OK, so thanks so much for having me here. Thanks to the University and to Jesse and Edouard and Jurģis and all of the organisers; it's a real pleasure to be here. OK. Let me get rid of some of this stuff [works on laptop] and make it look right. I suspect that my slides are going to be a little bit small and far away, but never mind because I'll just… hopefully anything that's important up there I'll say as well. OK. There's a few people I should thank on this one, most notably…most notably I think Karl Friston and Jakob Hohwy, they've probably had the greatest input to this particular talk.

OK. So the topic - what I want to talk about - is something actually slightly different to embodied and extended cognition, in fact to some of you it will probably seem enormously different. It will seem so different that you'll probably think that I've defected to the Cartesian opposition or something like that. But actually, I think that it will turn out to be the case that this is the perfect computational theory of what the brain's doing to fit together with embodied, and indeed extended, cognition. That's a kind of ongoing project of mine that I won't try and speak to today, actually. So, what I will try and do today is just two things; I want to try and get this view on the table -
hierarchical predictive processing. Just getting it on the table will take a little while and I’m going to allow myself the... the luxury of not offering you any evidence for the view. There is evidence for it, and I review some of it in a BBS paper that has just come out, but actually I think if I can just get the view on the table and then explore what it might mean, or suggest if it were true, that’ll be enough. OK. So, the particular issues that I want to look at are ones to do with the co-construction of perception, understanding and imagination; because I think this view speaks very directly to those issues.

14’08”
OK. So. Hierarchical predictive processing - most often known as hierarchical predictive coding - and again there’s sort of...a few reasons why I prefer predictive processing to predictive coding here, but we can get to those in discussion, if you like. So these views, what they do I think, is that they turn our sort of common sense understanding of perception pretty well on its head. This is a sort of Bayesian Inversion, if you like. Instead of trying to build up a picture of how the world is on the basis of lots of little incoming cues, kind of getting them together to form a bigger and bigger picture of what’s out there, you try to work out what the incoming cues are on the basis of what your existing picture of what’s out there. In other words the brain is trying - if you like - to guess the in-coming sensory data on the basis of how it knows about how the world is likely to be.

14’59”
So, there’s a more traditional view. This is how Egner and colleagues - who are opponents of the traditional view, but this is how they describe that view - they say:

“Visual cortex is traditionally viewed as a hierarchy of neural feature detectors, with neural population responses being driven by bottom-up stimulus features.”

Or, the visual system in those views was a passive analyser of bottom-up sensory information; these are views that Egner and colleagues want to reject. So, it’s a sort of Lego-type, building-block view of the route to sensory experience that is being challenged here, I think. What’s being challenged is a view of the brain as essentially passive and stimulus driven. So, I actually think of these views as something like the... the kind of culmination of the programme of active perception; this is about as active as perception could possibly be.

15’52”
OK, so. In these schemes, perception - and I’ve got a little caveat here that I do care about and I need to say it - perception in its rich, world-revealing forms - works from the top down. The reason why I’ve got the caveat in there is that there are all sorts of ways that perceptual systems can allow us to coordinate with the world, and do good things in the world, and they don’t all involve the exploration of a kind of... sort of... of the top-down use of a rich world model; not by any means. So one of the... one of the
challenges for integrating this work with work in embodied cognition is actually to see how we flip between these two different ways of dealing with the world and there’s a... there’s a nice story to tell - which I won’t be able to tell today, but I will mention it - where basically the flipping is accomplished in meta-Bayesian terms. So, you estimate which of the various strategies you currently have available is associated with the least uncertainty in a current situation, and you use that one. So it’s a sort of meta-Bayesian route to actually using either a quick-and-dirty strategy or a richer one. That’s a sort of nutshell version of how this is supposed to fit with embodied cognition for me.

17'05”
But let’s try and get the view on the table first. The idea here is that the brain - in particular the cortex - is trying to predict the present sensory signal by building it for itself using stored knowledge about the world. And I toyed with lots of different ways of trying to... of trying to introduce this; the one that I’m going to use is actually quite... quite informal. There are a lot of simulation-based ways of introducing it as well, but I think I prefer the informal way that I’m going to use, because I think it may give a better sense of what it would be like - if you like - to be a hierarchical predictive processing engine sitting in the head.

17'44”
OK. So, on these stories, what perception is trying to do is match the incoming sensory signal with a multi-layer, top-down prediction of that very signal, tracking it’s evolution at many different scales of space and time; and the thought will be when that works we perceive the world in this rich, world-revealing kind of way. And, when I say it’s rich and world-revealing, what I mean is that perceiving and understanding here are completely inseparable. To perceive in this way just is to understand; it actually is, I think, to be in a position to reason and it’s to be in a position to imagine stuff off-line as well. So, I think of it as a kind of cognitive package-deal. Instead of trying to get all these little bits, sort of, in different ways (we’ll do perception in this way, and then we’ll do reasoning in this way, and then we’ll do action in this way, and then we do imagination in some other way) you get to do it all at once using the same basic strategy; they’re co-constructed as I’ll try and suggest.

18'40”
So, there are two key elements in the... in the story here, the first is that perception and action depend on the learning and use of probabilistic, generative models. This is the bit of the story that I think is pretty secure; I’ll be very surprised if this turns out to be false. The other bit is less secure, it’s the idea that perception and action depend on the attempt to use those models to predict the current sensory signal, leaving only the unexplained bits of the signal to propagate forward; the residual error. That’s the kind of ‘predictive coding’ aspect of the... of the whole story. And that’s... that’s empirically
up for grabs, I think. That’s a rather... that’s the most radical way of telling this story; and because it’s radical I slightly prefer it because it brings some things into focus.

19'24"

So, I’m going to introduce these ideas informally, as I said. First of all, the most important bit of this story, I think, is the basic idea of using the generative model to predict the sensory inputs. So, here’s a... here’s an introduction that Dan Dennett gave to me, actually. So, I was stranded with Dan Dennett in a cabin up in Maine, because Hurricane Irene was preventing us going out on a sailboat at the time. So, during quite a lot of, kind of, enforced space up there, Dan remembered something that had happened to him back in the 1980’s. In the 1980’s a famous palaeontologist (who Dan won’t name for... presumably there’s some delicate reason for this) but anyway, a famous palaeontologist complained to Dan that his students were cheating at their stratigraphy homework. So, stratigraphy is that...it’s these sort of geological, layered strata drawings that you’ll have seen. A stratigraphy drawing literally is the drawing of layers, and it’s a geological cross-section that’s meant to reveal the way complex physical structures have accrued over time. So that the palaeontologist’s problem was that the students were cheating; they were just finding the right sort of picture in... in a book, and tracing it and bringing it in as their completed homework.

20'45"

What Dan imagined - and this was actually built, so a software engineer working for Dan at the time built this and they used it in some classes - was a little educational software package called “Slice”. Slice was a little, simple user-interface with a number of virtual knobs on it. Each of the virtual knobs controlled the unfolding of one basic geological process. So, for example, you’d have a little knob where if you twiddled it you could deposit layers of sediment; another knob where if you twiddled it you could erode; another knob for intrude lava; another knob for control fracture; and so on. And you can prob...so, this is a little bit like an Etch A Sketch in one way, and it’s exactly unlike an Etch A Sketch in another. So, those of you old enough to remember Etch A Sketch - I think it’s out again, actually - you know it’s a little thing where you can, kind of, twiddle knobs to make stuff appear on the screen. So, the new homework - I think you can see what’s coming now - the new homework is: you give the students the stratigraphy drawing (so they don’t have to go out and find it, they can’t copy it from somewhere; you’re giving it to them) but there job now is to create the stratigraphy drawing using this piece of educational software. The important thing is, of course, is that you can’t do this by tracing or simply copying. The little Etch A Sketch here won’t do that. It can’t... you can’t make a single pixel appear or move in any particular way on the screen here, all you can do is control the representation of these basic geological processes. Representations of intrude lava, fracture, etcetera. So what the device is forcing you to do is to match the picture that you’ve been given by twiddling the right geological
hidden-cause knobs in the right order. That means you have to find the right knobs, you employ them at the right intensities, and in the right sequence to yield the picture that you were given.

22'39"
So, Dan’s thinking was that: Look, if a student could do that, then she really did understand quite a lot about how hidden geological causes - sedimentation, erosion and so on - would conspire together to create a particular kind of physical scene; the one represented in the drawing. So, importantly remember, you can’t use it like a colouring book. The interactions here are complex, they’re non-linear, and you just have to get the hidden causes together in the right order to make the picture appear on the screen. A student who could do that would have to command a generative model; a model that enables her to construct various geological outcomes from their component parts as they interact in space and time.

23'22"
So, the next step in this - to push this up a bit - is now I think that the student has to command a probabilistic, generative model, because given a particular picture there’ll be a number of different ways of combining these different knob twiddlings to yield that picture. But, some of those will represent geologically much more likely events and combinations of events than others. So that means you need to find a set of knob twiddlings that correspond to the set of geological hidden causes that are most likely given all you know, that have brought about the observed outcome. So, we’re getting there know... this is very close to what the brain is supposed to be doing in sensory perception on these models. The next thing to do is eliminate the student (always a good move). So, Slice Prime now... now we go beyond, of course, that anything that Dan did - Dan didn’t eliminate the students; they all survived this process. Slice Prime is just like Slice, except that Slice Prime incorporates its own multilevel probabilistic internal model of how geological hidden causes combined to bring about the outcome. So, you have now put into Slice Star [think he means Prime] the knowledge that would otherwise be in the student. When Slice Star [Prime] receives a set of pixels specifying a geological image, it has to try to match that sensory input by unearthing the most likely set of geological causes, operating at many different scales of space and time, whose interactions are most likely to have yielded that pixel array. So this is now Bayesian; this is as fully Bayesian as you could want it to be. Slice Star [Prime] settles on a set of causes that, given what the system knows, make the current sensory evidence most likely. It tries to account for the sensory data by self-generating it, by using stored probabilistic knowledge about geological hidden causes. This is a very slim-line sort of, you know, one domain - only perception, no action here - it’s a very... you know, it’s a very slimline version of the trick that the brain uses (if the models that I’m going to be considering are right) to perceive the world.
So that brings us to that model: Perception as prediction. The idea here would be, well you know, in... in ordinary perception of...you know rich, world-revealing perception... then the incoming sensory signal needs to be matched by a complex, multi-level flow of top-down prediction. That match, in these models, is established by concurrent - it’s quite important that they’re concurrent, actually - concurrent bouts of signal passing in which some populations of cells send predictions sideways and downwards, and others send residual errors upwards. So, basically, you’re constructing your predictions with lateral interactions and downward-flowing causal influence within the brain; and what’s travelling upwards, forwards through the system on these models (these are the strong, predictive coding models) is just the residual error; just the difference between that...that prediction and the incoming signal at that level.

So we can think about this in fairly intuitive ways, you know, if you were a multi-layered system and part of what you knew about was words; and part of what you knew about was letters; and part of what you knew about was the pencil-strokes on a page; then your knowledge about words (at one of the higher levels here) would help you to predict which visually presented letter (say that’s at the level below) is currently most likely. Knowing about letters at that level would help you predict what combinations of pencil strokes are most likely at the layer below, and so on. It’s also important, I think, that people, you know, sometimes say “yeah, well where does the knowledge come from in the first place? How did... how did the right stuff get into this system to do this fancy thing?” But luckily the prediction task here is bootstrap heaven, this is a wonderful arena to pull yourself up by your own bootstraps. To predict the next word in a sentence, a really good thing to do is to know stuff about grammar. But, to learn stuff about grammar, a really good thing to do is to look for the best ways to predict the next word in a sentence. So what this means is that you can use the prediction task to bootstrap your way to the grammar, that you then use in the prediction task in future. And this is empirical-based now; you infer the priors from the data as you go along.

So now, and of course this is the handwaving moment - or maybe you thought it was all handwaving before, but now this is the real handwaving moment - now imagine this on a huge scale. By trying to predict your own evolving inner state, using a multi-level system, you’re driven to unearth the nested hidden causes (the latent variables, if you prefer) of the structure in the input signals. That’s the story we’ve been telling: To predict text; learn about words; and letters and strokes; and eventually meanings too. To predict all these other plays of sensory data that we get to grips with, at all those scales of space and time, [we] learn about other kinds of interacting hidden cause, like tables; chairs; cats; faces; people; hurricanes; football games; goals; intentions. This is a
huge *Etch A Sketch* of the mind now on these sorts of models where you’ve got hidden cause knobs that control predictions that are made on the basis of causes like that - causes like cats; faces; people; hurricanes; etcetera.

**28’45”**

And finally, although I’ve almost said this already (I was in two minds about whether to include this slide) but... yeah, focus for a moment on the data compression strategy here, the actual sort of hard-line predictive coding bit of the story. So predictive coding itself is a longstanding data progression strategy, it is used in motion compressed video and all kinds of places, and the idea basically is, you know: if all that’s changed between this frame of the video...

...is one little thing that moved from here to here...

...then in order to transmit *that* information, given that you had this before [Frame 1], all you need to do is transmit the difference. So, residual error then becomes the thing that you should care about. So, the kind of thought here is that this is a... this is a very active brain picture. The idea is then that the brain is, kind of, sitting there doing all of this ongoing prediction all the time. And that means that, in a way, it need only bother about whatever escapes the predictive net; wherever the residual error is. What you need to use here to drive more processing, to select a predictive model, or to nuance or amend the one you are using, or to learn a new one is to attend to the residual difference – the prediction error – between the current signal (at each level) and the predicted causes of that signal that you’re deploying from the top down. So, prediction error carries the news, if you like; that’s a common way of thinking about this. Prediction error is the kind of antihero – or hero – of the predictive processing family of models. So much so that, in fact, people talking about this will say “we actually don’t have a forward-flow of sensory information in these models. All we have is a forward-flow of prediction error.” (We can talk about that, if you like, as to what extend that’s a reasonable gloss on it.)

**30’34”**

So, in mammalian brains - if this is right - this all occurs in a hierarchical, multi-level, cascade of cortical processing. The classic references here are probably *Rao and Ballard’s 1999* simulation paper showing how this would account for some receptive field properties in early visual cortex and then *Lee and Mumford* have a, sort of, slightly larger-scale picture, and *Karl Friston* has the largest of all scale pictures in 2005.
So, summary so far:
Perception – rich, world-revealing perception – occurs when the downward cortical cascade can generate the sensory data for itself, using a linked series of models that have been acquired so as to capture regularity at different scales of space and time, basically. So what you see in the simulations here, is that the stuff that is higher up is tracking stuff that is working at longer temporal scales and larger spatial scales. So, closer to the peripheries you are tracking stuff – predicting stuff – at very fast scales of space...sorry of time...and small scales of space.

So all this recalls a dictum that I heard from Max Clowes, but I’ve also heard it from Rodolfo Llinás and have seen it attributed to Ramesh Jain (so if anyone knows where it actually came from that would be useful for me) but it’s the idea that “perception is controlled hallucination.” You might say that brains try to guess what’s out there, and to the extent that their guess matches - and as some say here “explains away” - the sensory data, we perceive the world. I’ll offer a sort of flip on that later, actually, but that’s what people say. And of course this same process can sometimes mislead – this is the famous hollow face illusion...

...this is Björn Borg, and that’s... what you’re seeing there is actually a hollow mask of Björn Borg’s face. So, you could do this with an ordinary mask if you just took it off your face and turned it around, lit it from behind, stood at a safe distance and observed. What you would see, even if you were looking in to the back of the mask, is an outward-looking face. It’s pretty obvious what’s going on there, it was Richard Gregory [who] made this argument back in 1980: The strong prediction of a convex face, there, the
strong bit of predictive knowledge that “noses stick outwards not inwards” here gets to beat – to trump – elements of the sensory input. There are elements there in the input signals specifying concavity, but they’re now being treated as noise due to the strong convexity expectation, and they’re simply ignored.

33'23"
So, I’ll say a word or two about the extension to action.
A big attraction of this story for me is that it has a very natural extension to action, and one that reveals action and perceptions as inextricably intertwined, constructed in the same sort of way, using the same basic computational principles. The basic idea here is that there are two ways that you can make your predictions come out right, corresponding (if you like) to different directions of fit. In perception you want to find the hypotheses that successfully predict the input. But, in action, you predict something non-actual and then you cancel out the error by moving your body around so that those predictions come true. So, the kind of thought here is that in action, the brain predicts a trajectory of proprioceptive signals that it’s not currently got, but then reduces the prediction error away by actually moving the body so as to produce that very set of... that very trajectory of proprioceptive inputs. So, that’s making the world conform to the prediction. So, we do it both ways and they’re supposed to interweave very delicately all the time. You want to find a model that fits the world and make the world fit the model.

34'39"
So, an intention... So this is a kind of version of the good old-fashioned ideomotor theory of James and Lotze, a kind of action here - an intention to act - becomes a sort of self-fulfilling prophecy. There’s a popular book by Hawkins and Blakeslee on these sorts of models called "On Intelligence" and they get this quite nicely, I think. They say:

“As strange as it sounds, when your own behaviour is involved, your predictions not only precede sensation, they determine it. Thinking of going to the next pattern in a sequence causes a cascading prediction of what you should experience next. As the cascading prediction unfolds, it generates the motor commands necessary to fulfil the prediction.”

So, if you put all that together, you get perception and action constructed in the same broad computational kind of way, and locked in a circular causal embrace. This is because your generative model, in these cases, will include expectations about how to use action to sample the world to determine whether the model you’re currently applying is adequate or not. So, in other words the generative model includes expectations about what will happen if you saccade around the scene in certain ways, for example.
OK, so some implications of this - and, the idea here is that these sorts of models perform this, kind of, Bayesian inversion of the standard story about perception: Instead of trying to build a model of what’s out there on the basis of low-level cues, we try and predict these low-level cues on the basis of our best model of what’s likely to be out there. And here are a few of the things I believe about this process. I think this is why we get to perceive a structured external world instead of sense data. We perceive a structured external world and not just plays of stuff across our sense organs, because we have to meet the transduced pixels (if you like) with this top-down cascade of interacting, represented distal causes. I think that the transparency of perceptual experience, the fact that we seem to simply see tables, chairs and bananas, falls rather naturally out of that; they’re the kind of things that are...amongst the sort of high-level knob twiddlings that would be most salient to the embodied organism. Perceiving and understanding – to the extent this is right – are co-constructed and percepts (you could say) are always at least weakly conceptualised. Secondly, top-down effects will be pervasive here. What we see, taste and feel will all be deeply impacted by what we (mostly sub-personally) predict and expect. And lastly perceivers like this are imaginers, too. They immediately have the resources to generate, from the top down, approximations for every scene they could currently possibly perceive. (And I'll return, later, if there’s time for that caveat about approximations.)

OK, so perception and understanding as co-constructed: I think that, you know, even from the *Slice* example it’s fairly easy why I would want to say that (if these models are right). To perceive the world – if these models are right – just is to use what you now to predict the sensory signal across multiple spatial and temporal scales. What you perceive is thus influenced at every level by what you know, but what you know is constantly conditioned, in that boot-strapping kind of way, by what you perceive. Perception and understanding, then, are inseparable. When the system like this perceives the world, it understands it. Because in perceiving it, it knows how the scene out there is likely to change across many different spatial and temporal scales; and I think that we can probably reconstruct the differences between basic and other forms of intelligence as a matter of the...the sort of extent of the... the spatial and temporal scales that you can get to grips with. As thought sensing and movement unfold in these stories, we don’t find any well-specified or stable interface between cognition and perception. We don’t find an interface at all. Believing and perceiving are co-constructed and constantly intertwined.
So, at this point it would be reasonable to worry why we don’t just see whatever we expect. You know, we don’t just see whatever we expect, it’s probably a good thing that we don’t and it probably wouldn’t be a great adaptive strategy...why not...

And then there’s another...so now there’s another element that I need to inject into the model and this is what most active research in this area is focussed on; this element. So in these models, despite all this emphasis on top-down prediction, the actual mix between top-down and bottom-up is not a fixed mix. There are three quantities in play in these models which are prediction, prediction error, and the weighting – otherwise known as the precision – of prediction error. So, that third quantity, the weight that is being given to particular bits of prediction error, varies according to how reliable – how noisy, certain or uncertain – different aspects of the signal are currently being taken to be. So, for example, prediction errors associated with the incoming, bottom-up input should be given less weight on a foggy day. While those associated with bottom-up information should be given more weight when you are visually attending to the fine detail of a [indecipherable word – possibly coin] or something like that.

So, this is normally really good news. We can still see very surprising things: The elephant smuggled onto stage and the, you know, revealed by the stage magician. We can see this, because under those conditions the brain assigns high reliability to an initially flurry of sensory prediction error. A high-level theory of an agent-unexpected kind now wins out to explain away that highly weighted tide of sensory evidence. It is important, then, that for the percept that for the brain is least surprising – the one that best minimises prediction errors or surprises – could well be for the agent some very surprising, very unexpected state of affairs; like the elephant on stage. The elephant is improbable, if you like, but not in the relevant informational sense - given everything that’s hitting the system at that moment and the context it knows it’s in - unacceptably surprising. In fact, it’s the least surprising percept of all the ones available, given the inputs, the priors and the current weighting on sensory prediction error. The upshot of this is that that third quantity, the precision of sensory prediction error reflecting estimates - changing estimates - of our own uncertainty, and estimates that vary with different modalities, is one of the big players on the cognitive stage in these stories.

These precision estimations are somewhat vulnerable. Mistakes in precision weighting sensory prediction error will have really, really big consequences here. So there’s a nice model of the co-emergence of delusion and hallucination in schizophrenia that basically turns on the generation of false prediction errors, but then the overweighting of those false prediction errors. If you assign too much precision to your sensory prediction
errors, you won’t be able to detect a faint pattern in a noisy environment. You might...and there are actually some accounts of...there are some accounts of autism that start from that sort of premise in these models. Or if you assign too little precision, you’ll start to hallucinate patterns that aren’t there, just because you expect to see them. And then, more subtly, if you assign high precision to the wrong aspects of current sensory prediction error, you’ll miss some things that are there and you’ll hallucinate some things that aren’t. And that’s probably the... the more familiar combination, actually. So there’s some famous examples of this. We can hear a famous song, if we’re told to expect it, in what’s really just white noise. So I think the famous song is “White Christmas” and students were told that “yeah, White Christmas is going to be hidden, very faintly, in this little sound stream.” And a good number of the students quizzed afterwards said “yup, I detected the onset of White Christmas.” And of course, we know that we can find faces in clouds and that sort of thing.

42'44"  
In cases like that, our active top-down model, just like in the hollow face illusion (I think) is causing us to discard some elements of the signal – we treat them as noise – and to amplify other elements of the signal; we say “Oh, that’s signal...that’s noise.” Normally, this is good, this is more accurate perception in more noisy and ambiguous circumstances; but not always. Here’s an example from Susanna Seigel; so this is how Susanna describes it:

“Jill believes, without justification, that Jack is angry at her. When she sees Jack, her belief makes him look angry to her.”

“Jill believes, without justification, that Jack is angry at her. When she sees Jack, her belief makes him look angry to her.”

[Hugh Laurie is my inclusion... I have no idea which picture Prof Clark used.]
I think it’s important to get these cases going - that the actually presented signal is, sort of... kind of on the... kind of neutral... sort of on the edge... that there’s... there’s scope for the uncertainty to really play a role here. So, that given, you know, I tried to find a sort of face that I could see as actually looking happy or angry. So I can do it with this one, if I... if I just prime myself in the right way, I can see that face either way. In the case of Angry-looking Jack, the thought is that our belief primes us to deploy a model that alters the precision that I apply to various aspects of sensory prediction error. In that way I discover visual evidence for the false and ungrounded hypothesis that Jack is angry. This will be just like hearing the song hidden in the white noise. But if that happens on these stories visual experience then represents Jack as looking angry. This isn’t a, sort of, add-on afterwards. This is... this is how peoples’ experiences is always constructed on these models. (And of course that’s fuel for the fire.) Action and perception here are locked together in a, sort of, mutually misleading cycle. The primed Angry Jack hypothesis gets to control actions that probe the world for confirming evidence of Jack’s anger: You saccade around that face looking for very subtle evidence; you look for tension in his limb movements; oddities in his words. And because – so, maybe I should have mentioned this: Precision is a zero-sum game here. If you up the precision on some aspects of the sensory signal, you have to down it on the others; it doesn’t work otherwise. So, because of that, you’ve upped the precision on subtle signs of anger and reduced it on veridical signs of normality. So, you’re very likely to find the evidence that you were looking for.

45'00
This isn’t just a philosopher’s thought experiment: Teufel, Fletcher and Davies have some nice... a nice review paper here, showing that our active top-down models of people’s current mental states and intentions make a difference to how we perceive them to be; to our perception of where they are looking; when they start to look there; the way that they are moving; etcetera. (This is, presumably, important in court cases.) And I think there are, sort of, other versions of this everywhere, you know, both in science and in every day life. We often seem to see what we expect, and we use that to confirm the model that is generating our expectations, and sculpturing and filtering our observations.

45'42”
It’s not all bad news, though. It’s intuitively obvious, I think, that a familiar song played using a poor radio receiver or heard in the shower, will sound to us an awful lot clearer than an unfamiliar song. You might if you were committed to a simple, sort of, feed-forward story think of that as a kind of add-on memory effect. But I think that it’s now reasonable to think of it as genuinely a perceptual one. The clear-sounding percept of the familiar song is now constructed in just the same way as fuzzy-sounding percept of
the unfamiliar one, it’s just that the familiar one invokes a better set of top-down predictions; has a better set of priors, so, it works better. Under those conditions I want to say that the familiar song really does sound clearer. It’s not that familiarity makes a fuzzy percept seem clearer; it’s not that memory later does some kind of filling in, affecting how we judge that percept to be. There are complexities here that I profess today, but one complexity is this: That, if your attention was drawn to the fact that radio has a bad sound, you’d actually be able to hear the fault in the signal, even in the familiar case, and in these stories that’s a matter of upping the precision on certain aspects of sensory prediction error.

[Sotto voce] I’ll skip that slide I think... yeah; I’ll skip that slide.

47'06"

OK, a word or two about linguistic self-scaffolding:
I think this offers a sort of framework that makes easy sense of various forms of language-based impact on cognition. So, words and phrases (whether we make them for ourselves; or encounter them in the world; or mentally rehearse them; or hear someone else say them; doesn’t really matter here) they’re all cues that can favour one hypothesis over another, they can alter what we perceive and they can make available new strategies for the control of thought and action. So there’s a nice experiment from Ward and Lupyan, using continuous flash suppression – this is a bit like the set up you use for binocular rivalry, the anaglyph glasses, kind of thing – and it allows you, basically, to one image to one eye and a different image to the other one. So, here you have an image continuously presented to one eye [gestures to slide] – this is in this case it’s that, sort of... angry-looking face – and then a changing stream of images being presented to the other eye. What Ward and Lupyan show is that an object that is marked from awareness by continuous flash suppression can be unsuppressed – unconsciously detected – if the right word is heard before a trial begins. So, if the... if the suppressed object was going to be a zebra, then if you hear the word “zebra” just before the trial the suppressed object isn’t suppressed, and you detect it after all. They say:

“We hypothesize that when information associated with verbal labels matches incoming, bottom-up activity, language provides a boost to perception, propelling an otherwise invisible image into awareness.”

And Lupyan has loads of other results that seem to point in that sort of direction. There’s also some nice work by Maloney and colleagues, very...one of those... actually it’s nice but it’s... I hate EEG experiments, to be honest; I find them so hard to understand, but anyway... It’s a complicated EEG experiment with a nice result, and the result is that expectation speeds the onset of conscious visibility.
OK. So, another thing about words: Words are clearly things that create expectations and can influence perception in the way you expect in these models, but I think there might be something special about words and it would be nice to know what it is... I wish I did, but I’ll... I’ll speculate slightly. I think one of the things that’s important here is that the effects of words can be very, very fine-grained. In a sort of a... a kind of... a sort of low-level demonstration of this kind of thing, Lupyan and Thompson-Schill show that hearing the word dog is better than just hearing a barking sound as a means of improving performance in a discrimination task. So, hearing the word does better than some other cue that, you might think, ought to be cuing the very same top-down model. I think this might be suggestive, regarding the role of language. Maybe language is a very fine instrument for artificially manipulating the precision of prediction error at different levels of neural processing. This might, I think, be a powerful way to think about a lot of the non-communicative, cognitive effects of language. Transient, targeted, subtle manipulation of precision could selectively enhance or mute any aspect of our world model, depending upon context and current purposes. So, the thought might be that there’s this, sort of, artificial second system here that language allows us to wrap around our own neural processing economy that lets us, sort of, reach in and fiddle with precision weighting – and the hence top-down, bottom-up balance – at any level of processing, so not just a simple, sort of, low level. That’s just pure speculation.

OK... Perception and imagination; the last one that I wanted to say a word about as co-emergent as well: So, an important feature of the... the probabilistic inner models that power these approaches is that they are generative in nature. That means that in the layer $n+1$ you’re capable of predicting the current sensory data at layer $n$, the layer below you, by means of downward connections. What that means, in effect, is that it means these systems can be run from the top down, generating virtual sensory data for themselves. I think that what this suggests is that, in creatures like us perception is co-emergent not just with understanding, but with something like imagination, and actually because of that, also with simulation (although, that’s another story; but, that’s how reasoning gets into the story.) That wouldn’t mean, of course, that a cat can deliberately explore possibilities in its imagination. I think that, for most of... most creatures on the planet, acts of deliberate imagining – in some strong sense of deliberate - is simply impossible. But, perhaps there’s a key role for language here, too, as tool to enable us to, kind of, artificially seed our own deliberate imaginings. At any rate, if these stories are right I think, any creature – if its able...if its perceptual organisation is working in this kind of way – if it perceives some state of affairs $x$, then it has the resources endogenously to generate an approximation for that state of affairs.
Why the caveat about approximation? There’s...there’s good evidence, I think, that in a lot of cases of mental imagery you’re only calling up on higher levels of those generative models. You’re not using those higher levels to drive prediction all the way down, as close to the sensory peripheries as you could. You drive it down only under particular experimental circumstances where the task really demands it. So, that the thought would be that we effectively ignore the lower levels, closer to retinotopic stuff, when we’re driving system top-down in imagery mode, if you like. Conversely, online perception might have some special features: It’s plausible to me that we can resolve prediction error in online perception at grains of level and detail that we just can’t do in imagination. For example, when we attend to the fine detail of the bark of a tree and so on. It might be that that’s impossible to do in dreaming and in mental imagery. If that were true, then maybe (although I hesitate to say this with epistemologists in the audience) but if that were true, then maybe when we’re awake we can actually know that we’re awake; even though when we’re dreaming (due to dampened critical reflection) we may not be able to know that we’re dreaming.

OK. Conclusions and puzzles.
Perception, understanding, imagination and action, if these stories are right, are co-emergent and they’re co-determining. They’re constructed using the same inner resources, and the same fundamental strategies apply. In fact, the difference between motor cortex and sensory cortex is pretty-well evaporated on these stories, they’re just... they’re predicting different kind of things, but apart from that, they’re doing the same sort of job. So the thought will be that, what we’re getting here is a kind of cognitive package deal; perception, understanding, imagination and action all co-emerging in systems like this at more-or-less the same time. Another implication is that top-down effects are going to be pervasive in the brain and will be felt at nearly every level and stage of neural processing.

So I’m going to end by briefly commenting on two common worries. The first is a worry about signal enhancement. Now this is a worry that you’ll have if you focus on the predictive coding side of this alone. So, the question is, can predictive coding - the technique whose signature is expectation-based signal suppression – so, you know, if you look at the fMRI stuff here, the... the key experiments that suggest that this might be the right model are ones in which expected elements of the signal, of the early sensory signal, are suppressed by higher-level expectation of that’s how that bit of the signal is going to be. So, expectation-based signal suppression, a sort of dampening of response when it’s predicted, that makes good sense just with regard to the straight predictive coding bit, but what about the enhancement of important or unexpected aspects of the
sensory signal? You can probably see where this comes from in these models. In these models, of course, it comes from the precision stuff. So, once you’ve got the precision weighting as a tool to vary the impact of precision error, then models like this can display these two superficially incompatible cognitive effects. You can have expectation-based signal suppression - lower-level stuff that’s well predicted by your current, top-level model - is explained away because there’s no news there. No signal propagates forward, all is left as it is, everything seems to be right; that’s the standard expectation-based signal suppression effect. But, you can also get salience-based signal enhancement. Targeted increases in precision-weighted prediction error here are a kind of attentional effect; they sharpen some neuronal effects, while dampening others. Where you predict reliable, important information in the world, response is sharpened. So, for example, if you detect a motion transient, the gain or volume of prediction error for that object or location in space will be increased making it more able to entrain action and higher level response. So there’s some nice work out there: Kok, Jehee and de Lange have a 2011 paper where they find exactly this combination of expectation-based dampening and then attention-modulated sharpening.

56'48"
And lastly, yes, something metaphysical. Doesn’t this sort of strategy threaten to cut us off from the world in perception, by inserting this rich generative model – or stack of generative models, technically it’s only one – between us and the distal environment? I don’t think it does, but you would be forgiven, perhaps, for thinking it does given the way that some people talk in this literature. Paton, Skewes, Frith and Hohwy – in a response to my BBS paper – talk about perception as generating a virtual reality, an inferred fantasy, about the world. And, of course, there’s the standard talk about perception as controlled hallucination. We can talk...I think you can see why you want to say something like this, because perceiving the world here requires meeting the incoming signal with this, sort of, model-based prediction; spanning many scales. But, of course, this is the very same process that reveals the world; that lets us get at the world. This is the process that enables us to sort out the signals from the noise, unearthing the various distal objects casting their sensory shadows (if you like) on the wall of Plato’s cave. So I think, instead of thinking of perception as controlled hallucination, we should think of hallucination as uncontrolled perception. In hallucination (if these models are on track) all the apparatus of perception is brought to bear, but with disturbed precision estimation, leading to inappropriate balances between top-down prediction and the incoming sensory signal. But it’s only because we’re normally capable of this very balancing act that we get to encounter the world at all, under ecologically prevalent conditions of noise, imperfect information and ambiguity. So, I think this is perceptual openness, but it’s perceptual openness for real agents operating in a real and noisy world.

That’s the end, thank you.
58'56"
[Dr. Jārgis Škilters]  Thanks a lot, so, time for questions, comments.

[Questioner 1]: I have a question about elephants and surprisal (back to the first part of the talk) so... so, take a case where you see in a circus a surprising elephant or something surprising is going on about an elephant... and you’re suggesting that an elephant is the least surprising percept in that context, so that still seems to be – even if we grant that to be the least-surprising percept in that context, it still seems be in tension with the claim that we perceive the world to the extent that it matches our predictions. It seems like some cases of surprisal are cases when we’re aware of discrepancies or disconfirmations; the very fact that we’re surprised indicates the fact that we’re not just perceiving what matches but we’re also perceiving mismatches. That’s one question, but I was also wondering about...

[Prof Andy Clark]: Maybe I should just take them in order, otherwise I'll forget that before I get to....

[Questioner 1]: Yeah, sure...

[Prof Andy Clark]: So...so, I think the right thing to say about that is prediction is still prediction in these cases, it’s just that your first prediction, as it were, the ongoing prediction that nothing much is going on on stage over there, turns out to be false. So, you have a flurry of prediction error that has to recruit a new and better hypothesis. So, maybe when I say it’s the least surprising percept, sometimes I prefer to say it’s the least “surprisal-ing” percept. That’s to say, you know, surprisal is this measure of the mismatch between your world model and the current sensory input; and, obviously, once you’ve found the right hypothesis to bring to bear, now there’s no longer a mismatch between your world model and the current sensory input and it’s not “surprisal-ing”. But, you the agent are probably sitting there thinking: “well, that was a surprise.” So...

[Questioner 1]: Oh, so you think that surprise is merely cognitive in that case...

[Prof Andy Clark]:... is what?

[Questioner 1]: You think that its just a cognitive surprise, not a visual surprisal, it’s not genuine visual surprisal...

[Prof Andy Clark]: Oh... hang on, so what’s visual surprise then? I’m...

[Questioner 1]: ... when you stare at something longer, you know... just a bunch of physiological reactions. So, in that case it seems like... the fact that there is visual surprisal going on, you’re aware that something is surprising it seems like you are aware of the disconfirmation; you’re aware of the mismatch...

[Prof Andy Clark]: Well, OK, now there’s an interesting question here that I don’t have a... that I don’t have... I don’t have a response to, so I probably shouldn’t even mention it. But, there’s this, sort of, question about whether we’re actually... whether we experience prediction error. Whether we experience the mismatch, if you like. And I don’t think we do, I think prediction error is just this thing that functions within us, to
bring the world into view. But some people working in music and prediction error, for example, think that we experience prediction error and that it makes a difference.

[Questioner 1]: Thanks.

62“00”

[Questioner 2]: That was a great talk, Andy. So, I’m kind of wondering, what do you include in the purview of your generative models? Because in a lot of Bayesian models you create these latent variables just on the basis of the statistical regularities that you’re extracting. Whereas, I could imaging that in more grounded generative models you’d also include highly constrained systems based upon, like, the physics of what you... what you’re seeing, so you’re only seeing this... collision of billiards, because you have a particular physical process which would predict these responses to these collisions. So, and...so, if you’re allowing both of these to be part of your generative models, then is it still...Bayesian?

[Prof Andy Clark]: Yeah... I mean, so OK, the... the... I use the word “Bayesian” with... with a kind of a... a kind of.... deep fear and angst. It’s...it’s... I think it’s Bayesian in spirit, but of course, because everything that is being deployed here involves huge amounts of, kind of, approximation and the model that we bring to bear to deal with some particular state of affairs in the world might be a huge simplification - or even a falsification – with regard to how things are actually operating in that part of the world. I think it’s Bayesian in spirit, because we’re trying to predict the...you know, we’re trying to explain the data on the basis of the model, rather than the other way around. But, beyond that, I don’t have any axe to grind over it, so it’s hugely approximation based. There’s... one other thing that... that is probably worth mentioning, is that when Friston does this sort of stuff everything gets to be part of the generative model, so your... your morphology would be part of the generative model you bring to bear on the world. Whereas another way of thinking about it is that that’s a, sort of, background condition that the generative model that is formed in, say, in...in cortical processing can just take for granted when it has to work out, you know, what to predict out there in the world. And I...I actually slightly prefer the non-Friston way of talking about it, because, once... you know...the bacteria is a generative model of its environment for Friston, and I...it disturbs me slightly, I know what he means but... but I don’t know what... the... the thing you were asking about physics at the beginning; I don’t think I got to grips with it, so I didn’t quite get it, what was the...?

[Questioner 2]: [initial part barely audible]...you could’ve have...you could’ve have completely statistics-based ways of creating these latent variables, and then when you’re trying to explain something, you’re going to be combining, partaking of these different latent variables in different combinations to explain what you’re seeing. But, there’d be a way that seems a lot less pure probabilistic reasoner, which would be, say... well we actually have models of how bodies move, that are subject to physical constraints, not just these probabilistic constraints. Or, with the billiard ball example,
it’s... we have these mechanical models of what happens during a collision, and that’s
constraining how we see things, rather than just pure probabilities...

[Prof Andy Clark]: [Overlaps] Oh, ok...Right, I see... So, this...this is, kind of, a question
about the content of the... of some aspects of the generative model: Is it all best
understood just as probabilistic expectation all the way down? [Pauses; thinks].
I suppose what I kind of think, but I can’t prove it, and we should talk about it is... is that
the... what we... what we gloss perfectly correctly as a model, rather than the
probability-density distribution, is actually implemented as a group on a probability-
density distribution. So, I think that the, kind of, the model picture here is one that we
should avail ourselves of as cognitive scientists (and Aaron Sloman has been arguing
about this stuff a bit) but... Then you ask the question: “How do you implement that
model?” and I... I think that the... our best clues about implementation are all
probabilistic, generative models that are dealing in density distribution. Can you...is ...is
Anyway, we can talk about this afterwards, I guess... it’s a good case...

66'25”

[Questioner 3]: I wonder how the perception/cognition distinction works on this kind
of... this kind of model, I mean...I take it the idea is not... the contents of perception aren’t
actually the contents of the sensory signal, they’re the contents of the model that
generates and predicts the sensory signal, but... On the face of it, the... the model that
predicts the... generates and predicts the sensory signal is going to have all kinds of stuff
that doesn’t correspond to stuff... to things we don’t normally think of as perceptual
content, right...you know, there’s stuff going on behind the sc... behind the curtain, my
best world model would say that people are walking out from behind the curtain, and
that’s part of what going to generate my predictions of the signal, we’d normally think of
that as a... more of a kind of cognitive representation, rather than a perceptual
representation. On your model, why don’t these just collapse?

[Prof Andy Clark]: Yeah...so there’s, sort of, two issues there, I think...or at least two
issues. One... one is a sort of question about how to, sort of, pin the phenomenal, kind of,
tail on this... on this donkey, because - you know - there’s all sorts of probabilistic
prediction going on here, and some of it... some of it doesn’t actually seem to us as if it’s
part of our conscious experience, like – you know – zero crossings, or something like
that. So, I... you know, that’s... I think will come down to stuff about the sorts of... the
sorts of embodied interactions that for systems like us make salient information
available. So, the stuff that will become highlighted in phenomenal space is the stuff
where an agent-controllable interaction can do something good to the... to the incoming
signal. So, that’s a, sort of, quick thought about that...As for the question in general
about perception and cognition, another way to think of... so... one way to think about it
is that... that it, kind of... that the higher levels here are...could be thought of as high level
percepts. There’s a kind of Gibsonian... [see also here, here, and here.] Gibsonian way of
thinking about the higher levels here where they’re... they’re... they’re, sort of... (if this
makes any sense) abstract perceptual states. So, a very, very high level percept predicts a whole a whole multi-modal flow of activity.

[Questionner3]: [Response is indecipherable, does not have microphone.]

[Prof Andy Clark]: [Some of this overlaps with the questioner’s response]

Yup...yup...yup that’s true.
Well, if I... If I... If I don’t...
If I... If I think of you without the...the kind of... the kind of linguistic overlay, then I think I can get away with saying something like: What you’re describing here is just the fact that your... your model involves high-level perceptual states that have, sort of, flowing... flowing sets of multiple temporal and spatial scale predictions coming from them, so, you know, even if Rob was in the coffee bar, you could still have, you know...that particular thing going on. [Pauses; thinks.] I know there was something else I was going to say, but I can’t think what it was...OK...so...

[Dr. Jorģis Šķilters]: So, perhaps, two last questions...and, well...

69’41”

[Questioner 4]: So I can see how the model is extremely effective... So I can see why the model is extremely effective for a lot of... for perception, and I was just a little bit worried about bodily sensations, and particularly pain, for a couple of reasons, and what... Firstly, it seems that it’s hard to understand pain in terms of just purely indicative contents, you want something like imperative contents, ways you should respond. And secondly, the very notion of perception as hallucination; I mean, it seems tricky even to make sense of the idea of hallucinating pain as opposed to actually perceiving pain... people might disagree on that, but certainly pain doesn’t seem to be easily captured within a mere prediction about how things in the world are.

[Prof Andy Clark]: That’s not obvious to me, I mean, the pain signal seems to me to be as... as much a part of the set...you know, nociceptive prediction seems as good as any other kind and as apt for the, sort of... as apt for going wrong if you have, you know, the wrong balance between top-down prediction and bottom-up sensory signalling. I... I think the pain case should... should fall quite well out of this, actually, given all of the... you know the, sort of... top-down effects on how we experience pain. Yeah...But it’s an interesting case, I mean, another one of course is... is smell and there is some nice work on prediction and smell. And I know that, also just... I did remember what I meant to say to Dave just before there, which is there’s a whole other story there to tell, Dave, about... about what language does to this. So, you know, when I try and predict your behaviour part of my model is that your a propositional attitude type of agent, that has a grip on your self as an agent in those terms, and that’s going to...that going to make things complicated.

So, the pain case: Let’s talk about it some more afterwards, ’cause I think its a...It would be... it would be very bad news if it didn’t deal with that case, I think.
[Dr. Jorģis Šķilters]: OK, and now the final question...unfortunately we are a little bit out of time, so we’ll start the parallel sessions ten minutes later, but the final question is [?] Burgess

72’03”

[Questioner 5]: Thanks for a wonderful talk and a great way to start the conference, Andy. I want to ask something about the grandeur of the picture, because you’ve given us a, kind of, programmatic formula – Perception is Prediction. And you’ve said that almost like it’s a general theory of the mind, like connectionism, classicism and now we have prediction theory, forward coding, all this kind of stuff. But the problem is perception does all kinds of things pervasively: It sharpens; it calibrates things that occur in the nervous system at different points in time. It does something that’s almost the opposite of prediction – it uses... it uses the... the... sort of present - if you think about these, sort of, time cases where the past is re-written by the future or something like that. So, there are all kinds of ways in which perception is integrating information at different timescales. So, we... we could talk about general principles of perception, and suppose – just take the simplest one where there’s sharpening, where there’s filtering to get rid of noise. So, why not just say that perception is noise filtration? Why pick one of the general things perception does, maybe contingently, I mean... these are ways to improve perceptual information; use prior knowledge to improve perception, but you could also use resolution filters to improve perception. All these seem not to be the essence of perception, but among the important and pervasive tools in the perceptual tool-belt.

[Prof Andy Clark]: So. I can see a challenge, you know, a reasonable one about... you know, why pick that particular phrase out of all the phrases you could have used to describe this whole space of models and I think that the response to that is that, the good thing about the prediction machine kind of story is that I think it gives you a framework in which you could easily slot all of the things that you mention. All of the things about, you know, salient space signal enhancement and, you know, embodied action, direct perception kinds of stuff. I think we can slot all of those in very neatly by, for example, thinking that a really important job of the rich-world model is to let you know when a thin-world model will do, that kind of stuff. So, I just use it as a, sort of a... I like perceiving as predicting, because I think it gives us a handle on an overall economy in which there’s loads of things going on, but they all have to make sense within an economy that is based on probabilistic, generative models. That’s... So, I could call it “Perception as always based on probabilistic, generative models”, but it didn’t sound quite so punchy...[Laughs.]

[Dr. Jorģis Šķilters]: OK, thanks a lot, and we are....[Applause].