so I am not even sure that I would want to separate off philosophy of mind from the sciences of the mind. I am not sure that that is possible.

If there is no clear separation, is there something still distinctive about philosophy or is it just one big continuous area?

Perhaps it is a matter of degree. One way of thinking about it would be to evoke Wilfrid Sellars' well-known characterisation of philosophy as concerned with things in the widest possible sense and how they hang together in the widest possible sense. For instance, thinking about how and why we came to adopt the Cartesian conception of the mind requires a broad scope of inquiry that includes historical and philosophical reflection, and that is not something that would normally be expected of someone working in a particular area of cognitive science or psychology. So part of the answer is scope, and, of course, there are differences in method. Putting aside experimental philosophy, many philosophers do not carry out experiments. So perhaps there are differences of degree in terms of scope or method, but I am not sure there is a sharp dividing line between the two.

There is a big variety within the philosophy of mind as well.

That's right! Certainly, there are some people in the philosophy of mind who would not agree with the approach I am taking, as I challenge certain theoretical commitments of working cognitive science. This might be seen as being at odds with a naturalistic perspective, which holds that the deliverances of our best sciences are our best guide to the world. I am perhaps less chastened by that accusation than I should be, simply because cognitive science and philosophy are fairly closely connected, and cognitive science is also a relatively recent discipline. So I do think there is still a role for philosophers to question the assumptions that are made within the cognitive sciences.

Do you think that there is progress in philosophy? You seem to be quite fond of older philosophers like Gilbert Ryle and Wilfrid Sellars. Are we doing now better than they did around 75 years ago or are we still facing the same problems again and again.

What I admire in those thinkers you mentioned are two features in particular. One feature has already been mentioned by referencing Sellars' characterisation of philosophy as understanding how things hang together in the broadest possible sense. Of course, a lot of philosophy nowadays is fairly specialised and there may well be good reasons for specialising in certain areas, but I do think we lose something by that, and so I do admire the broad scope that you often find within older traditions in philosophy. I also very much admire Ryle's writing style and particularly his avoidance of jargon. Along with specialism nowadays very often goes a tendency towards specialised language that makes philosophy increasingly separate from wider culture. Again, there are good reasons for that in certain areas, but a lot of the questions that I am interested in are not distinctively questions for a specialised group of people called philosophers, but are questions that occur to anyone in their lives. So, I do value and respect the way that authors like Ryle write for a general audience and I think that there should be a little bit more of that way of writing philosophy nowadays.

The idea of progress goes hand in hand with a specific conception of philosophy: if you think that philosophy is about stating the final truth about knowledge or beauty, then you may hope that we could arrive there one day. As I said, the conception of philosophy that I follow is one that sees it as a matter of thinking about the concepts that are woven into our particular form of life (to use a Wittgensteinian term) and how they give rise to philosophical puzzles and problems. It is part and parcel of this conception that the work of philosophy will never be done, and that when new forms of life give rise to new confusions and new problems, there will always be that therapeutic work of the philosopher to engage in.

Why should the general public take interest in the philosophy of mind? Does it have any relevance beyond the ivory tower of academia?

I very much hope so, yes. Perhaps particularly with questions in the philosophy of mind as its central questions are ones that many people ask throughout their lives. Questions about the nature of human beings or the existence of the soul or the afterlife are not so distinct from the questions that are asked in the philosophy of mind. More broadly, our conceptions of mind are woven into our conceptions of human beings and their relationship to others

and to the external world, and in that sense they are also connected to moral and political life. Reflecting on the nature of mind and the nature of human beings is something that many reflective people inside and outside philosophy do all of the time. The wonderful thing about academic philosophy is that it can give you tools to try to think about those questions in a more rigorous way and try to make some progress with dissolving the particular problems that you have at that particular time.

Do you already have more or less concrete plans about what kind of philosophical issues you would like to tackle next now that the *Mind as Metaphor* is published?

Yes, I do. The book that I just finished was an inquiry into our current concept of mind. It is part of the view that I have put forward that we should expect notions of mind to vary between different cultures and to change over time. So I am particularly interested in trying to tell a historical story about how it is that we came to a particular conception of mind, particularly the Cartesian view of mind as inner world, and how that is woven more generally into a way of life.