ECNP Congress programme spotlight

Keynote Lecture – How we build internal models of the world

Ray Dolan is the Mary Kinross Professor of Neuropsychiatry and Director of the Max Planck-UCL Centre for Computational Psychiatry and Ageing. He is a Fellow of the Royal Society (FRS). In 2017 he won the 2017 Brain Prize, along with Peter Dayan and Wolfram Schultz. He will give the ECNP Congress Keynote Lecture **How we build internal models of the world** on Saturday 2 October 2021. Here he is interviewed by ECNP Press Officer Tom Parkhill.

"*We tell ourselves stories in order to live.*" Joan Didion.

"*When the facts change, I change my mind.*" attributed to John Maynard Keynes.

Tell me about your background. You're from Ireland, how did you end up where you are now?
I'm from Galway, in the west. I went to a village school, with a husband and wife team as teachers, and around 100 pupils. It was very deprived and heated by turf fires, with pupils having to bring a contribution to keep the fires going in winter. Eventually, the village parents went on strike, refusing to send their kids to school until there was an agreement for a new school to be built. After that I went to boarding school – my parents valued education above all else – and then to medical school, in Galway. When I completed medical school I was unsure about what to do with myself. Medicine seemed very conventional, and many of my friends at university were doing arts, which seemed more interesting and appealing.

It was very much an attitude of the period, the early 1970s.
Yes, I got caught up in the whole post-68 counterculture, and it was a really exciting time. I know people say if you can remember it you were not there – but I remember it well. Psychiatry offered an interesting point of reference within medicine, but also it was part of a big counterculture debate at the time, related to the whole ideal of freedom, madness, incarceration, etc.

The R. D. Laing period?
Yes, he was a Glaswegian, like you. I met him a couple of times; he was bright, brash and deeply opinionated. He came to a party I gave once at a house I was renting in Golders Green. Around this time he used to run birthing rituals, a sort of primal scream therapy, in a hall in Belsize Park, London. In fact there was an infamous play about this at the Royal Court Theatre. I went along a few times, but I was always really scared I'd be picked out as a subject for rebirthing. A colder eye now tells me it was all an act in group histrionics. I was training in psychiatry at the time. Despite the prevailing perspective of the counterculture, movement I gradually became interested in the brain. It was really hard to get traction on a research career in psychiatry based on the brain, and the discipline suffered from a weight of descriptiveness and the absence of a plausible scientific explanatory framework.

So how did your research career start?
I was fortunate to get a Wellcome Training Fellowship, which allowed me to do full-time research for the first time. This was freedom. Around this time I also connected with people at the Hammersmith Hospital who were beginning to exploit the potential of PET. This immediately seemed a way forward. Even in the early days it provided the first hints that it was possible to ground cognition in measurable brain activity. Out of early work here I developed an interest in computational neuroscience, and was taken by its ability to provide an important link for understanding brain activity and cognition.

I was fortunate that a few doors down from where I worked in Queen Square, we had the Gatsby Neurocomputational Centre, directed initially by Geoff Hinton, and subsequently by Peter Dayan, who has been a long-term collaborator. The collaboration was fuelled by our interest in the function of dopamine, its role in reward processing, as well as it providing the basis for a reward prediction error. More broadly, Peter and I were also interested in decision-making, particularly the duopoly of goal-directed versus habitual decision-making, and this became the frame of reference for a lot of work that has come out of my laboratory over the past 15-20 years.

A lot of this work involved the use of fMRI. One of the problems using this technique is that the time resolution is slow; you can only resolve neural events over the order of seconds. Consequently, over the last five years I've become more interested in capturing neural activity as it relates to cognition in real time. It turns out that the technology enabling this has been around for some time, magneto-encephalography (MEG). The interest in the lab has centred on the general question of how do you build a mental model of the world? An assumption here is that we all carry a working model of the world in our head, a model that is very useful for lots of things. For example, a mental model enables you to ground your assumptions as to how the world should be, what expectations we might have of future states of the world, as well as enabling you to simulate what might happen when you do something you have never done before. So over the last five years, we've been using MEG on how to understand how we build and maintain mental models.

This work led us to an interesting connection with earlier work carried out primarily in rodents, related to neural replay. The core concept goes back to John O'Keefe, who discovered that cells in the hippocampus encode the position of an animal in an environment (work for which he won the Nobel Prize). This discovery of place cells suggested the hippocampus is important in building a model of the environment. Subsequently, others found that after an animal has performed a spatial task you get this phenomenon called replay, whereby cells responsive to position when the animal moves in an environment replay back this trajectory when the animal is subsequently at rest or sleeping. This has been thought of in terms of consolidation of this recent experience in memory. Ideas on this have now shifted such that evidence indicates this might allow an animal to update a more generic model of the environment, as well as simulate future behavioural trajectories.

Working with a number of collaborators – Tim Behrens, Zeb Kurth-Nelson, Nathaniel Daw, as well as my PhD students Yunzhe Liu and Matthew Nour – we have shown that we can measure replay in the human brain. Working beyond the constraints of spatial navigation, this has enabled us to show that replay is not just a recapitulation of prior experience, but reflects an inference on the structure of that experience.

So it's a feedback model?
What we observe played out in replay is not just a recapitulation of what people are presented with in the context of a task, but an inference on the structure of the task. We've now taken this further and shown that replay not only updates your local experience, based on feedback as to what has happened, but also provides a prioritised assignment of value to likely future experience. Here we found two types of replay, one that updates your local experience, and the other that assigns credit to experiences you haven't had but are likely to have in the future, enabling you to a make better future decisions.

We have begun to address how replay might relate to psychiatry, starting with schizophrenia. Here the challenge has always been how to understand its key symptoms – hallucinations, disorganised thinking, delusions, and so on. We have begun to think about this in terms of a disordered cognitive map of the environment; and that patients with the condition do not update their assumptive model of the world. In fact we have shown that patients with schizophrenia do not have intact neural replay and this relates to how well or badly they do in a learning task we have given them. And I suppose this takes me back to where I started, with the challenge posed by mental illness. Now, I am more optimistic than ever that we can begin to make inroads into one of the most puzzling conditions in all of medicine.

Psychiatry has never really had many biomarkers.
Exactly. And it has been dominated by two fields, genetics and psychopharmacology. For example genetics has yet to live up to its long promise of providing a mechanistic basis for the disorder. Likewise, the pharmaceutical industry has also been dominant in the field and whilst there are therapies of variable effect, none are entirely satisfactory, with therapeutic interventions very much being hit or miss. The bottom line is psychiatry lacks any treatments that are grounded in any understanding of underlying biology, and that's a fairly unique situation in medicine. The recent work we have done on replay, including our finding in schizophrenia, is what I will talk about in my ECNP presentation.

You are still very much a practicing psychiatrist?
I've been working in clinical psychiatry since 1978, but I'm easing out of it now. I want to focus my energy on research and building on the work we have been doing over the last 10-20 years. It's difficult to do multiple jobs well, so I now want to focus on the challenges posed by serious psychiatric disorders.

In the abstract for your ECNP talk you mention Kenneth Craik. He qualified as a philosopher before moving into neuroscience.
Yes, he also had a psychology and engineering background. His book was a discovery for me (I only read it some 10 years after I was awarded the Kenneth Craik Prize). It's delightful. He seems to have had the notion of mental models well before Tolman, and I don't think he's been given enough credit for this. Some important notions on the field, such as expectation, prediction, prediction errors, and simulation are all there in his work. Tragically Craik died young. While cycling in Cambridge he was struck by a car, in 1945. The 20s, 30s and 40s were really interesting times in Cambridge, particularly in relation to the nascent field of decision-making, with people like Keynes, Frank Ramsay, Turing, Craik among others who came at the problem from many different perspectives. If only fate had allowed them to meet.

The idea of mental models interests me generally, not just in neuroscience. We all fit the world according to our models, emotionally, politically. Do you have any thoughts on how the neuroscience imposes a model at that level?

Interestingly enough, a former PhD student of mine, Steve Fleming, has been doing a lot of really interesting work on this. If you have a good model of the world, and if the model doesn't quite fit with reality, you should adjust your model to this reality. Steve has been probing this question in relation to political views, sampling people with opinions that cross the spectrum. He has shown that the degree to which people incorporate new information into their decision-making relates to where they are on this spectrum. People at the political extremes, left or right, are less likely to take on board additional information that might help them modify their previous decisions. It is difficult not to see this as enacted out in the large proportion of Trump supporters, including large chunks of the upper hierarchy of the Republican party, who reject evidence of the election outcome and subscribe to an unfounded belief in the 'big steal'. They don't want to update their model of the world to the new reality and this is deeply troubling in terms of how mass movements might compromise what we have come to assume are the very pillars of freedom and democracy. The very essence of science is that our beliefs should not be fixed but instead they should be endlessly malleable to adjust to new evidence.

Keynote Lecture

KL.01 – **How we build internal models of the world**

Saturday 2 October 2021

Register to the 34th ECNP Congress *Hybrid* via the button below:

Register here