[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: So today, we're going to finish talking about motion perception. So we got this started last time. So motion happens when things in the world change position over time. And the problem of motion perception is understanding how we see that motion.

And so we talked about the evidence that there are motion detectors in the visual system in the way of the motion aftereffect. We experienced a motion after effect. And we talked about simple models for how you would construct such motion detectors. So the simplest and most intuitive is what's called the Reichardt detector, where there are spatially displaced inputs that arrive at a downstream neuron with different time delays, such that if there's something that's moving through space, such that the spatial offset matches the temporal offset, the inputs end up reaching the downstream neuron at the same time. And if the downstream neuron is set up as a coincidence detector, then you'll get a direction-selective response.

We also looked at an actual direction-selective neuron and the receptive field that you can see if you measure the spike triggered average of such a neuron. I mean, that consists of, in V1, in this example, of these orientation selective receptive fields that kind of shift over time. And so then we talked about the idea that another way to think about motion detection is that you can think of motion as being orientation in space time. And so to detect motion you really need filters that are oriented in space time. So that's another way to think about this.

And these receptive fields that are measured in actual neurons show this orientation in space time. So remember, one of these receptive fields is like a movie right. So you got the x and the y dimensions and the t dimensions. So it's like a bunch of image frames. But you can project those down to one spatial dimension, just so it's easy to look at. And then you see this orientation, in this case, in x and t.

So then we talked about-- so this is evidence that these mechanisms in primary visual cortex for detecting image motion get oriented inputs. So you could kind of imagine of building this from a bunch of simple cell inputs. And the consequence of them being oriented is that there's an ambiguity in the motion that they signal.

And so specifically, we talked about the idea that a motion detector that's set up like this, it can tell you how much something is moving, in this case, in the horizontal direction. It's not going to tell you how much it's moving in the vertical direction. So they're measuring a component of the velocity rather than the velocity in its entirety.

And so the consequence of that is that if you observe a response in one of these local motion detectors, that response is giving you a constraint line in the space of velocities. So if you think of velocity as a vector quantity, so it's got two dimensions. There's an x component and a y component. And the constraint line means that the data that you observe, so the response that you observe, is consistent with each of these possible vectors.

And what do these vectors share? They share the horizontal component in this particular example. But the vertical component is unconstrained. So each one of these local motion measurements is giving you information about motion. But it's not uniquely specifying the 2D motion.

And then we talked about this idea that if you have multiple detectors that are tuned to different orientations, they will give you distinct constraint lines. And so if both of those detectors are stimulated by the same thing, then you can take the intersection of those constraint lines. And that would tell you the two-dimensional direction of motion that's happening in the world.

So you can think about this in terms of neurons as taking a whole bunch of different local motion detectors, each one of which is oriented, and adding them all up, adding all of the ones up that are consistent with a particular 2D motion direction. So each one of these things has a particular constraint line. And so this particular combination here is all the ones that intersect at this particular point.

So you could build something that is selective for that particular 2D velocity with this particular combination of these simple motion detectors. And if you wanted something that was tuned to some other velocity, like up here, that would give you a different combination. So this is kind of a theoretical idea, suggesting that there are these two stages of motion processing, an initial stage where things are decomposed into one dimensional components, and then a second stage when those are combined, potentially using this intersection of constraints idea that we talked about.

And so then we began to talk about the evidence that these two stages that are postulated on theoretical grounds actually have a physiological basis in the visual system. And the story here is in this area called MT, which is a visual area that's kind of midway up the dorsal pathway. Remember, we think of the visual system as being organized into these two pathways. The ventral pathway mediates object recognition, the dorsal pathway also often called where pathway. But it's also where motion analysis seems to take place. And so this is area MT, just part of the dorsal pathway.

And so we saw how in area MT, virtually all of the neurons are selective for the direction of motion. So they're tuned to direction. So that's one indication that it probably is important in the perception of motion. And I didn't show you direct evidence for this. But I asserted that if you lesion MT, you get deficits in motion perception.

And so when we left off last time was by introducing the idea that you could study this problem using this stimulus, which is known as a plaid. So a plaid is a stimulus that is composed of two sine wave gratings. So here's one sine wave grating. And here's another. And you get the plaid by just adding them together.

And so the idea behind the stimulus is that each of the sinusoidal components will stimulate a different elementary motion detector. And then when the gratings are superimposed, the question is whether or not the outputs of those two motion detectors gets combined, and if so, where in the visual system. So this is what each of those looks like on its own. And then when you add them together, the interesting thing is that you see a single pattern that moves in a direction that's different from the direction that you see with either of the components on its own.

AUDIENCE: Gratings are a little bit more different. Like, for example, if you return colors, [INAUDIBLE], and I guess what might explain those differences.

JOSH MCDERMOTT: Yeah, that's a good question. So the extent to which the two gratings cohere, so look like a single pattern, will depend on the property of those gratings. So for instance, if the spatial frequency of the gratings is more different, so if one of them is very low, and one of them is very high, you'll be less likely to see the pattern direction of motion.

I don't actually know whether differences in color will have an effect. But other things that you might think would be related to the extent to which those belong together do affect this. So it's not like there's some obligatory integration and it's affected by principles of perceptual organization. So yeah.

AUDIENCE: Is there a need for this [INAUDIBLE]?

JOSH MCDERMOTT: Plaid motion. Yeah. So this is a very famous stimulus in the history of vision science. So the point is that this is a stimulus. Perceptually, you see it as moving in this direction of the pattern. But it's composed of these two components. And the idea is that those components are what elementary motion detectors are going to see because they're tuned to orientation.

And so the question, is there a place in the visual system where the outputs of these elementary motion detectors would get combined and respond to the pattern direction? So just to confirm the intuition or to illustrate the intuition a little bit more, so we've got this stimulus here that consists of these two components. And it creates a plaid.

And so if you have a neuron that is selective for the direction of motion, you might, for instance, see tuning like this. So this is a tuning function in directions. This is polar coordinates. So the direction here is indicated by the angle. And then this is a plot of the response of the neuron at each direction of motion. And so the fact that this is way out here for this direction is like an indication this neuron is tuned for horizontal motion that is to the right. So that's a directional tuning curve.

And this is what you would measure with a single grating. That's why it's called the grating response. And so there are two possibilities for what you might get if you show this, a plaid, should you show this neuron a plaid that's composed of these two gratings. So one possibility is that the neuron is going to respond to each of the gratings individually, in which case, there will be this bilobed response function, because what happens is that you rotate the plaid around, and you move it in every possible direction. And then there will be two directions in which one of the gratings is aligned to the tuning of the neuron.

So this is what you would expect, like, say, in area of V1, that there's two directions where you get a big response, where one of the gratings is aligned with the preferred direction, or where the other one is. But the other possibility is that the neuron might be responsive to the plaid direction, so the direction that you actually see when you look at this. And if that was the case, then you would get a single-lobed response that would be aligned to the direction of the tuning that you would get with a single grating.

So those are the two idealized possibilities. And the question is, what do you see? So in area V1, you almost exclusively see neurons that seem to be responsive to the individual gratings of the plaid. So in the way this is evaluated with that same experiment, so you measure the directional tuning with a single grading.

So this is a neuron that prefers this kind of direction that's downward and to the right. And then you measure the direction tuning with plaids. And the directional tuning that you get with plaids has these two lobes. And one of these is the actual response. I think the solid line is the response. And the dashed line is the prediction of the response, just from the response to gratings under the assumption that it just responds to the gratings.

So again, the idea is that these neurons in V1, they're orientation selective filters. And so they essentially just see one orientation component at a time when one of them aligns with the receptive field.

And so this can be quantified by measuring the correlation of the observed response function with these two predictions here, with the prediction of what you would expect if the neuron was just responding to the components, and by contrast, the prediction if it was just responding to the plaid. So you get the correlation between the actual tuning curve and these two predictions. And you can plot that in this plane here. So this is the correlation with the component prediction. This is the correlation with the pattern prediction.

And so the idea is that if this is a neuron that is really just responding to the components, then it's going to be in this region of space because it should be highly correlated with the component prediction, whereas, if it's responding to the pattern direction, it should be kind of up here. And so this is a plot of the results of doing this analysis on neurons in area V1. So each dot here is a neuron. And what you can see is that pretty much all of the dots are in this region here.

And what that means is that the directional tuning curves of V1 neurons to plaids have this bilobed shape that resembles the prediction if they were just responding to the components. So the point is that V1 is doing what you would expect and not doing something super Interesting.

Now the question is, what happens downstream? And so this area MT. And so some of the neurons in MT look just like neurons in V1. So here's an example. So this is the tuning curve to gratings. And this is the tuning curve to plaids. And you get again this bilobed tuning response function.

But here's a case where the direction tuning to plaids looks a lot like the direction tuning to gratings, and in fact, deviates quite significantly from what would be predicted if it was just responding to the components. So this is an example of a neuron that is tuned to the direction of motion that you see when you look at that pattern.

So this is just two examples. This is the result of this analysis, where every dot, again, is a neuron. And although you can see, some of the neurons here are in this lower region, indicating that they're responsive to the individual components, just like in AV1. There's a bunch that are up here in this top region, indicating that they're responsive to the pattern direction.

So this is evidence for this two-stage model of motion perception. There's this initial stage, where you measure responses to 1D components. Those then seem to provide inputs to the second stage, potentially located in area MT, where those responses get combined. And you end up with neurons that are responsive to the pattern direction.

So as I mentioned last time, this was a fairly influential story in the history of systems neuroscience, sensory neuroscience, in that it was a nice example where there was convergence between both perception-- so this stimulus of plaids was introduced and shown to exhibit these interesting properties, these theoretical ideas for how you would measure motion-- and then some experimental evidence where different stages of the brain kind of mapped onto these different theoretically motivated computational stages in a pretty clear way, so a classic success story in systems neuroscience. Any questions about how this works? Yeah.

AUDIENCE: So if the plaid motion response neurons are only seen half of the grading, will they respond partially or not at all?

JOSH MCDERMOTT: By half of the grating, do you mean by only one grading?

AUDIENCE: Yeah.

JOSH MCDERMOTT: Well, that's what's shown here. So this first column is the tuning function if they're only presented with one grading. And so they respond to the single grading.

The key thing that they do is that the direction tuning to a single grading is the same as the direction tuning to the plaid, even though when the plaid is moving in this direction, which is the direction that it responds a lot to when it gets a grating, the individual gratings of the plaid, one of them is moving this direction, and one of them is moving in this direction. So that's what's kind of remarkable about it is that it's doing this integrative operation that's nonlinear. And that gives you this emergent property.

All right, so story so far is that we've got this two-stage process for detecting the direction of motion. There's this area called MT, where neurons are able to detect 2D motion. And what we're going to talk about now is this additional challenge that is present. And so previously, we've been of talking about this challenge that happened in some sense because of the way the visual system was set up. So you start out making these motion measurements that are oriented. And then that has this consequence that you have to combine across orientations in order to get 2D motion signals.

So there's this other problem, though, which is that there's an ambiguity that exists in the world. And so the issue here is that neurons typically are making local measurements. We talked about this idea that the neuron looks at the world through a receptive field that is restricted in an area. And on your homework, on your problem sets, you've done this exercise of looking at images through apertures and verified how things are frequently pretty ambiguous when you just have local evidence.

And so another type of this ambiguity is due to the fact that many things in the world, in particular, the edges of things, are locally one dimensional. And as a consequence, they're inherently ambiguous, their motion is inherently ambiguous, independent of how you measure it. And this is known as the aperture problem.

So here's the idea. So we have an edge here at time 1. And the edge moves. So this is where the edge is in the image at time 2. But if you're viewing it through this aperture, which could be like a receptive field, again, all that you can measure in principle is the component of the velocity that is perpendicular to the orientation. And just due to the geometry here, you don't really know what the component is that's parallel to the orientation.

And so the consequence is that the edge could be moving in any of these particular directions. So this is called the aperture problem. So here's just an example of this. So these three motions, they're all different. But they look the same when you view them through the aperture. And I can show you some actual examples of this that involve motion.

Oh, no. This happened before. Do you remember this from last year? What did I--

AUDIENCE: It's a flash drive plug-in [INAUDIBLE].

JOSH MCDERMOTT: Yeah, no, this is it. But it's just that it's-- oh, there we go. So this is a demo of a simple scene with some moving objects that you can view through different apertures. So this is one aperture. You're looking at this thing, and it kind of looks like it's moving diagonally. And where's my-- there's my mouse.

So this is what is actually there. So everything here is moving horizontally. But the edges are ambiguous. Now not everything is ambiguous locally. So there are certain features that are two dimensional, which are unambiguous. But some things are highly ambiguous. And that's the aperture problem.

So given this issue, that edge motions are ambiguous, how does the visual system determine their velocity? And so there are two main answers that we'll talk about. The first is to make use of unambiguous 2D signals, so like the corner that you saw there to generate an unambiguous solution. And the second one is to combine edge motions across space. So if you have different edges that are different orientations, you can combine those and resolve the ambiguity. So each edge is ambiguous on its own. But together, they uniquely determine a velocity.

So this is a really pretty cool example of a real-life consequence of the aperture problem and our reliance on 2D features. So I think this is at a football stadium in Spain or something, something in Europe. And these are exit ramps to get out of the stadium. And so they're just these big long ramps that are spirals. And so people just walk down the ramps.

But because of the aperture problem, the motion of these edges is ambiguous. And it gets captured by the motion of the people who are walking. And so the whole thing looks like it's spiraling. So what's actually happening there is that this huge gazillion ton structure in the world that, of course, is not moving.

The people are just walking. But because of the smoothness of the edge and the aperture problem, the 2D motions of the people capture the whole thing. And your visual system ends up thinking the thing is spiraling down into the ground or something. So quite amazing.

Now one of the-- so we often rely on these 2D signals. Now another challenge that comes up as a consequence of this is that not all of the 2D motion signals that are present in images are real motion signals in the sense that they correspond to an object that's actually moving in that direction. And one of the main causes of this is that you can get lots of 2D motion signals from occlusion.

So this is the same situation where we have two squares. They're moving horizontally. But the intersection, the place where one occludes includes, another will actually be moving up or down. And so you can see this in that demo that I was just trying to show you.

So yeah. So this thing looks like something that's moving up and down. But it's actually just this. So anytime things occlude one another and move, this tends to happen.

And so you might expect that in order to see things moving correctly, then the process of determining or inferring motion would need to be sensitive to occlusion, the fact that sometimes one object is in front of another. And so this is-- and there's now lots of demonstrations that this is the case. This is one that's quite powerful, where I'm going to show you a stimulus that consists of these two bars, one that's moving vertically and one that's moving horizontally. And when you see both of the bars at the same time, you can interpret the motion in one two different ways, depending on whether there's evidence for occlusion.

So if you see the bars on their own, they look like they're moving separately. One's up and down, and one's back and forth. But then when we put this little frame around them, which plausibly could be occluding the endpoints, you now see the thing as one object that's kind of moving in a circle. And so the plausible explanation is that your visual system supposes that the image motion that's happening here is due to occlusion. And that no longer really has a big impact on what you see.

So this is an example of that. And I should just say that these demos-- hang on a second. There we go. So this is the basic stimulus. And so this is what happens if the bars are just on their own. Hopefully, everybody just agrees, it looks like two different things, right?

We add this frame around them. And most of you probably see a single thing that kind of moves around in a circle. So the thing to emphasize here is that the image motion is exactly the same in these two cases. But you arrive at these two different interpretations, depending on whether there's evidence for occlusion or not.

I just wanted to mention, these demos, these were created by an MIT UROP in like 2002 or something. And there used to be a website that had all of these things. But they were programmed in flash. And flash doesn't work in browsers anymore. So you can download a flash player, and they still work.

But this is just to say that you never know. The things that you do in your UROP might still remain relevant in 20 years. So it's kind of cool. So the point of this is that motion analysis seems to be informed by information about depth and occlusion. Motions that occur that where you have evidence that they occur at points of occlusion tend to be discounted. So if we go back to the outline that I was just showing you, we talked about this idea that the aperture problem can get resolved in two main ways.

One is to make use of unambiguous 2D signals. We saw some examples of that. The trick there is that you need to discount 2D signals that come from occlusion, which means you have to take into account information about occlusion. But then the other way you could overcome the aperture problem is by integrating edge motions across space.

And so the challenge there is that some times, local motions arise from different objects. So if you just take this same display again, with these two squares that are moving in opposite directions, here's the situation. So let's suppose you have a receptive field here and a receptive field here and a receptive field here. So the consequence of the aperture problem is that each one of these measurements is ambiguous. It's not useless. It's not uninformative. But it's ambiguous. And so what that means is it gives you a constraint line.

So with this measurement, you get this particular constraint line. So this local motion is consistent with any velocity anywhere on this line. And this particular local measurement is consistent with a different constraint line. So the intersection of those is unambiguous and gives you the correct direction of motion, which is this horizontal motion So life is good.

But let's just suppose that you, instead, decide to combine the measurement that you make here with the measurement that you make here. So this one, again, gives you that same constraint line here. But this one gives you a different constraint line. And now the intersection of the constraints is vertical. And nothing in the scene is moving vertically.

All right, so this business of integrating information really only makes sense if you know what to combine with what. And one possibility here is that you would make use of form information to determine that. And there's now lots of cool demonstrations of this. This is a classic one. This was introduced by Maggie Shiffrar and Jean Lorenceau.

So again, there are these moving bars. And this is a situation where, when there is evidence for occlusion, that could potentially account for the fact that there would be a single diamond here, just the corners of which happen to be covered. You tend to see these things as moving as a single thing. And without that evidence for occlusion, you tend to see separate things. And that's a super powerful effect. So here's how that works.

OK, so two sets of bars, they look like they're moving independently. Then we introduce occlusion. And suddenly, you can see a single thing that's kind of moving around in a circle. Importantly, there's certainly nothing locally that moves in a circle. So that circular motion that you perceive has to result from combining information across the edges. And you combine it in this setting, but not in this setting.

All right, and so intuitively, one possibility is that this process of integrating the motion of the edges might be linked to contour completion. So a couple lectures ago, we talked about the fact that when things are occluded, there's this process of amodal completion that allows you to sense the presence of one object behind another. And you might think that is happening here and that the integration of the motion is related to that.

And in fact, there's a bunch of pieces of evidence for that. So here's one variant on this stimulus that can be used to test this. So here, the idea is that there's a manipulation of the occluders so that in this case, there's room beneath the occluders for there to be a full diamond for these contours to connect. And the idea is that here, there really isn't room.

And so if you think that really, the process of integrating and combining the motion of the edges is dependent on the contour completion, you might expect that this thing would cohere as a single object a lot less than that one. And that is, in fact, what tends to happen. You can verify this for yourself.

So this is a case where there's room for the diamond to complete behind the occluders. And most people tend to say that they see a single thing moving around in a circle. But if we make those occluders much thinner, that integration probably breaks. I see some people nodding. Is that working for all of you? Yeah, good. If we make them thick again, you're probably able to see a single thing moving around in a circle.

We make them thin, possibly breaks. This is another manipulation. So if we make a small change here that now allows you to see the occluders as these extended surfaces again, people tend to be able to then see a single object moving around in a circle again. Yeah, working? Yeah, OK.

And so these are just demonstrations that you can evaluate from looking at them. But you can do experiments to measure this. These are just bar graphs that plot the proportion of time that people say that they see this, what's called the coherent interpretation, which just means that there's a single thing moving in a circle. And when there's room for the thing to amodally complete behind the surfaces here or here, people tend to report the coherent interpretation. And then when there isn't room, they don't.

Now you may remember, back when we were talking about contour completion, that another thing that affects completion is this property of relatability. So it has to do with whether the contours are positioned so that they can complete. And so here, the four white lines would be considered to be relatable because you can connect them with a smooth contour. And here, they're really not relatable. You have to have a really wacky looking shape in order for those things to all be connected.

And so if you think that the motion integration would be affected by this, you might expect that there would be a big difference in the extent to which you would see coherent motion. Even though the local motions in all these cases are the same. You're just kind of moving them around a little bit. And so here's an example of that.

So here's a case where things are relatable. How many people are seeing this as a single thing kind of moving in a circle? Raise your hand. Most people. In fact, another thing that affects us a little bit is the contrast. So we can even make it a bit lower contrast. Is everybody seeing that as moving in a circle? There's one person who doesn't. But what about this case here?

Yeah, so that causes it to break. Now you may say, OK, well, that's just impossible to see the thing moving in a circle in that case. But then I would say, well, what about this? So we add these little 2D-features to the thing. And now everything looks like it's moving in a circle. So this is a demonstration that it is, in principle, possible to see that as a single shape moving around in a circle. But without that local evidence, you don't really interpret it in that way.

We move back to here, much more likely to integrate that. And so you can do experiments that verify what hopefully you just saw. These are big effects. So they're not very subtle. So again, this is evidence that these mid-level processes of contour completion seem to be intimately related to the interpretation of motion.

Here's just another demonstration of related things, where information about grouping, in this case, in depth, really seems to affect things. So I'm going to show you four stimuli here. These are each kind of a plaid that consists of a couple simplified gratings. This is the basic plaid. And then we introduce this gray thing that can be in front or behind or in between the two gratings.

And so the idea is that when you position this thing in between them, it becomes impossible for the things to be connected as part of the same thing, or much less likely. And that ends up having a big effect on the motion interpretation. So here's this one.

So hopefully, most of you are perceiving horizontal motion here. So now we just add this other thing in back. And you still see something horizontal. Now we put it in front. Still probably looks horizontal. But now, you probably see the two things is moving separately.

So again, the local motion information is fairly similar in a lot of these cases. But the way that it gets interpreted is very differently, depending on the other kinds of evidence that these two sets of lines could be part of the same thing.

And again, you can run experiments to verify all of this. And things work out more or less like you like you would expect. So the big picture here, these are all demonstrations that there are these big interactions between the perception of motion and the perception of form and grouping and shape.

And so we classically think of the visual system as consisting of these two streams. And the machinery for processing motion looks like it's in this dorsal stream. And naively, you might expect that all the stuff for dealing with shape and so forth would be in the ventral stream. And maybe that's true in some way. But there's pretty clearly some very tight linkages. And we don't really know how these perceptual phenomena that I just showed you are actually mediated. But there's got to be very tight interactions.

And so it seems like the visual system tends to infer an explanation of motion in terms of surfaces and objects, again, based on implicit knowledge of the way the world works. All right, any questions about those sorts of phenomena? All right, so another thing that's worth knowing about-- and it's thematically related to what we've just been talking about.

So we just talked about these examples where form information seems to constrain the interpretation of motion. There's also some very famous examples, where motion alone can define form. And so the original demonstrations of this came from experiments where this scientist put little lights on the joints of humans and then filmed them in a dark room. So the idea was that the only thing that was visible was the motion of these a few points on a person's body.

And when you do this, you can see stuff. So the first thing I'm going to show you is this, which is probably not going to look like a whole lot. This is something meaningful. But I played a little trick on you. And this is the one that's probably recognizable.

So you see this, and you immediately see that there's a person walking. And the one that I showed you first is the exact same thing, just flipped upside down. And so there's a presumably a prior that kind of favors upright walking just because that's something you see all the time. So that kind of constrains this.

But the key point is that this is, again, ill posed. There's lots and lots of possible explanations for the motions of these dots. But they are consistent with the person walking. And they cause you to see a person walking. And there's lots of other examples of things like this. So motion can define form.

This is another cool example. So this is a little movie that resulted from running an old-school edge detection algorithm on a movie. And this is pretty interesting, just because each individual frame of the movie is really hard to interpret. But then when you see them all as a movie, you can easily recognize that you got some dogs playing. You can see people in the back, walking along the fence.

But if I like stop it in the middle-- oops, I shouldn't stop it there. If I pause it here, each individual frame here is really hard to interpret. So lots of cases where motion really helps you see shape.

So what I want to turn to now, though, is a puzzle in motion perception. And the answer to that puzzle really comes in the form of a Bayesian model of motion perception. And so this turns out to be a pretty nice application of Bayesian theories of perception that we've talked about repeatedly over the course of the class.

So we talked about this idea that you can take local motion measurements and combine them according to the intersection of constraints. There's lots of cases where the local motion measurement gives you a constraint line. And to get the true 2D direction of motion, you would combine the constraint lines. And so those constraint lines, those could come from different 1D measurements that are made in the visual system in V1. They could come from different 1D edges of objects in the world, situations where you've got the aperture problem, same issue.

And so this is a case where what you see is actually not the intersection of constraints direction. So that's really the puzzle. There are these instances-- so I should back up and say, a lot of the time, the direction that people do see, like when you're looking at a plaid, is given by the intersection of constraints. So historically, what happened was plaids were introduced as a stimulus for studying motion. This was back in the early '80s.

And there were these initial demonstrations that the direction of motion people see when they look at plaid is typically the intersection of constraints direction. And then everybody got excited. There were like a million different experiments on plaids. And people started to realize that there are conditions where sometimes, what you see is actually not the direction it's implied by the intersection of constraints.

And so this is a stimulus that exhibits this, one of several. So these are rhombuses that come in different variants. And the idea is that the edges of the rhombus are each locally ambiguous. And each one gives you a constraint line. So for the thin rhombus, you get these two constraint lines. And the intersection of the constraints is horizontal for the case where they're moving horizontal. In this case, you get these two different constraint lines. But again, the intersection of the constraints is still horizontal.

But in some conditions, the direction that you see is actually closer to something that's pretty far off of the intersection of constraints. So I'm going to show you that. And the thing that we're going to manipulate here is the contrast.

And so the contrast is being manipulated here because that helps to control the degree of ambiguity of the local motion information. And so this thing is always going to be actually physically moving horizontally. And in some cases, like now, you probably see it as moving horizontally. But we are going to manipulate some things, making it thinner and thinner and thinner.

And you'll actually start to see the thing moving diagonally and downwards. So even though it's physically moving horizontally, you're seeing it as moving in a different direction, moving downward. Does that work for people? Yeah, OK. Good. So that's the puzzle. Yes.

AUDIENCE: I was just going to ask if you can make it dark to see if the effect changes at all.

JOSH MCDERMOTT: My pleasure.

AUDIENCE: Oh, wow. That's cool. Yeah.

JOSH MCDERMOTT: So it's actually moving horizontally. Yeah, thank you for asking. OK, so that's the puzzle. And as I said, there were a lot of demonstrations that nature. And so the question is, how can we understand that?

So to try to make sense of this, I want to go back to thinking about perception as a probabilistic inference. So this is a slide that you have seen several times. So remember, we think of the task of perception as inferring the most likely state of the world, given the sensory data that you observe.

So really, the quantity that we're interested in is what's called the posterior. So that's the probability of this variable here, which indicates something about the state of the world, given an observation. That's code. So the idea is that there's lots of possible states of the world. You observe O, and you want to figure out the most probable state of the world, given O. That's a general framework for thinking about perception.

So Bayes' rule tells us that this posterior probability is equal to this. So there's three terms here. There's the prior. So that's how likely the different states of the world are in the absence of any evidence. Then there's the likelihood. And that's the probability of the observed data, given different hypotheses. So that's how consistent the observed data are with what you would expect, given different possible states of the world.

And then there's a normalization factor, which is the probability of the observation. And for a given observation, that would be fixed So then this is an example of how different possible perceptual interpretations can vary in the prior and the likelihood. So you have, for this observation, this particular image, the idea is that we intuitively think of this as being a good explanation of the image.

There's a background with these two colors and then a letter. But you could also account for the image with this state of the world, these two objects. One looks like that and one looks like that. And that perfectly accounts for the data. But it's a priori very unlikely. So the prior is very low here. But the likelihood is perfectly reasonable.

On the other hand, in this particular case, the prior probability of this state of the world is perfectly fine because it's like an intact letter and a reasonable looking surface. But it doesn't account for the image data. So the likelihood would be low. So a good explanation is something typically that's got high prior probability and also high likelihood.

So that's the key idea is that the posterior the thing that we're trying to maximize were somehow optimized when we're doing perceptual inference involves these two quantities. So that's the general framework. How can we think about this in the context of motion? Yeah, all right.

So I want to walk you through this in detail, just because it's a nice concrete application of these ideas that otherwise can seem abstract. So remember, there are these two ingredients in Bayesian accounts of anything that kind of matter, the likelihood and the prior. And those get combined to yield the posterior.

So let's talk about the likelihood. So remember, the likelihood is the probability of the observed data, given a hypothesis about the world. So we're dealing with motion perception. And so now the data is going to be an image sequence. In the simplest case, it's just two images, frame one and frame two. And something moves between frame one and frame two.

The hypothesis space, because we're dealing with motion, is going to be velocity. So we're going to think about-- we're going to pretend that there's only one thing moving in the world. It's defined by one velocity that is a two-dimensional vector. So it's a point in a 2D space. There's an x component of the velocity and a y component of the velocity.

So here's our stimulus. And we want to look at the likelihood for one particular measurement that's being made on that image. And in this case, it's going to be an edge of this moving diamond where that circle is.

So the image sequence is what's ever observed, which in this case, would be two frames of that moving diamond. For a given image sequence, the likelihood is a function of the unknown velocity. So we're going to imagine, all right, for every possible velocity, how probable is the observed image sequence through that little aperture?

And so what you get here for this particular stimulus is this. What is that? Well, it's a constraint line. And what this tells us-- and I should say that the gray level here represents probability. So the likelihood is the probability of that image sequence, given a velocity, which is a point in that 2D plane.

And so what that is saying is that for all velocities that are on that line, the probability is pretty high. You can also see that the line is kind of blurry. So what that means is that as you move off of the line, there's this graceful degradation in the probability of the velocity. So there's a range of velocities. So if you're right on the hot spot of the line, you know the probability is pretty high. And then you move off of that, and it goes low and then goes down to something close to zero.

And so what that is saying is that there is this family of hypotheses that are all consistent with the image sequence. And that's because of the aperture problem and the intrinsic ambiguity of the local evidence there. So now, what's manipulated here-- and this is actually because of this projector. This is not going to be very clear to you. But what you're supposed to see as you go from this to this to this is that diamond becomes lower contrast.

So I think, in fact, maybe you can't see this at all. But in fact, there's a low contrast diamond here. And there's an even lower contrast diamond here. And the consequence of decreasing the contrast is that the sensory evidence becomes noisier. So there's always noise in the measurement process. And when the contrast is lower, the image data is less clear.

So when you increase the noise or decrease the contrast, what happens? Well, this is what happens to the likelihood function. So we still have this line here. But the line is fuzzier. So because the image data is not as diagnostic, there's now more hypotheses that can account for the image data reasonably well. And if you decrease the contrast a lot, this thing gets really fuzzy.

So that's the likelihood at one location. So if we're going to do Bayesian inference, so we're going to try to infer the most likely velocity, given this observed image sequence, well, we have to take our likelihood and multiply it by the prior. And in particular, we also have to take the likelihood at different locations and combine those.

And so this is how this works. So we've got our image. We're making measurements at a bunch of different locations. Again, the assumption here is that there is a single thing in the world with a single velocity. So we're just inferring this one velocity. So we've got likelihoods at different locations. This is the high contrast case where the likelihoods are these very crisp, thin, not very fuzzy constraint lines.

So we've got one at this location, one at that location. And then we've got a prior. And so what we're going to do is multiply these things together. And that's going to give us the posterior. So the posterior is, again, a function of velocity. It's telling us the probability of different velocities, given an image sequence.

So what happens? Well, in this particular case, when the contrast is really high, the likelihood is very peaked. And so the consequence of that is that when you multiply the likelihood at these two different locations, you essentially get a dot. You get a point, because these are really thin lines. And so it's pretty much just this intersection of the constraint lines where the probability is significantly above zero.

So then you multiply by the prior. And the prior, in this case-- and we'll get back to this-- is just assumed to be this kind of Gaussian blob here. And you, again, get this one point that's kind of significantly above zero. Now things get more interesting, in this case, where the contrast is lower. So when the contrast is lower, the sensory evidence is not as good. It's more ambiguous. So there's a bigger family of velocities that's consistent with the observed image sequence.

And so now we get these constraint lines again at the two locations. But they're now much blurrier. They're more spread out in space. So now what happens is that when we multiply the two likelihoods at the two different locations, you get a blob, a pretty good sized blob. And so now, when you multiply by the prior, the prior actually has a much bigger effect.

So the blob that you're going to get from multiplying these two things, again, would be centered at the intersection of the constraint lines, which would be horizontal. But it's a big blob. And so now you multiply it by this thing that's peaked at the origin. And the estimate gets pulled towards the origin. So the posterior now has a peak that's pretty far off of the true horizontal velocity. Yes.

AUDIENCE: Does it matter that this is a very thin diamond shape? Or if it were a thick line, would the effect not work anymore?

JOSH MCDERMOTT: So what does matter here is essentially, the relationship between the constraint lines. And essentially, yeah, I mean, so as you change the orientation of that stimulus-- and we saw a demonstration of that in a little while ago-- what you see changes. And that's in part because the relationship between the constraint line changes. And so the way that this whole thing works out is different.

But I mean, there's nothing magic about this particular shape. It's just, what's actually doing the work here is the fact that the constraint lines are not that different. And essentially, what happens is when you multiply the constraint lines, you get this blurry thing like this. And that means that when you multiply by that, you get a point that's down here.

But if they were differently shaped, then the product of these two things would look different. And it would interact with the prior in a different way, essentially.

So the key thing that is doing the work here is this idea that when contrast gets low, the sensory evidence is more ambiguous. And than the prior exerts a bigger effect. Now the other key part of the hypothesis that's being tested here is the prior. And we haven't talked about that so far.

And so what's being proposed here is that in our heads, we have a prior that favors slow speeds. So this is a Gaussian that's centered at zero. So that means that when speeds are slower, close to the origin, the probability is higher than when speeds are faster. So that's a hypothesis. And I haven't motivated that hypothesis at all. But that's being proposed as something that is consistent with the observed data. And we'll see we'll see some evidence for that.

So other evidence for that, and a prediction of this, is that if you just take a regular stimulus, and you reduce the contrast, it should look slower, because what happens when you reduce the contrast is that things become more biased by the prior. And if the prior actually favors slow speeds, things should look slower. So again, I'm sorry about the projector here. The contrast of the projector is not so good. This looks OK on my screen. But maybe, when I make this thing move, you'll be able to see it.

So if you look at the center of fixation-- yeah, you can see this thing. So you have a very high contrast grating on the left and a low contrast grating on the right. They are physically moving at exactly the same speed. But if you are looking in the middle, the high contrast one probably looks like it's a fair bit faster than the low contrast one.

And so that's just a fact about motion perception. So if you take patterns, and you make them lower contrast, they look slower. And that's consistent with the idea that there's a prior that favors slow speeds. And then when the sensory evidence kind of gets poor at low contrast, you're biased by the prior.

And so this is a graphical depiction of that principle. So here, it's shown in 1D, just because it's a lot easier to look at graphs in that setting. So we've got the prior, the likelihood, and the posterior. They're both functions of velocity. In this case, we're just looking at one dimension.

And so the hypothesis here is that we've got a prior that favors slow speeds. So that's what this thing is. It's this bump that's highest at the origin. We've got a stimulus that's got some particular velocity. And so there's a likelihood that's centered on the true velocity of the stimulus. But then when we multiply the prior and the likelihood together, we get the posterior. And so you can see the posterior shifted ever so slightly in the direction of where the prior has the highest value.

However, because so this is a situation where the contrast is high. When the contrast is high, the likelihood is very peaked. There's a narrow range of velocities that are consistent with the stimulus. And because it's very peaked, it's not really influenced very much by the prior. It just shifts a little bit. Whereas, in this case, this is the low contrast case. So as a consequence of things being low contrast, the likelihood is now very fuzzy. It's very spread out. There's a very wide range of possible velocities that's consistent with the observed image data.

And so now, when you multiply by the prior, the posterior gets shifted way over from where the likelihood peak is. So the prior has a much bigger effect at low contrast. So again, a general principle of the way these Bayesian frameworks work is that when the evidence is degraded or worse, the prior tends to have a bigger effect. And so things tend to get pulled towards the mode of the prior.

And so that's essentially the proposed explanation for these otherwise funny features of human motion perception, that in some cases, you actually don't perceive the intersection of constraints. And so the explanation is that you're in a regime where the sensory evidence is fairly ambiguous. And the prior has the ability to pull things towards its mode. And that causes you to see a velocity that is slower than the true velocity. And in these situations, that actually causes the direction of the velocity to be different from what it actually is.

And so this was a whole thing about 20 years ago. And there was an extensive set of experiments showing. So this is like a manipulation of the angle of the rhombus, which has this particular effect on the perceived direction of motion. So you go from seeing the horizontal direction to seeing something that's kind of diagonal. And this is the model prediction. And it matches up pretty well with what people do.

So the take home message here is that this is a canonical application of Bayesian theories of perception. So the proposal is that when you are seeing motion, your visual system is estimating the most likely direction of motion, given the image evidence. And in a Bayesian framework, we can think of that as the product of the prior over the velocity times the likelihood.

When the image data is very high quality, so high contrast, the likelihood tends to dominate. And you'll just tend to see what's at the peak of the likelihood. When the quality of the image data is worse, the prior has a bigger effect. Things get pulled towards a slower velocity. And then we saw other evidence for that prior, in the sense that when you take any pattern, and you and you lower the contrast, it tends to look slower. OK, questions about that.

So what I haven't told you here, and what's missing from this story is why we have this prior. So there's a lot of cases in perception where the prior is more clearly linked to properties of the world. So for instance, like the fact that edge orientations are more likely to be vertical and horizontal, because of gravity, essentially. So that's another example of a prior that you can link to the world.

In this particular case, the prior is less obviously related to the world. It's actually hard to measure. So it's more like it's a hypothesis that was proposed because people realized that it would explain a whole bunch of data. It's been built into models. And it does explain quite a lot of data. And so that's of the evidence that we have this prior in our head. It's not as clearly related to statistics of the world. So that's the piece of the puzzle that's a little less satisfying than it could be.

So the last thing that we'll talk about today is the fact that motion is also very closely related to depth perception. And so there's lots of these really cool classic phenomena and illusions that are related to this. And I'm going to give you a bunch of these. And again, this can seem a little bit like a laundry list. And part of the reason that I'm giving you these is that I think they are potential inspiration for your illusion labs. So they're good things to just know about. But I don't expect you to memorize every single one of these.

So we talked about these examples where you can perceive form from motion, like with the point light walkers. But there's also a lot of really interesting cases where you see 3D shape from motion. And so this is a pretty simple display, where you can take an array of random dots. And you move each row of dots left or right with a sinusoidal motion.

And if you vary the amplitude of the motion appropriately, it can be consistent with a 3D surface that oscillates kind of left and right. So you can watch that. And you're probably able to see that-- you can see this as a corrugated surface that's moving around. But of course, it's also consistent with something that's just kind of wiggling around in a 2D plane.

It's ill posed. There's lots of different ways to explain this in terms of shape and motion. People tend to assume interpretations that involve rigid objects. So that's a prior that seems to be exerted in these kinds of cases.

Another related examples is this thing called the kinetic depth effect. So here, this is a display where there's a bunch of random dots. And each one just moves sinusoidally back and forth. And the consequence of this is that the motion-- again, I don't know how-- can you see this at all? Maybe I can turn down the lights a little bit. Yeah, it's funny. It's very clear on my screen.

So you look at this thing, and just obvious that it looks like a sphere that's moving around. The bottom half is the dots are not moving. But each one of these things is just rotating sinusoidally. So again, you tend to interpret the motion as rigid objects. The direction of motion here is ambiguous. So if you look at it for a long time, the direction will sometimes flip.

Here's another one. This has these pretty striking interactions between motion and depth. So this is called the stereo kinetic effect. So a long time ago, before there was the internet and MP3s and stuff, people would listen to music on vinyl records. So you'd have a record player. It actually have a platter. It would spin around.

And so mostly, the function of the record player was supposed to be to help people listen to music. But you can also do other things with it. And so people realize that you could create patterns like this and put them down on the record player and then look at them and amuse yourself in this way. And so this is a pattern that is just going to rotate around in a circle. But when you look at this thing, you perceive this three-dimensional structure.

So again, you tend to interpret these patterns as these rigid objects that are moving around in particular ways. So again, it's ill posed. There's a lot of different ways to account for the image motion here, in terms of things in the world. And you tend to arrive at a particular interpretation by imposing a prior that favors rigidity.

Here's another cool example, the same sort of thing. So this one's got a lot of different interpretations. If you stare at these things for a while, you'll find your visual system flipping back and forth between different interpretations. Not all of them are completely rigid. So the big picture here is that there's these complicated joint inferences of motion and depth. Yeah, question.

AUDIENCE: I feel like a well-known reason for why we see multiple things if we keep staring at it, wouldn't it make sense that once we see something, we stick with that interpretation, unless something were to change?

JOSH MCDERMOTT: Yeah, so the question is, why is it that sometimes these things are bistable? And I think the presumptive answer is that the situations where things are bistable are those where the posterior is actually multimodal. So there's multiple good explanations that both have reasonable prior probability and that explain the data.

And it tends to be-- the world is very rarely bistable. You tend to see this one thing. But like it tends to be these kinds of stimuli, which are a little bit impoverished. They're super clean, black and white. So things are just a little bit more ambiguous. And so there tend to be multiple good explanations.

And so the question is, OK, well then why not just stick with one? And well, there is evidence that you do kind of stick with one. So the first thing that you see, you tend to see for a little while. But some people have postulated that the bistability is actually related to sampling.

So if you think there's a distribution in your head of the possible interpretations, then maybe perception is sampling from that. That's a proposal. But yeah, I mean, I think it's not-- I would say we don't have a definitive understanding of that, other than the fact that it tends to happen a lot when you have slightly impoverished stimuli like this.

So another respect in which motion interacts with depth is in the way that the depth layout of the world and our self motion creates optic flow patterns. So they're characteristic patterns of motion that are induced when we move around in an environment. And this helps us figure out where we're going.

If you're moving in different ways, you could get a flow pattern like this, or this if you're moving horizontally. Or if you're going to crash into the ground, you get a flow pattern like this. Here's just an example of that, where the motion makes it look like you're headed down into doom. And you can do experiments to actually verify that the visual information that you get from optic flow is important for posture.

So here's an experimental setup, where you got a person who's standing in a room. But it's an unusual room, because the walls are detached from the floor so that the walls can be moved independent of the floor. And so the consequence of that is that you can induce this image motion by moving the walls a little bit towards or away from the person. You can induce this image motion without introducing any vestibular cues that you might think would also play a role in posture.

And so you move the walls a little bit towards the person or away from the person. And the idea is that there are these optic flow patterns that would get created based on what the person is looking at. And then what the arrows here are indicating is that the person subtly adjusts their posture.

So when the room moves toward them, you get this optic flow that, well, it's consistent with the room moving toward them. But the natural assumption is that you're actually falling forward a little bit. So then the person shifts back on their heels to try to compensate. And the same thing if it's the reverse.

And so supposedly, you can do the same experiments with children. And you get this horrible result where the kid falls over. So these are not my experiments. So optic flow is used to calibrate posture, even in children.

And so we talked a little bit about this area about area MT. And you can see, so this is, again, the dorsal stream of the parietal pathway. Here's MT. And there's some downstream areas. And if you move one step downstream from MT, you get to this area that's called MST. And that contains neurons that are selective for optic flow patterns.

And so this is a paper from Duffy and Wurtz of an example of neurons. So A, B, and C are different neurons. And so A is one that's selective for planar motion. B is one that is selective for a circular pattern of motion. And C is selective for radial motion. And so the arrows off to the side are depicting the motion flow field of the stimulus. And then this is a plot. The graphs here plot the response over time in response to the stimulus.

And so the leftward motion here produces a big response, whereas, all these other ones don't. The circular motion counter-clockwise produces a big response. Whereas, all the other ones don't in this particular example.

And then this particular inward motion gives you a big response. So you do see sensitivity to these motion patterns downstream. And so these are presumably-- the tuning functions here are presumably the result of neurons that are getting input from a bunch of different local motions at different positions that collectively form that pattern.

Last thing that I want to talk about in the context of motion perception is eye movements. So we've talked about how eye movements are necessary to direct the fovea to image regions of interest. This is the pattern of eye fixation positions that a person would produce if they were looking at that image of a face. So you can see that people are moving their eyes all around. They look at particular parts of the image more than others.

The reason that we're talking about this in the context of motion perception is that making eye movements causes massive retinal image motion. And so the question is, how do we distinguish the retinal motion that is induced by eye movements from the motion of things in the world? So anytime you move your eyes, there's this huge shift that happens across the retina. And you don't want to mistake that for something actually moving.

So eye movements come in several flavors. There are saccades. So a saccade is a fast movement from one image location to another. We make these things all the time, like three or four times a second. You typically don't realize it. You're just constantly looking around the world.

There are microsaccades. These are small eye movements that occur whenever you try to fixate a point. So you think when you're fixating, your eyes are actually fixed. But they're not. They're kind of moving around just very slightly.

And then there are smooth pursuit eye movements. And so these are made when you track moving image features. You can't really make these intentionally. You can only really make these if something actually moves.

This is a pretty crazy movie that I found on the internet, where this person stuck these things to their eyes. So you can actually see the eyes move. And you can observe for yourself.

[VIDEO PLAYBACK]

- You know what's crazy about having these things in my eyes, right? It lets you see how the eyes actually work. Your eyes don't move very smoothly. Like when I move my head, you can see how my eyes try to track what I'm looking at. But sometimes, you can move your eyes smoothly if it's following something. So if I hold my finger out, and I move my finger, you can see my eyes moving smoothly. But if I try to just move my eyes like that, you can't do it smoothly. Isn't that crazy?

[END PLAYBACK]

JOSH MCDERMOTT: All right, yeah, so those are pursuit eye movements. They are crazy. So every time we make an eye movement, huge amount of motion is induced on the retina. How do we discount this motion? All right, in brief, there are two main ways in which you might expect this would be dealt with. One is that you would make use of proprioceptive signals from your eye muscles.

So in other words, anytime that you move your eyes, you could have a little sensors in the eye muscles that would send a signal to your brain saying that things have moved. Another would be that you would use efference copies from your motor system, so the thing that's actually sending motor commands to your eyes. And you could get a copy of the motor command that would get sent to your visual system. It's called an efference copy.

And so we can test the proprioceptive theory by poking our eye. We're going to quickly do this experiment. And then I'm going to send you home. So close one of your eyes. And position your finger on the eyelid. See what I'm doing?

And so you're just going to give your eye a little poke, gentle, not too hard. And what you should see when you poke your eye is the world moves. So you're moving your eye. You're causing changes in the tension in the eye muscles. And so if this was really something that you would discount with proprioception, you would expect that you wouldn't really see that as motion. So that's evidence that it's not really proprioception.

Efference copies seem to be an important part of the story. This can be tested by paralyzing the eye muscles. So when you do that, you tend to get sick because every time you try to move your eyes, your eyes don't move. And it looks like the world moves. So that's part of the story.

Summary, motion measurements are made by local detectors early in the visual system tuned to orientation and spatial frequency. Local motion measurements are believed to be made in two stages, 1D measurements in V1, followed by 2D measurements made in MT. This has been studied with plaids, using both psychophysics and physiology. The aperture problem is a geometric ambiguity that occurs with the motion of edges. We believe it's solved using local features, informed by occlusion and grouping. The visual system seems to construct explanations in terms of surfaces and objects.

We discussed how motion perception provides a key success story of Bayesian models, using a prior favoring slow speeds. We saw examples of how motion and depth perception interact, how optic flow is important for motor control, and how motion from eye movements is discounted, apparently using efference copies. That's all I got for you. Have a great weekend. Midterm is on Tuesday. I'll see you soon.