[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: So we got three classes left. Today we're going to finish talking about attention. And then we're going to talk about touch, taste, and smell. OK?

So last time we got started with attention, talking about the fact that it's kind of an ill-defined thing. It's a word that we're all probably comfortable using. And the idea that it's mostly been studied in the context of a pretty wide range of phenomena that don't necessarily always have a unifying framework, and that as of now, we're mostly lacking, I would say, a rigorous computational understanding. But there are a lot of interesting phenomena, and I think the time is ripe for actually making progress on a real, better theoretical understanding of attention. And these sorts of phenomena, I think, will provide the backbone for that.

So we started off talking about cuing, and the idea that you can present a cue which provides information about the probable location of a subsequent stimulus that you have to either detect or recognize, and the fact that cues help you to detect something. So you get this dot here that indicates where something is likely to occur, and then you're faster at identifying what the letter would be. So the benefit of cuing is typically measured as a reduction in reaction time.

So one of the findings that falls out of experiments like this is that the benefit that you get from a cue is localized in space, right? So there's some thing that happens to the subsequent processing of the stimulus that is spatially localized. And so it's often thought of with a spotlight metaphor. So there's this spotlight that gets kind of shown on some region of space that in some way, in some unspecified way, improves the processing of something that cashes out in the experiment as a faster reaction time.

And then we talked about the difference between exogenous and endogenous cues. Where the exogenous cue, an example of which is shown here. So the dot pops up. It's like a physical transient. There's this involuntary reflex where attention gets drawn to that location and makes you faster at that location. But then there are also these cues that really require interpretation. So the arrow is indicating that something is likely to happen in that direction. And you have to know what the arrow means in order to do that.

And so you can see differences between these two types of cues and plausibly two distinct types of attentional mechanisms in the latency at which they kick in. So we looked at how long it takes you to get a benefit from the cue. And this is an experimental paradigm where the time between the cue and the target is varied. And so what you find is that when the stimulus onset asynchrony, or SOA, is really short, you don't actually get a benefit from the cue.

So that's what this graph is showing. So the y-axis is benefit. And this is the stimulus onset asynchrony. And so when that gets really, really short, obviously, you can't get a benefit of the cue. It takes some time to kick in. But when you have an exogenous cue, the benefit comes online more quickly than when it's a symbolic or endogenous cue, indicative of some distinct internal process.

So we then transition to talk about evidence that the spotlight metaphor of attention is limited. So we talked about these multiple object tracking tasks where there's some subset of objects on the screen that are identified as things that you need to keep track of. They then start moving. And then at the end, you're queried as to whether or not a particular item was in that original set that you were supposed to track. So we tried this ourselves and then saw some data indicating that people have an ability to do this for four or five objects. It's a little bit harder than if you just have one or two, but you can definitely track multiple things at the same time.

We saw an example of auditory attentive tracking, which might be necessary in cases where you're listening to multiple concurrent sources that are occupying the same approximate space. And you might have to track one of them over time. And we think that might be part of what happens in the cocktail party problem. And so we saw evidence that you can measure attention and attention to a thing by testing whether people are better at making a discrimination about the thing that they're attending.

And so in this particular case, what you're discriminating is the presence or absence of the vibrato on one of the sources. And so what the results show is that for the source that you're tracking, your vibrato detection is better than the source that you're not tracking. So that gives you some way of actually testing the extent to which you've successfully attended to a particular thing. We noted that you can attend to features, right? So you can attend to the red things, or the blue things, or the horizontal things, and you become aware of their organization.

We then talked about visual search where you have to detect some target that's maybe defined by features, and saw that there are some visual search displays where the target is very easy to detect. We say that it pops out. So we talked about how we typically measure performance in a visual search task by plotting the reaction times of how long it takes you to determine if the target is there or not, versus the number of items on the screen, the set size, OK?

So these cases where the target so-called, "pops out" are indicated by results graphs that are flat. So the reaction time doesn't change with the number of items in the display. So these are typically situations where there's some unique value of an elementary feature. But then we also saw cases where the things don't pop out, right? And so a classic example of that is where the target is defined by a conjunction of features. So the target here is blue and vertical.

And so in these cases, the results of the visual search experiment where you measure reaction time for displays of different sizes, look something like this. The reaction time grows with the set size. So the slope is not zero. It's got some number of milliseconds per item. So we saw a whole bunch of different examples of searches like that.

And so the classical explanation for these slow searches is that the target is defined by a combination of properties, and so in order to detect the target, you have to combine those properties. And the combination is by hypothesis not automatic. And so one of the functions of attention is to facilitate the registration of the different properties that make up the target. And as a consequence, these searches are serial. So you have to move your attention around the display each time. Whatever is in the focus of attention gets combined. And then you check whether that's the target. And so it's slow. All right? And in particular, the time that it will take you to detect the target scales with the number of items. So you can measure kind of how fast the spotlight moves around.

And so then we noted that visual search in the real world is often inefficient because a lot of the things that we in practice want to look for are defined by combinations of elementary features. And so the study of visual search has got a lot of real world relevance, both in terms of stuff that we do on a daily basis, but also some very important tasks, for instance, like finding abnormalities in like a radiology image if you're a doctor, or if you're working for TSA and you're looking at pictures from those scanners and you're looking for particular kinds of things that might look like a gun or whatever, that's essentially a visual search task, so lots of practical relevance for this kind of thing, and lots of interest in understanding what it is that causes problems for people and what you could potentially do to make it easier for people.

And so the theoretical framework that was offered back in the early '80s to explain these kinds of phenomena was known as feature integration theory put forth by Anne Treisman. And the key idea is that there are distinct features that are represented separately, often posited to be in maps, that are registered to the retinotopic location. And then attention has this special function of allowing you to register the different features across locations, so combining things and binding them together.

And one of the pieces of evidence for this notion came from the finding of illusory conjunctions where if you in part divert attention, in this case out to the periphery by asking people to pay attention to the numbers and then query them on the stuff in the middle, people will often make conjunction errors. So they'll miscombine the color in the letter, so they won't see some color that's not there. They won't tend to see some letter that's not there. But like they might say a green x or a red t. OK?

And so this was taken as support for the idea that one of the things that happens when you're paying attention to something is that you're able to bind distinct properties of stimuli together, and that when attention is diverted, that binding is disrupted in some way.

And so as I said last time, these days, it's hard to find people who are stalwart believers in feature integration theory for a bunch of different reasons. One of which is that the simple division between things that pop out and things that don't has proven to be more complicated. There's examples where you can get pop-out for properties that seem pretty high level and very distinct from the elementary features that were the cornerstone of the theory. But there's nonetheless, like the set of phenomena that kind of inspired that theory remained very real and things that I think demand explanation.

And so that includes illusory conjunctions. It includes these differences between feature searches and conjunction searches, and so forth. All right. That's where we left off last time. Before I proceed, are there any questions about any of that?

OK. All right. So to continue our tour of interesting phenomena related to attention, we're going to talk about the possibility that when you pay attention to something that improves the spatial resolution of the representation. And the idea behind this set of study and some others that are kind of like it, really goes way, way back right to the notion that rooted in introspection, that when you're really attending to something you can see it better. And that's kind of the whole reason why we pay attention, is somehow the processing is kind of improved.

So the question is, in what respects do we see things better and can we actually measure that? And this is one example that tried to get at this. And I also like presenting this because it relates to some other things that we've talked about in the class, in particular texture segregation. So here we have a texture display. There's a little patch where the line segments are oriented differently from the rest of the texture. And you're able to detect that. And so this experiment makes use of this texture segregation task.

Now, one of the peculiar things about this particular task is that unlike most other visual tasks, this task, texture segregation, is not easiest when the target is presented at the fovea. So for most kinds of things that I would ask you to do, you will do best at that task if you fixate the thing that you're being asked to detect, or discriminate, or whatever. But for texture segregation, that's not the case. You actually are best at texture segregation if you look a little bit to the side of the thing that you need to detect.

And so the presumptive explanation for this phenomenon has to do with the dependence of the receptive fields in your visual system on eccentricity. So you remember how one of the core organizational principles of the visual system is that resolution changes with eccentricity? What's eccentricity? Yeah?

AUDIENCE: Distance from the fovea.

JOSH MCDERMOTT: Distance from the fovea. Yes, exactly. Eccentricity, distance from the fovea. So when you're at the fovea, the receptive fields tend to be small. And then as you move away from the fovea, they get larger. OK? And that's true in the retina. It's true in the visual cortex.

OK. So receptive field sizes increase with eccentricity. If you think that the way that you solve this task is by using the outputs of filters, like oriented filters, Gabor-like filters that are processing this image, and if you suppose that those filters can vary in their scale, well there's going to be like a size of the filters that's optimal for this particular texture, where the size of the filter is about the same size as the elements in the texture.

And so the idea is that there's an eccentricity at which the receptive fields are just right for that particular texture. And in many cases, if you're foveated it, the filters will be too small, and so you don't do quite as well at the task. So that's the premise of all of this. So that's just an interesting fact about texture segregation.

So now in this experiment, they wanted to test what the effect would be of drawing attention to a location. And they're going to utilize the fact that performance on this task varies with eccentricity in a way that's plausibly dependent on the size of the receptive fields to then investigate how cuing attention to a location might change the size of the receptive fields.

So here is a schematic of the task. So this is time. You got these different frames that people would be shown. So there's a fixation dot. And the organization of the trial, so this is one trial and a trial is defined by you giving one answer in response to the sequence of things. And so there's two intervals to the trial. All right? So you get a first cue and a first display followed by a brief mask, and then a second cue and a second display followed by a brief mask.

You have to have the first display and the second display, and only one of those is going to have what's called the target. So that's a little segment of the texture that's oppositely oriented to the rest of the texture. That's the thing you have to detect. So your job when you do this experiment, is to look at this and say, the target was in the second display in this case. Or maybe it would be in the first display in some other case. So it's first or second.

Now for both intervals there is a cue, which is a little thing that pops up. It's an exogenous cue. It draws your attention to a particular location. And so the cue is always valid in the sense that when the target appears it's at the cued location. But in one of the intervals there isn't going to be a target. All right. So you get the cue and then a display, another cue and then a display. And so you can see in this case, the cue is at the same spatial position as that target. OK?

So the idea here is to compare what happens when you are cued to the target's location, when you're not. So I already told you that when you're not cued, when it's just a vanilla texture segregation task, this task has this funny property that performance is best not at the fovea, but at some intermediate eccentricity. And so the question is whether that is altered by the presence of the cue. And so if you think that maybe somehow the attentional cue is serving to improve spatial resolution by shrinking the receptive fields, then you might expect things to change.

All right, before I go on, any questions about the way that this experiment works?

So here are the results. All right. It's a little bit complicated. So let's walk through it. So for the time being, just look at the one on the left, in this particular panel, panel a. So the graph is plotting percent correct, so it's how well you do on the task, as a function of eccentricity. So this is 0. That's the fovea. And then this is moving out to greater eccentricities. OK?

So for the moment, just look at the open symbols. So that's the case where the cue is neutral. So the cue is not telling you the location of the target. And so in this particular case, what happens is that performance is best at these intermediate eccentricities like four or five degrees. So this is the phenomena that we just talked about. So when the target appears right at the fovea, you're worse than when it appears at these intermediate eccentricities. And then when it gets too far out in the periphery you're again worse.

So this kind of unusual trait of the texture segregation task. So now the question is, how is that curve altered by being cued to the location of the target? And so that's what the red curve shows you. So the red squares here are showing you performance on this texture segregation task is a function of eccentricity when you're cued to the target's location.

Now the first thing to note here is this. So in this particular case, you're being cued to the location. And the cue makes you worse. So that's pretty weird. So normally we think of attention as actually making you better. But this is a case where the cue pops up, and that actually impairs your performance.

But the other thing that you can see from this is that if you look at the overall shape of the red curve, like the eccentricity at which your performance is the best has actually shifted a little bit towards the periphery. This is real data, so it's kind of jaggedy. But if you think about smoothing out those two curves, the peak of the red curve is more eccentric than the peak of the white curve. And that's consistent with the idea that when that cue pops up, the receptive fields get a little bit smaller. And as a consequence of that, the eccentricity at which the receptive fields are best matched to the texture moves out to the periphery a little bit, so pretty wild.

So what's up with the other two columns? The other two columns are a manipulation of the viewing distance. And the idea behind that is that when you move closer to the display or further to the display, the size of the texture elements on the retina is changing. And the consequence of that is that one would predict that the eccentricity at which the receptive fields will be optimally matched to the texture is also going to change. And that's more or less what happens.

So you can see that when you're close and you're far, those curves have different peaks But in all of the cases, in all three of those viewing distances, you see the same kind of characteristic, which is that close to the fovea being cued to the location makes your performance worse. And the overall peak of the curve for the red condition where you're cued shifts out to the periphery.

All right. So this is consistent with the idea that attention is altering the receptive fields, effectively shrinking them, which maybe in general, like for lots of things, might be helpful. In some sense, it improves spatial resolution, maybe. But in this particular texture segregation task, this leads to performance decrements when the filters are already too small for the texture, namely at the fovea. OK?

So the texture segregation task is being used here as a diagnostic to understand what's changing in the visual system when you get cued, when your attention is drawn to a particular location. And so this is a really unusual result, both because it's one of a-- I can't think of any other cases. There's probably some, but I can't think of other cases where attention actually makes performance worse under some conditions. So that's really unusual and interesting.

And then also it's a notable result in the sense that it provides some suggestion of how attention is actually altering visual processing. All right. So this is another experiment. This is actually from the same group. This is the lab of Marisa Carrasco, who is an expert on visual attention at NYU. And this was an attempt to ask, again, something about the extent to which attention changes the representation of things visually, but in this case, in a way that might really affect subjective appearance.

And again, it has its roots in speculation from over 100 years ago where people would observe that when paying attention to something, there was a sense in which things look clearer. The question is, what does it mean for something to look clearer? And so here they're asking whether attention has the effect of altering perceived contrast, in other words, making things appear to be higher contrast than if you weren't paying attention to them. So again, they have a clever design to get at this.

So here's the sequence of things that happens on a single trial. So you start out, you're fixating. OK. And then there's a cue. And the cue can either be neutral which means that it's at fixation, or peripheral, meaning that it will be at a location of a subsequent stimulus. So then you get the stimulus display. And the stimulus display consists of two gratings here. And so the task here is to judge the orientation of the grating that is higher contrast. And then you press a key to indicate what the orientation is.

So here they're not actually interested in the judgments of orientation. The judgments of orientation is just a way to get people to report on the thing that's higher contrast without explicitly getting them to make a contrast judgment. So in this case, one grating is fixed in contrast. That's going to be called the standard And the other is varied across trials

All right. And so what they're interested in is whether being cued to the location here will make people more likely to choose the thing that's not actually higher in contrast. So the idea is that when you're paying attention to something, the perceived contrast gets boosted, then you're going to erroneously potentially choose the one that is potentially actually lower contrast.

So here's the results, guys. Let's walk through this. OK. So the y-axis here, so just focus like on this particular panel. So the y-axis here is plotting the proportion of trials on which people report that the contrast that they perceive of the test is greater than the standard. So the standard is the one that's always fixed and the test is varied. And the x-axis here is the contrast of the test. So what do we expect? Well, we expect that when the contrast of the test is really low, people will always choose the standard. And so this should be close to 0.

As the contrast of the test increases, well, once it becomes higher contrast than the standard, you would expect that people would choose the test. And so you end up at 100%. So you get this psychometric function that's going to go from 0 up to 100.

Now the point at which the curves have a value of 50% on the y-axis is what we're going to call the point of subjective equality. That is the contrast of the test stimulus that effectively looks the same as the contrast of the standard. And so we expect that when the cue is neutral, the point of subjective equality should be actual equality. It's just when the test is going to be the same contrast as the standard. So the black curve is what happens under the neutral cuing condition. And so you get some point of subjective equality that is very close to the actual contrast of the standard.

Now the interesting thing is what will happen for the blue and the red curves. So the blue curve is where the test location is cued. So that little dot pops up at the location of the test. And the red curve is where the standard is cued. The dot pops up at the location of the standard. That's my advisee. So what happens here?

So let's look at the point of subjective equality. So that's where these curves hit 50%. So for the blue curve the point of subjective equality is a lower contrast, right? So that's when the test is cued. And so the point of this is that when you cue somebody to the location of the test, it can be a little bit lower contrast and look equal to the standard. And when in fact, they're equal in contrast.

When it's cued, people are more likely to actually say that the test is higher in contrast than the standard. Whereas when the standard is cued, the point of subjective equality is shifted to something greater. So that means you've got to crank up the contrast of the test a little bit to get it to look the same as the standard. All right. So the inference here is that when your attention is involuntarily drawn to some location, the perceived contrast at the location is a little bit higher, which might have the consequence of making something look clearer, or crisper, easier to see.

This other graph here is just the exact same thing at a different spatial frequency. So CPD is cycles per degree, so this is essentially just a replication. And then the bottom curves here are a control experiment where this interstimulus interval, or ISI, so that's the time interval between the cue and the stimulus. So in the graph that we were just looking at, that was pretty short, like 50 milliseconds. And in this control experiment, it's extended to 500 milliseconds. And there you can see the effect goes away.

And so what that means is that whatever happens when that attentional cue pops up there, is transient. So you get a temporary change to visual processing. And then it goes away. OK, so this is a pretty cool result. It's confirming this intuition that when you're paying attention to something, things are kind of easier to see, and suggesting that one correlate of that or one consequence of that is that things look higher contrast.

What questions you got about that? Yeah?

AUDIENCE: I'm still unsure what the cycles per degree part.

JOSH MCDERMOTT: Just say it louder? Sorry.

AUDIENCE: I'm just unsure about the cycles per degree part. Is that how many times they show a certain orientation?

JOSH MCDERMOTT: It's just the spatial frequency of the grating. So remember, you measure spatial frequency in cycles per degree. So it's just how many cycles of the grating happen in one degree. And so you just have to set that to some value. And they just tried two different values. It's just a replication to show that the effect seems to be somewhat general. So it's just a property of the stimulus. Yeah?

AUDIENCE: I'm sorry, you might have covered this, but is there an explanation for why [INAUDIBLE] any reason to have more variation?

JOSH MCDERMOTT: Oh, no, I don't think so. Yeah, I look at those graphs and I say, well, there's qualitatively the same effect at the two. But yeah, it's true that it's bigger in one than the other. No, I think there's an explanation for that. Yeah.

So can anybody tell me why did they design this the way that they did? So what's the point of having people do this orientation task? So if really what you're interested in is whether the cue is going to alter perceived contrast, why not just directly ask people about the contrast? Can anybody think of why? Yeah?

AUDIENCE: Is it because it tests the more natural and unconscious way we view the world as opposed to focusing on it? Is the idea that, if we tell them to focus on contrast, it will skew the results to something that's more synthetic than natural?

JOSH MCDERMOTT: Well, I mean, maybe in a sense. What about anybody else, got any other ideas? Yeah?

AUDIENCE: Is it because contrast can be like an ambiguous term, people will slightly interpret in different Ways

JOSH MCDERMOTT: Well, but they are asking them about it in the sense that the task is to report the orientation of the grating that's higher contrast. So I presume they're showing people some examples of high and low contrast, just so they get the hang of that. So hopefully that's not an issue.

But I think there's another thing that could be a reason why they did this. Anybody else have another idea? So I think one reason to do this is that you if the main point of the task is to actually discriminate the contrast of these two things, people are going to pay attention to both of them. So you'll get this cue, but then people will be dividing their attention to look at both of them. And presumably that might weaken the effect. So I think the reason to actually just have them do the orientation task is to mitigate that to some extent. Right?

So this is one of the things that's always tricky about studying attention is that it's very difficult to query the nature of what you're not attending to behaviorally, because if you ask somebody about something, then they pay attention to it, right? So yeah, I think that's the logic behind the design.

OK. This is a pretty interesting effect where attention appears to alter the perception of time. So I'm going to play you this movie. And it's going to show a bunch of stuff on the screen. Things are going to flash up. Everything's going to be the same duration. But you're going to notice that there'll be one thing that appears to not be the same duration as everything else. And that's an illusion.

[VIDEO PLAYBACK]

[END PLAYBACK]

So each of the disks is up for the same amount of time. And the one that expands is also up for the same amount of time. But people normally say that it looks like maybe 50% longer or something like that. Right? So you can actually measure this. So you can do experiments with this, where you vary the relative durations of the thing that expands and the static ones, and find the duration at which things look to be the same. And then you can calculate what they call a temporal expansion factor. So 1.5 means that the expanding thing looks like it's about 50% longer than everything else.

And so that was true across a range of durations here. That's the thing circled in red. So the oddball seems about 50% longer. And there is this weird thing here that is unexplained that at very short durations, you get what looks like an opposite effect. I don't know what the explanation is of that and neither do the authors. But this is not about things that are expanding seeming longer. So if you do the kind of opposite design, so you have a series of things that expand, and then the oddball is something that doesn't expand, you get about the same effect, maybe a little bit smaller. But again, it's the thing that is unusual or distinct in the sequence that ends up looking longer.

And so the presumptive explanation of this is that the thing that is unusual or deviant attracts attention. And that causes the thing to appear longer. Now is the attention that gets attracted to the eyeballs, is that the same kind of attention that we've been studying all these other experiments? That's totally unclear. So we always use the word attention to apply to all of these things. And again, it's very much unclear whether we're talking about the same thing across experiments. But it's a phenomena that is often described in terms of attention.

Similar things also happen with sound. Maybe it's a little bit weaker, but the lower line here is the same kind of effect with sound, where in this case, I'm pretty sure I think these were beeps that occasionally changed in frequency.

So some of you may have been in a car accident at some point, or encountered some other thing that caused things to seem like they were slower for some brief period of time. It's very commonly reported. People will say that in these situations where they're like highly aroused, that the experience of time seems to slow down. And so this phenomenon is believed to potentially be related to that kind of everyday experience that people sometimes report. So when your attention is drawn to something, it seems to last longer.

So it's an effect that's been documented. We don't really have a great understanding of what happens or why. There's lots of open questions. Does the fact that things seem longer mean that you're doing more computation during that time? Is the sense of time actually maybe related to some kind of mental work? Is time supermodel? We don't know. These are all kind of interesting things that would be fun to explore in the future.

You can also select things with your attention in time. So you can be paying attention to something and have a sequence of stimuli occur and to have to pay attention to one of them. And you can either have some temporal expectation of where that's going to be, or you can be looking for a particular kind of thing. And there's some very interesting and well-known and robust effects that characterize our attentional selection in time. And one of them is what I'm about to show you here.

And this arises in a setting that's shown here, where there is a sequence of stimuli that pop up on the screen. In this case, they're letters And your job in this experiment is to look for white letters or X's. So here's a white letter. And you would press 1 key. And then here's an X And you would press a different key. And the effect that I'm going to show you relates to the effect of your having detected the white letter, which is known as T1, on your ability to detect the secondary target known as T2. And that turns out to be a function of how separated these things are in time

All right. And so what this graph shows is the percentage of the second target that's detected, given that you correctly detected T1, which is that white letter OK? And it's plotted as a position of T2 in the list relative to T1. So 1 here means that the X comes right after the white letter, 2 means that it comes two letters after, and so forth. So look at the red curve here. This is like what somebody like me would do. And what this is showing is that in the few positions following the white letter, people are impaired at detecting the black X. And this is called the attentional blink right.

It's like there's a refractory period on your ability to select or detect stuff that you're looking for. Now this particular graph, it's got a second line on it, the blue curve, which is people that spend a lot of time playing video games. And I think maybe a lecture or two ago, somebody asked me some question and I mentioned how there's a body of work indicating that people that play lots of video games have slightly different visual systems, and in particular, their performance on a lot of visual attention tasks is different, usually better. So in this particular case, the attentional blink is reduced in people that play lots of video games, I guess because you have to do lots of detecting of things in rapid sequence when you play a video game.

This is like the metaphor, again, we don't really have rigorous mechanistic accounts of this, but it has been very well documented. But this is the metaphor is that there's this first fish that's swimming by. And you've got to detect that. And that involves doing something to it. You grab it with your net and then you empty it. And it takes some time to empty the net. And during that time when you're emptying the net, the second fish comes along. And if the net hasn't been emptied yet, you don't grab it. So there's something that happens when you're detecting stuff, and it takes it a little while to reset. So it's called the attentional blink.

So this also happens for sound, behaves very similarly. So it's the same kind of task where in this case there's a sequence of sounds. And I can't remember what distinguished the different sounds here. But there's two different types of targets. And again, it's the exact same graph. So this is the proportion of the second target that you correctly detect, conditioned on having detected the first target. And when it follows that in short succession, you're worse than when it comes some period of time afterwards.

So that's the attentional blink. And it refers to the dynamics of attention over time. And again, is what we're talking about here when we use the word attention, is this the same thing as when we're talking about exogenous and endogenous cuing? I think it's not clear. Why you might ask? So well, why do we use the same word to talk about this stuff? And that's a good question. It probably has its roots in the colloquial usage of the word attention, where we often talk about the fact that we're paying attention to something.

A lot of these tasks have the common character that they feel effortful. So volitional attention, often you feel like you're trying to do something. But whether or not it's really the same thing happening in your head, not totally clear. OK? All right. Questions about the attentional blink? Yeah?

AUDIENCE: What does [INAUDIBLE]

JOSH MCDERMOTT: So the black line is like a control where the T1 didn't actually occur. So this is showing how good you would be if you hadn't detected the first target because it wasn't there. And this is just showing that you're worse relative to that. But then that difference diminishes and maybe is eliminated by the time you're at position 8 or whatever. Yeah.

So another of the many fascinating phenomena related to attention is what's known as change blindness. So how many of you have encountered change blindness in other settings? OK, a handful of you. Yeah so this is a pretty well-known effect. And it again, has a ton of real-world relevance. So we think that magicians are utilizing our susceptibility to change blindness all the time. And change blindness refers to the fact that people are in many contexts, surprisingly poor at noticing certain large-scale events that go on around us.

So this was initially discovered in experiments using saccades. So people noticed just by accident that if you make a change to a display while people are in the midst of making an eye movement, so moving your eyes from here to here, if during that brief period of time you make a big change to the display, people just often wouldn't notice. And that got this large research program started. And much simpler demonstrations have since been developed.

So I'll just show you a few examples of this. So we're going to alternate between two images OK. And there is something that is different about the two images. And I want you to raise your hand when you see what's different. Here we have two hands, three hands, four, five, six. If we keep doing this, eventually everyone will see it, right. But it takes a very long time. OK, so just to speed things up, here I'm going to put the two images at the same location. And everybody can see that there's that notebook thing that has been taken off of the desk. OK?

Here's another case. So again raise your hand when you see the change that's happening between the two images. So we still have about half the class that hasn't seen it. So I'll just point to this. And then everyone says, oh, because once you attend to it, it's very obvious.

Here's another one. Raise your hand when you see it. Some of you are still missing it. Look at this. Yeah. All right. OK. How many of you have seen this before? Yeah, just a couple of you. So there's some of cheesy music that accompanies this, which is beside the point, but it's a different kind of demo where there's a very, very big change that is going to be made to this that happens kind of gradually over the course of the video. And let's see if you see it. If you've already seen this and know what it is, don't say anything. All right?

[VIDEO PLAYBACK]

[END PLAYBACK]

So did anybody see anything change? OK, a couple people. So note that what's the color of the sweatshirt of this guy at the start.

AUDIENCE: Green.

JOSH MCDERMOTT: Green, good. And what's the color at the end?

AUDIENCE: Blue.

JOSH MCDERMOTT: No, it's turquoise. And then now they're back in the starting position. So what happened here is the color of the sweatshirts morph. All right, and then by the end, they've rotated around the circle, right? So at the start, this guy is that turquoise blue. And this guy is green. And then they slowly change the colors. Until the guy in front is blue and the person to the right is purple. And then they switch around places

So this is a case where, again, the change is made very continuously. And if I just go like this, well they're dancing so it's kind of hard to get them exactly the right spot. But it's a pretty salient change, right? And you can see that the brown person, the guy with the brown shirt is turned into green. These are big changes. But then they're done continuously and you don't see them.

OK. And then there's also these really famous examples where stuff like this happens in real life. So there's this legendary study that was done where on a college campus, there were experimenters who would stop random people on campus and ask them for directions to some building. And so they'd be talking to them for a brief period of time. And during that conversation, there would be a couple people who would walk by with a door, a big door, and they would walk in between the person who was being asked for directions and then the experimenter.

And while the door was between them, this person would hop onto carrying the door, and the person carrying the door would take this person's place. So the person that this unsuspecting participant was talking to would change. And the finding is just that a lot of the time, they just don't notice that the person that they're talking to has changed. So there's lots and lots and lots of examples of stuff like this. OK.

And it's just quite surprising because you look out at the world, and you have the sense that your visual perception is very rich. I feel like I can see all of you, right? And these experiments suggest that actually there's lots of stuff that we're either not seeing, or not remembering, or not representing in some way.

So the standard explanation of change blindness is that you need attention to see change. So without it, observers are change blind. And so the intriguing implication is that this impression that we have of rich sensory representations is largely illusory. So we think that change blindness phenomena suggests that the actual sensory representations that we're representing in our heads are fairly sparse, and that relatively little is retained over time.

All right. So in everyday life, we usually see change. If we didn't, it would be a problem. OK? But what enables this typically is usually when things change in the world, there's some low-level transient in the image. There's a motion signal. So something moves or something flickers. So if we go back to this example, you see the book disappearing because there's this big thing that causes flicker in the image. And so what prevented you from seeing that was having the image displaced so that you didn't have that local signal to take advantage of.

And so all of these examples that I've shown you of change blindness, something is interfering with attention being drawn to the change. So there's flicker. So in those examples where we alternate between the two images, there's a blank screen in between. And so that prevents these local motion signals that would normally be created. Sometimes there'll be occlusions that get put into place. And in the case of the color, it's that the change is very, very slow. So you don't get this rapid change that would normally be detected by your visual system.

So if something interferes with the drawing of attention, then attention doesn't go to the change and you won't see it. So in those examples where the images were alternating back and forth, like the one with the plane engine missing, so we had waited a little while and a bunch of people hadn't seen it. And so then I just point to the engine and then of course, everybody sees it. Because that draws your attention to the thing. OK?

So another important thing to note is there doesn't seem to be much accumulation of built-up representation. So we could repeat those demos of the pictures alternating back and forth and let you look at the first picture for a while first. So you could look all over the image and then start the same demo, and it wouldn't really make a difference.

It seems that there's this representation that's built up by attention, which typically moves with your eyes. So you look at a particular place, you see the stuff that's there, but it stays built up only as long as attention is at that location. So it seems like it dissolves once attention is withdrawn.

So again, we talk a lot about these loose metaphors in the world of attention. So think of attention acting like a hand. It grabs the stuff, holds it together. You can then look at it and make discriminations about it, but then it moves on, and the hand lets go, and the things go back to not being represented in the same way.

So really interesting question here is that if the conclusion to these observations about chain blindness are correct and we're actually only seeing a few things at a time, why do we have the impression of seeing such a rich world out there?

So I think one of the really deep insights to come out of this line of work is that whenever we need to see something, we look at it. So because we can just look at whatever it is that we want to see, in some sense, there's no need to have rich representations in your head. So this is often called "just in time" vision, another key idea and catchphrase associated with this is the idea that eye movements are cheaper than memory.

So memory is expensive. You have to do stuff in your brain, expend metabolic resources, maybe use up synapses. So there's a cost to actually remembering stuff. Whereas eye movements are plausibly pretty cheap. They don't cost the organism very much. And so if you're living in a world that's stable and this world is pretty stable. Most of you haven't moved around a whole lot over the last 10 minutes. You can actually use the world as an outside memory. You want to know what's there. You just look at it, OK?

And so we have this illusion of seeing all this stuff because anytime that you want to see something, you just look at it, and it's there. So again, another useful analogy is like the refrigerator light, right? So when you were little might have thought, well, the light in the refrigerator is on all the time because every time you open the door the light's on. But in actuality, the light is mostly off and it just turns on every time you open the door. So it looks like it's always on.

So that's the general conclusion from that particular literature which is that you rely on attention in order to detect changes. And whenever you attend to something, you can see that, and represent it, and notice the changes. But there's a lot of stuff that is not represented in the absence of attention. What questions you got about that? Yeah?

AUDIENCE: Is this the representation of our surroundings, something that's trainable at all, or are we just incapable of encoding that kind of stuff?

JOSH MCDERMOTT: Yeah, I don't actually know of good attempts that have been made to alter this with training. My guess is that it would be pretty hard, but I don't know for sure. Yeah. Yeah?

AUDIENCE: Is change blindness limited to the visual modality, or would you be able to replicate this in like an auditory way?

JOSH MCDERMOTT: Yeah, there are some demonstrations of what's called change deafness, where you can play people an auditory scene with five sounds in different locations, and show that they're not great at noticing, one of the sounds disappearing, for instance. But then if you, again, you draw their attention to it, then of course, they hear it, right? So yeah, I think it's probably pretty general. Yeah. Any other questions on that?

So all of this suggests that representations are sparse, that you don't actually remember a whole lot about what you are seeing. And I want to give you an interesting example that points in the opposite direction. So I want you to look at every single image in this array, just one at a time, just briefly 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14. 15, 16. All right. Got them all. So this is a brief gap to avoid image transients.

So now I want to ask you which of these did you just see? So raise your hand if that one was there before. How about that? How about that? How about that? How about this? How about this? How about this? How about this? How about this one? This one? This one? This one? This one? This one? This one? This one? All right. So quite a lot of agreement.

This is the answer key. So you all did very well. All right. And what we just observed is the fact that visual memory for images is spectacularly good. And that had been known to some extent for a long time. But there was very influential work that actually was actually done in this building by Aude Oliva and Tim Brady and others to document this rigorously and with modern standards. Finding that people can remember, a remarkably large number of images just from a single exposure and often with a fair degree of specificity.

So this is an indication that, well, some image properties are very easily retained, just from one exposure. And that seems at odds with these findings of change blindness, where we were making these large changes to images and people weren't noticing. I think the way to reconcile these things is that typically the changes that are made in these change blindness demos, they often don't alter the meaning of the image. So you have that image of the plane, and yeah, the engine pops on and off, but it's still a plane. There's still people. You've got the image of people having dinner, the railing moves around. It's like a visually salient change at some level, but it doesn't really change what we often call the gist of a scene.

And so it may well be that what you're really good at remembering is something related to gist, and there's a lot of visual details that are not really critical to that. And those are things that are actually not really represented. But again, we don't yet have complete theories about this. But it's pretty striking.

So another angle on attention, in fact, one of the very first attempts to study attention was in the auditory domain with selective listening. This was pioneered by Colin Cherry. And this used a paradigm called speech shadowing. So in this experiment you got headphones on. And there are two different speech streams or two audio streams, one to the left ear and one to the right ear. And the task is to repeat or what's called shadow, what's being said in one of the ears. And the two audio streams are independent of each other.

So you might hear this and if this is the ear you've got a shadow, you have to say fruit, morning, planet, door, swan, for instance. And then there's all this stuff coming in the other ear. And the point is that this is a really hard task. It's so hard that you really have to focus your attention on the stream that you're supposed to be paying attention to.

So this is the title page from the original papers from back in 1953. It's called "Some Experiments on the Recognition of Speech with One and with Two Ears." OK. Using speech shadowing, and the kind of key result from these experiments is that when people are successfully performing this task, they're shadowing one of the streams. They recall almost nothing from the unattended stream.

So here's an excerpt from the results section. It says, what factors of the rejected message are recognized? Rejected means the one that you're not like supposed to be listening to and shadowing. "In this series of tests, the listening subjects were presented at their right hand ears with spoken passages from newspapers, chosen carefully to avoid proper names or difficult words, and again instructed to repeat these passages concurrently without omission or error. Into their left ears were fed signals of different kinds for different tests, but each of which started and ended with a short passage of normal English speech in order to avoid any troubles that might be involved in the listener's 'getting going' on the test. The center, major portions of these rejected left ear signals thus reached the listener, while he or she was steadily repeating the right ear message."

OK, so these were some of the conditions in this experiment. The center major portions of the left ear signals were they could be normal male spoken English, which was what was going in the attended ear; female spoken English, so this notably has a higher pitch so it sounds really different; reversed male speech, so same kind of frequency content but without words or semantic content; a steady 400 cycle per second oscillator, so that's a pure tone at 400 hertz, so a beep.

And so after any one of these tests, the subject was asked, did the left ear signal consist of human speech or not. If yes, can you say what it was about or even quote any words? Was it a male or female speaker, and what language was it in? So here's the description of the results.

"The responses varied only slightly. In no case in which normal human speech was used did the listening subjects fail to identify it as speech; in every such instance, they were unable to identify any word or phrase heard in the rejected ear and, furthermore, unable to make definite identification of the language as being English. On the other hand, the change of voice-- male to female-- was nearly always identified, while the 400 hertz pure tone was always observed. The reversed speech was identified as having something funny about it by a few listeners, but was thought to be normal speech by others. The broad conclusions are that the 'rejected' signal has certain statistical properties recognized, but that detailed aspects such as the language, individual words, or semantic content are unnoticed."

All right. So you're able to hear some aspects of the unattended signal, but they're kind of basic and don't include things like language. OK? Yeah?

AUDIENCE: I'm curious. Did they measure all the performance of the shadowing task because they have a start and then like a middle important [INAUDIBLE]. I'm curious whether they know when that change happens, when the shadowing tasks kind of slowed down for a second like to pick it up and then switch back.

JOSH MCDERMOTT: Yeah, I don't think that was explicitly measured. I mean, they made sure that people were doing the task right. And my guess is that if they mess up, they just start over. Yeah. And I should say, I mean, this is a massively influential study. As an auditory experiment, it's pretty complicated because the person's talking at the same time as they're listening to this stuff. And so, I mean, they can probably hear themselves talk to some extent. And what effect that has on things is hard to know. And this has been revisited a few times, but yeah, probably not as much as maybe it should be.

So from the unattended channel you notice large changes in volume, change in the gender of the speaker, not a change in language. So some things seem to draw attention to themselves, these very salient low level properties of the signal. And we think of them as being processed more or less without attention. Other things are not.

All right. Let's talk a little bit about attention in the brain. So in general, when you pay attention to something, it could be a part of space, it could be a particular type of feature, it could be a type of object, we generally observe enhancement of neural activity in the region that is processing the thing that is attended. So this is an example of attention to spatial locations. This is an fMRI experiment. And there are two conditions attend 2, where there are these two locations on the screen that you have to attend to because there's stuff happening that you're going to have to respond to.

Here's a case where there's just one location that's attended. And this location is different from these two. So this is a retinotopic map. So it's a flat map. It's obtained with fMRI showing the primary visual cortex, the left hemisphere and the right hemisphere. This is the foveal representation. And then these are the regions of the visual cortex that are representing the four locations at which things could happen in the experiment.

And the yellow, things are color coded yellow when there is an increase in the attend 2 condition relative to some baseline. And they're colored blue when there's an increase in the attend 1 condition relative to baseline.

And so the locations here, we've got the upper left hemifield and the lower right hemifield. So remember how in the visual cortex things are flipped, right? So you have the left hemifield represented on the right. And they're also up-down reversed, like in the visual cortex. So the upper hemifield, the representation of that is below the representation of the lower hemifield. So that's why you get a blob here and here that correspond to these two locations. And then, in this case where there's just this one location that you're attending, the blob is up here. So in this case, you're getting these spatially specific increases in response to attending to particular locations. So that's one example.

This is another pretty cool example from our own Nancy Kanwisher a long time ago at this point. So in this experiment, people were shown these stimuli here. So what these are, are superpositions of images of faces and images of houses. And so these two types of stimuli, they were chosen because Nancy had identified these two brain regions that selectively respond to faces and houses, so the fusiform face area responds a lot more to faces than to other stuff. The parahippocampal place area, or PPA, it responds a lot more to houses and other kind of landmark-ish things than other kinds of images.

And in the experiment here, people, they had to make discriminations about one of the two stimuli. And in this particular case, either the face or the house moved back and forth along one axis. And so people were instructed to either attend to the thing that was moving or to attend to the thing that was static. Because then they were performing some task on them, probably like detecting whether it repeated compared to the previous trial.

OK. What do these graphs show? They are plotting the responses to these stimuli in these two different brain regions, the FFA and the PPA. So the y-axis here is percent signal change. So when it's higher, that means there's a larger brain response in that region. So there's always both a face and a house on the screen. But people are attending to one of them.

So in the stimuli, either the face could be moving or the house could be moving, and the task was to pay attention to the moving thing or to attend to the static thing. All right. So what happens here? So in the situation where the face is moving, when you have to attend to the moving thing, the response in the fusiform face area is higher than when you have to attend to the static thing. So same stimulus, but you direct attention to one or the other, and the response gets boosted.

By comparison, in that exact same stimulus condition, when the face is moving in the PPA, you see the opposite effect. So there the response is higher when you have to attend to the static than when you have to attend to the moving. By comparison, when the house is moving and the face is static, you get the opposite effect. So there in the fusiform face area, the response is higher when you're attending to the static because that's the one that's the face. And the response is lower when you're attending to the moving one because that's the house.

OK. So the idea is that attention is directed to either the face or the house, even though they're spatially superimposed. So that's what's kind of interesting about this. So you can select one of these two things even though they're at the same spatial location. So attention is directed to one or the other. And the response in the brain region that we believe represents that type of thing then gets boosted.

OK. So big picture here is that attention to something, so part of space or a type of feature or a type of object in this case, generally enhances neural activity in the region that's processing whatever is attended. OK. So the next thing that I have to tell you about is going to involve effects of attention at the single neuron level.

I think I'm going to call it quits for today. And we will talk about that on Thursday. So on Thursday we'll wrap up talking about attention and then get into touch.