The mind as an improviser

When we think about how our minds work, we tend to think that there is a lot of hidden stuff in there. If we really could think very carefully about all the things we know and why we think them, there'd just be an enormous trail of justifications and reasons. We can't necessarily probe all those, but if we start to think about them, some of them seem to appear. I think, "Why did I do that one thing? Oh, I did it because of this... and I did *that* because of something else", and off I go down a series of complicated justifications. It's easy to think that it was just tucked there away in my mind, and that what I'm doing is exploring this internal space, this mental world.

But think about it for a minute. Supposing I ask you (as indeed was asked in quite a well-known psychological study many years ago) how some everyday piece of equipment works, for example, the air conditioner? So you think to yourself, "air conditioners, well, I must pretty much understand how they work". Perhaps you could draw a little diagram, maybe a little exercise for you right now, just to see how you would draw a diagram to explain how an air conditioner works.

If you spend a few minutes having a go at this, you'll probably find that you are amazingly baffled. It's really hard to think, "Well, warm air goes in and cold air kind of comes out, and there's something pushing it through and it sort of just gets cold". How does that work? Well, after a bit of pondering, you might start to feel you don't really understand this at all.

Interestingly, large numbers of people, when probed on this kind of problem, tend to think they do understand it, and they feel their diagram pretty well encapsulates what an air conditioner does. But really, they don't. None of us do (or at least people who have done lots of physics and thought about it do, but it's pretty puzzling).

Now, air conditioners are one thing. We're not very familiar with those. But this is just a general fact about the way the mind works, that we have an illusion of understanding, which is much more rich and varied than the actual understanding inside us.

Say, for example, I ask you why you eat a particular breakfast cereal or why you live in a particular house. You may think, "That's like the air conditioner. I must have a lot of thoughts about that and I can give you some". But if you start to probe those thoughts and analyse them, you might find that they're actually just as feeble, just as incoherent as your thoughts about how air conditioners work. That's not surprising because human decisions, and the people who make them, are obviously far more complicated than air conditioners are.

Just one intuition to make you feel unsettled about the things you think you know and the justifications you think you have for your behaviour: Everyone who's had a small child or has contact with a small child will remember the phase in which there's a constant sequence of 'why' questions. The really puzzling thing about these 'why' questions is that you can answer very few of them, so things like, "why is the sky blue?", or "why do we drive on this side of the road and not that side?" –

almost any question a child will produce for you – has an immediate sense of, "that's a simple question. I know the answer to that, it's easy"... followed by utter bafflement.

I want to suggest that this is a very general problem. We're seeing the world not just through rose tinted spectacles, but almost as if they were panoramic spectacles. We see a world which we think we understand very well. We think we see lots of it. We think we have a full view of what's going on. But in reality, we don't.

So if it's the case that we understand much less than we think – we have an *illusion* of understanding about the world around us and ourselves – how is this illusion sustained? The answer is that whenever we ask ourselves a question, we produce an answer. If we ask ourselves another question, we cook up some answer to that as well. The trick is that we're very good at creating these instant answers and it gives the impression that the answer is, as it were, lurking in our minds all along. But the instant answer shouldn't be trusted. We just made them up.

Here's an interesting example, a very old example from a great social psychology experiment of Festinger and Carlsmith. They asked people to do a fairly boring task: they had some sort of peg board, and you had to take the pegs and twist them a quarter of the way around, and then do it again and again and again with many different pegs –very dull – and you were paid a small amount of money (I think it was $1) for doing this task. At the end of the task, you had to get someone else to do it. So your task was to go out and find another recruit. And you either got paid a small amount of money to get another recruit to do the task, or a really large amount of money (I think it was $20, which – when this experiment was done – was a very large amount of money indeed).

When you go out to get that person (which you're getting paid a lot of money to do), what do you suspect? You think, "the task must be horrendous; I've got to persuade some poor, unsuspecting person to do something which really has got to be fairly unpleasant because they're paying me $20 to get someone to do it. I'm going to have to pretty much fool them, deceive them, into doing the task because if I told them how terrible it really was, then they'd never do it". So it's almost bribery from the experimenters that you've been given the task of recruiting some hapless subject, so it must be a really miserable thing to do.

Of course, the person you actually try to recruit is, in fact, a stooge of the experiment, as it's all a con, but off you go. You're paid a little or a lot to get another subject, and then you are just asked in passing what you thought of the original task. And it turns out that, as you might suspect given the way I framed it, people who have been paid a small amount of money to get another subject think, "well, it wasn't too bad. It was a bit dull. That's OK. It was perfectly fine". But when you've been paid a lot of money to get someone else to do it, you think, "Well, it must have been awful". So you say "it was a terrible task. I hated it. It was very boring indeed, in fact, quite unpleasant".

Why do you think that? You think that because, as I've suggested, you're trying to make sense of the fact that you've just been given $20 to encourage someone else to do this. If they're paying you that much, it must be a pretty miserable task you're getting someone to do, so you must not have enjoyed it, it must have been a miserable experience.

Now, that's an interesting example of the general phenomenon. You might think as you go through a task (or any experience), "it's good", "it's bad", "it's indifferent". But according to the view that when you're asked the question, "what did you think of that experience", you have to cook up an answer. You don't really know the answer. You think, "I don't know. Let me think". Then what happened afterwards – in this case, whether you were paid a lot or a little to get someone else to do it – can affect your view. So it turns out that in this experiment, as in many others, you infer what you must have experienced based not just on the experience itself, but on things that came later.

A rather different, but interesting, related phenomenon, which you may have noticed yourself, is called hindsight bias. So supposing you are wondering whether something's likely to happen, or whether someone's going to succeed in something, or whether you're going to enjoy an experience.

Afterwards, you look back, and now you know the answer. Did it happen? Did you enjoy the experience or whatever?

Now, suppose we ask you what you thought in the first place. Suppose I go to a restaurant – I don't know whether I'm going to enjoy it, it turns out to be awful. Then you ask me, "What did you think it would be like before you went?" Interestingly, and very consistently, I tend, if I had a bad experience, to think I probably might have had a bad experience. I was tending to be suspicious that it was going to be a bad experience. On the other hand, if everything went well, then I tend to have a positive glow about what I thought in the first place.

If you're asking me after the event, what I thought before the event, I tend to shift and shuffle my thoughts in the direction of what I actually know to be true. For example, if you or I make a decision and it goes horribly wrong, we tend to think, "I should have known", or "I half knew", or "I feared this would happen". If it goes right, I tend to think, "I always had a feeling that was going to work out, that new hire was going to be a success, or those builders were going to be great".

The same story that we talked about earlier is going on. So it's not the case that as I go through my life, I have a clear idea of what I think and how strongly I think it. When you ask me what I thought about something, I can't look back in my mental record and say, "hang on, I've got a detailed record of all my thoughts, and there they all are written down, I'll just have a quick look". I can't do that. When you ask me a question about what I thought in the past, I have to cook up an answer now. I have to infer what I probably thought. And of course, if you ask me whether I thought something would be a success or a failure, and I know it was a success, perhaps a failure, I tend to think, "it did work out really well, I know that, and that probably was something I suspected, because normally, I'm a fairly sensible person". So I tend to twist my interpretation of what I thought in the first place towards what actually happened. I tend to be biased by hindsight.

Now, of course, this is something we see all over the place in the news and in daily life. People are always saying, "I knew that would happen", "I knew that was a mistake". And you may respond, "No you didn't – we talked about it and there was no mention of this at the time". But of course hindsight bias is a very powerful illusion and it fools all of us.

The reason is that we don't have, as we flow through our lives, a clear record what we believe, of what we value, of the nature of our experiences. When you ask me a question about what I thought in the past, I cook up the answer now. If I know stuff now (like it was a disaster, and it all went horribly wrong) I can't get that out of my mind, or ignore it to try and get back to what I really thought in the first place. It is quite possibly the case that there really is no answer to what I really thought in the first place. And we'll see that in a second.