LAURA SCHULZ: So what you've heard by now is that the hard problem of cognitive science turns out to be the problem of commonsense reasoning. Our computers can drive. They can perform fabulous calculations. They can beat us at Jeopardy! But when it comes to all of the really hard problems of cognitive science, there's only one organism that solves them, and that's a human child.

And those are problems of face recognition, scene recognition, motor planning, natural language acquisition, causal reasoning, theory of mind, moral reasoning. All of those are solved in largely unsupervised learning by children by the age of five. And those are the things our computers don't do well. And that is largely because the problem of human intelligence, common sense intelligence, is a problem of drawing really rich inferences that are massively underdetermined by the data.

So to make that point really clear, I'm to give you all a pop intelligence quiz, OK? So here's the pop intelligence quiz. Can I see a show of hands, please-- how many of you think that I have a spleen? Can I see a show of hands? Excellent.

How many of you would care-- keep your hands down if you're an M.D. But other than that, how many of you would care to come up here and diagram, for the class, a spleen, and explain what it is, where it is, and what its exact function is in the human body? Anyone? How many of you have met me before? A few of you. But most of you, without knowing anything about spleens, and not knowing anything about me, are nonetheless extremely confident that I have one. So you have a lot of abstract knowledge in the absence of really much in the way of specific, concrete facts.

OK, let me give you another problem. Of course, it's a kind of classic one from-- by the way, for those of you who are curious, that's a spleen. All right, classic-- these problems crop up in every aspect of human cognition, all right? What's behind the rectangle?

AUDIENCE: [INAUDIBLE]

LAURA SCHULZ: Right. You all know. You can't articulate it, but you know. And the fact that there are infinitely many other hypotheses consistent with the data doesn't trouble you at all, doesn't stop you from converging, collectively, on a single answer, which means there has to be a lot of constraints on how you're interpreting this kind of data.

Here's another one. Complete the sentence.

AUDIENCE: Very long neck.

LAURA SCHULZ: Very long neck-- could be, or temper, or flight to Kenya. There are many things that it could be. In this case, it looks like a frequency issue. Oh, well, "neck" is a very common word. Others aren't. But if I said, "giraffes are really common on the African--" you would say "savannah," not "television," OK? So you're using a lot of rich information to take a tiny bit of data and draw rich inferences.

And that is the problem, and the hard problem, of common sense intelligence, and I think, a real dissociation between what we do and what our children do. Human intelligence uses abstract structured representations to constrain the hypotheses and make these kind of outrageous inferences that we shouldn't be able to make. That's all well and good, but where do these abstract structured representations come from?

And there's only two possibilities-- we're born with them, or we learn them. Liz Spelke has already told you a lot about the reasons to think that we are born with many of them, and so true for anything that might be stable over evolutionary time, that might emerge early in ontogeny and broadly in phylogeny. And that's true for many aspects of folk physics, folk psychology, causal reasoning, navigation, number.

But there's a lot of other things you know. Shoes, ships, sealing wax-- basically everything else-- they are not plausibly innate. And so how do you learn them? So Piaget, the founder of our field of developmental psychology, says you build these up from experience. You build them up starting with sensory motor representations very, very gradually. You progress through a lot of concrete information until finally, somewhere around the age of 12, you get to abstract representations of the world.

But it turns out that just like you and your spleens, children have a lot of abstract knowledge before they have much of this concrete information. They know, actually, almost nothing--

even less than you know, it turns out-- about anatomy and biology. But they know all of these kinds of things. Animals have insides. Similar kinds of animals have similar insides. Plants and objects don't. Removing those insides is a bad idea usually. They can go on and on without really understanding anything about anatomy or biology.

And so for you. If I push you hard on all kinds of things-- how does a scissors really work-- you know, most of you would be like, ahh, and [INAUDIBLE] pretty quick, OK? And we know this.

So we have these intuitive theories that seem to constrain our hypothesis space. But it's a really hard chicken and egg problem, because we've said we need these rich abstract theories to constrain the interpretation of data. But how do we learn them if we don't have concrete information? And nonetheless, some of them are learned. We're going to return to that at the end of the talk. And really, I'm going to do that to set up Josh and Tomer, who are going to talk about that.

But first, what I'm going to talk about is-- you know, OK, why am I talking about this as an intuitive theory? Where is this argument coming from? I'm interested in learning. And it's a research program that emerged against the backdrop of two revolutions in our understanding of cognitive development.

One was the infancy revolution. You've heard a lot about it, so I'm going to go very quickly. Babies, it turns out, know a lot more than we thought about objects, and about number, and about agents, and their goals, and their intentions. It's not just infants though. It turns out that very young children, preschoolers, also represent knowledge that is not plausibly innate in ways that are abstract, that are coherent, that are causal, that support prediction, and intervention, and explanation, and counterfactual reasoning in ways that seem to justify referring to them as intuitive theories.

And together, these two revolutions, the infancy revolution and the revolution in our understanding of early childhood, dismantled Piagetian stage theory. There was never a time-- there is never a time-- in development when babies are only sensory motor learners. There is never a time when there is not some level of abstract representation going on.

But fundamentally, neither of these revolutions was about learning per se. And I think I can make that point most clear by pointing to a popular book that came out at the time. The subtitle here is *Minds, Brains, and How Children Learn.* But in fact, if you look in the book, this was a publisher's title. There's nothing about brains. And there's very little about learning. The

titles are what children know about objects, what children know about agents.

And I can say this with great reverence and some authority, because the first author's my thesis advisor. And the book came out in 1999, which is the year I started graduate school. So literally and metaphorically, this is where I began. I began with this metaphor of the child as scientist.

And it's a really problematic metaphor, because science is a historically and culturally specific practice that is practiced by a tiny minority of the human species and is difficult even for us, right? So it seems a really odd place to look for a universal metaphor for cognitive development. But science has this peculiar property, which is that it gets the world right. And if you really want to understand how new knowledge is possible, how you could get the world right, you might want to understand how scientists do it, what kinds of epistemic practices might support learning and discovery.

And the answer to that is both that we do and do not know, which is to say we can say a lot of things about what scientists do. Here are some of them. They'll all be familiar to you. They would be familiar to you if you were a physicist, if you were in aero-astro, if you were in paleontology. These are the kinds of scientific practices that cut across content domains and arguably define what science is, which is to say, if you did all of these things, you couldn't necessarily do science.

Science requires bringing these inferential processes to bear on really rich, specific, conceptual representations of individual content domains. But arguably, if you had all of that rich, specific content knowledge and you didn't do these things, you couldn't learn anything at all. These are the kinds of epistemic practices that seem fundamental to inquiry and discovery. And the argument that I've made in my research program is they are fundamental not just in science, they are fundamental in cognitive development. These are the processes that support learning and discovery.

And there's good evidence that each and every one of them emerges in some form in the first few years of life. And that is because these are the only rational processes we know of that can solve the hard problem of learning, which is exactly the problem of how to draw rich abstract entrances rapidly and accurately from sparse, noisy data.

I said we could characterize these practices both formally and informally. And indeed, as

you've heard, I think, from some of our computational modeling colleagues, for each and every one of these practices, we can begin to characterize something about what it means to distinguish genuine causes from spurious associations or to optimize information gain. But with all due respect to my computational modeling colleagues, and much as we want really simple models-- Hebb's rule, Rescorla-Wagner, Bayes' law-- that would capture it, none of these do justice to what children can do.

Because children can do all of these things. And we don't yet have a full formal theory of hypothesis, generation, inquiry, and discovery. That remains a hard problem of cognitive science. But it's a problem to which I think our theories should aspire. Because there is good empirical data that this is the kind of learning that humans, including even very young children, engage in.

So normally, at this point, what I would do is I'd say, and I'm going to show you a few examples of this from my research program. But the talk I'm giving here is a sort of funny throwback talk in some ways. What I was asked to talk about today was the child as scientist. And that was a research program from a few, few years ago.

And you know, at that point, I was a junior professor. And you know, when you're a junior professor, like when you're a graduate student or post-doc, it's all idealistic science for the sake of knowledge, pure science. And then you get tenure. And it's all grants, and money, and administration, and allocation of resources.

And in the years since, I have started to think not just about the pure science of learning, but about the cost associated with gaining information and the trade-offs between those costs and rewards. So I'm going to take these same practices now, and situate them in a world, a real world, that has certain kinds of trade-offs in how you think about information and talk, along with the child as scientist, about what those costs do and how they are also, in themselves, a source of information about the world. So I'm going to talk about this as sort of inferential economics.

Children know information is valuable. I'm going to show you a couple of examples of how they reason about it. And children selectively explore in ways that support information gain. So I'm going to show you some old work, but I'm also going to throw in a few new studies, because I can't resist.

But information is also costly. And the costs themselves are informative in a variety of ways.

I'm not necessarily going to get through all of these studies, although I might try to. But I want to give you a kind of feel for the kinds of things children can do.

All right, so let's start by talking about a really basic problem. It's a problem that's basic to science, but it's also a problem basic to human learning, which is the problem of generalization. How do you generalize from sparse data? In science, we do this all the time. We have a small sample. We want to make a claim about the population as a whole.

And of course, we can use feature similarity and category membership to say things that look the same or belong to the same kind are likely to share properties. So if some of these Martian rocks have high concentrations of silica, maybe they all do. If some of these needles on Pacific silver fir trees grow flat on the branch, maybe they all do.

But in science, we can also do something a little more fussy and suspicious. We can say, well, you know, it kind of depends on how you got that sample of evidence, right? If you randomly sampled from the population, yeah, sure, the properties, you can generalize. But if you cherry-picked that data in some way, maybe the sample isn't going to generalize quite as broadly.

So do all Martian rocks have high concentrations of silica or just the dusty ones on the surface? Do all Pacific silver fir needles lie flat or just those low on the canopy, right? These are the ones that are easy to sample from. How do I know how generalizable the property is?

And if I think that you cherry-picked your sample, I might constrain my inferences only to things near the ground. So how far you're going to extend a generalization in science depends on whether you think that the sampling process was random or selective. And we wanted to know whether this was true for babies as well.

So this is how we asked. We showed babies a population, in this case, of blue and yellow dog toys. They're in a box. The box is transparent, has a false front so it stays a less stable representation of what looks like a lot of balls. And we're going to reach into that box. And we're going to pull out-- there are many more blue balls in this box than yellow balls. We're going to pull out three blue balls one at a time and squeak them-- and squeeze them-- and they're going to squeak.

And then we're going to hand the baby a yellow ball from the same box. And the question is, does the baby squeeze the ball and expect it to squeak? Well, there's nothing very suspicious about pulling three blue balls from a box of mostly blue balls. And this has a lot of feature

properties in common. So it looks like the others.

We predict that children should generalize. They should try squeezing this ball and should squeeze often. And the question is, what happens if you do exactly the same sample in exactly the same way from a different population? Now, it is very unlikely that you sampled three blue balls from a population of mostly yellow balls. In this case, it's much more likely that you were sampling selectively. So maybe only the blue balls had the property. Yeah?

AUDIENCE: And they can see the population?

LAURA SCHULZ: They can see the population-- transparent box, transparent front. So they can see the population. And if children understand that it's not just about the property similarity but something about how that evidence was generated, then, in this case, children should say, well look, you just looked like you were cherry-picking your sample. Maybe it doesn't generalize predictions that fewer children should try squeaking and children should squeeze less often.

So I'm going to show you what this looks like. By the way, the yellow one has that funny thing at the end so that children could do something else with the ball, right? So they can bang it, or throw it around, or other things like that. So let me show you what it looks like.

Kids are always going to see three squeaky blue balls. They're always going to get a yellow one. But you'll see that they do different things depending on whether they think--

[VIDEO PLAYBACK]

LAURA SCHULZ: --the evidence was randomly sampled and possibly generalizable or not.

So the child at the top is squeezing, and squeezing, and squeezing, and squeezing.

[END PLAYBACK]

So these were-- this was true both of the mean number of squeezes and the number of individual children who squeezed at all. I'm just going to show you the mean number of squeezes. But what you'll see is that children are much more likely to squeeze, and squeeze persistently, in the condition where the evidence looks like it was randomly sampled than selectively sampled.

But what you could worry about here is children are sensitive to something about the

relationship of the sample and the population. But maybe they will just generalize from a majority object to a minority, but not the reverse. Maybe they won't generalize from a minority object to the majority. So they don't really care about whether the evidence was randomly sampled or not, they just care about that aspect of the sample and population.

So we ran a replication of the yellow ball condition. Again, we're going to pull an unlikely sample from that box, three blue balls in a row. And we're going to compare it with a sample that's not that improbable. You could easily randomly sample just one blue ball from the box.

So in this case, children are going to see much less squeezing, right? They're only going to see one blue ball squeezed. And we squeeze it both once and three times, but it's only one blue ball in two different conditions.

And the prediction there is that even though the children themselves are seeing much less squeezing, they should say, well, that's not an improbable sample. And they, themselves, should squeeze more. And that's exactly what we found in both of those conditions.

It's graded, by the way. If you do two balls, they're intermediate. And if it's a model-- well, not going to talk about-- but yeah, what happens if you just pour them upside down and drop them? Now this is a really improbable sample I said, three blue balls from a yellow box. But you've just given positive evidence that you shook the ball, and it just happened to fall out. And they don't know we're MIT, and we can do sneaky technological things like have a trap door.

So in this case, it's an improbable sample, but it was randomly generated. And the prediction is, in this case, the babies themselves should squeeze more. Because as I say, if you can pour out any balls that squeak, probably everything squeaks. And indeed, they do. Indeed, they do. All right?

So 15-month-old babies' generalizations take into account more than category membership and the perceptual similarity of objects. They make graded inferences that are sensitive both to the amount of evidence they observe and to the process by which that evidence is sampled. Is that clear?

All right, let me show you another example of sort of child as scientist. It's going to start with a hard problem of confounding that we all have, which is that we are part of the world. So in one-offs, when things go wrong, we may not know if we were responsible or the world was responsible, right? This is a chronic problem in relationships, right-- you or me? So it's a hard

problem of confounding, and you might need some data to disambiguate it.

So here we're going to give babies a case where they cannot do something. And the question is, can we give them a little bit of data to unconfound that problem and convince them either that the problem is probably with the toy or the problem is with themselves? And the argument is, if they think that it's themselves that's the problem-- it's the agent state and not the environment state-- they should hold the toy constant and change the agent. But if they think it's the toy, then they should just go ahead and reach for another toy.

So in both cases, they're going to have access to another person, their mom. And they're going to have access to another toy. And so the question is, what do they do if we give them minimal statistical data to disambiguate these, OK?

So this is the setup. We're going to show the babies two agents. In one condition, I am going to succeed one time at making that toy go and fail one time.

And Hyowon Gweon, my collaborator on this project, is also going to fail and succeed once. So this looks like this toy has maybe faulty wiring or something. It works some of the time, not all the time. It's just not a great toy. The babies can have another toy. The parents are going to be there. If they think it's the toy, they should change the object.

In the other condition, Hyowon is always going to succeed, which is generally true in my experience. And as is always true of my experience in technology, I am always going to fail. And in this case, the children should conclude that there's something wrong with the person. And if that's the case, they should hold the object constant and change the agent. This is what it looks like.

[VIDEO PLAYBACK]

- [INAUDIBLE]

LAURA SCHULZ: We've lost sound. Well there's an audio here, but [INAUDIBLE] we are going to want that later.

- Cool What happened to my toys?

LAURA SCHULZ: In any case, were just showing them the data at this point. And the babies are going to get the toy in each condition. One toy is on the mat. By the way, there are lots of individual differences

in any two sets of clips between the exact positioning of the parent, the child, the toy. All of these were coded blind conditions for all of these other variables to make sure those were evenly matched across conditions, and they were.

So what you're going to see now though is that in the condition where the babies think it is probably the toy, they engage in a very different behavior than when they think it is probably agent.

- [CRYING]

- [INAUDIBLE]

[END PLAYBACK]

And that's what we found overall. The distribution, overall, of children's tendency to perform one action versus another differed across conditions depending on the pattern of data that they observed. So 16-month-olds track the statistical dependence between agents, objects, and outcomes. They can use minimal data to make attributions. And they help them choose between seeking help from others or exploring on their own. Clear? OK.

I've just shown you kids' sensitivity to the data that they see. But of course, data isn't handed out all the time. At least disambiguating data isn't handed out. One of the really important, hard things that you have to do if you want to learn is sometimes actually figure out what data would be informative and go get it. And that is a really characteristic thing about science. And the question is, is it, in some sense, a characteristic thing about common sense.

So that's what we're going to ask here. We're going to jump to much older children here. These are four and five-year-olds. And we're going to give them a problem where instead of showing them the disambiguating data, we're going to ask if the kids themselves will find it.

So what we showed children-- when you were little, you possibly played with some beads that snapped together and pulled apart. These are like toddler toys. So we gave them these snap together beads. They're each uniquely colored. We place each bead, one at a time, on a toy. And in one condition, every bead makes the toy play music to each bead you put on. And the other condition, only half the beads did. So the only difference between these two conditions is basically the base rate of the candidate causes. One works for every bead, the other only

works for some of the beads.

And then we took all these training toys away, and we showed the children either a pair that we had epoxied together-- it was stuck. We tried to pull it apart, we showed we couldn't. The children tried to pull it apart, they couldn't. It's a stuck pair of beads, or it's an ordinary, separable pair of beads. And then, as a pair, we placed each pair, one at a time, on the toy, and the toy played music.

In principle, this evidence is always confounded. One bead in each pair might be the responsible party activating the toy. But as a practical matter, if you just learn the base rate is that every single bead activates this toy, there's not a lot of information to be gained here. You should just assume that all of these beads work. And in that condition, we expected that kids would play indiscriminately with the two toys.

But in the condition where only some of the beads work, there's genuine uncertainty. Right? Maybe only one of these beads work. Maybe they both did. And if that's true, and if kids are sensitive to the possibility of information gain, only one of these beads affords the possibility of finding out. On only one can you isolate the variables. And that's with the separable pair. So we thought on this condition, the kids should selectively play with the separable pair. And in particular, they should place each bead, one at a time, on the toy.

So that's, in fact, what we find. In the obvious condition, the kids basically never separated the pair. And in the some beads condition, about half the kids did it and performed the exhaustive intervention. That was cool, but my graduate student at the time said, they're doing something really interesting even with the stuck pair of beads. We should look at the stuck pair. And I said, what can they do with the stuck pair? There's nothing to be done with the stuck pair. It's stuck. And she said, well, let's just try it again with the stuck pair.

So we did the same thing. They got introduced the fact that either every bead worked, or only some of the beads worked. And this time we introduced just the stuck pair, and we placed it on the toy, and the toy made music. And let me show you what the children did.

[VIDEO PLAYBACK]

- All right, I'm going to do this one, and then it'll be your turn a little later. But now, can you just watch and see watch happens? All right. This one makes the machine go. How about this one? This one doesn't make the machine go. What about this one? This one doesn't make the

machine go. Let's try this one. This one makes the machine go.

LAURA SCHULZ: She goes over that a second time, and then she hands the child the toy.

- Just a minute.

- Look at that.

[END PLAYBACK]

LAURA SCHULZ: The child plays around, does just what we did. And then she does something we'd never done. She rotates the position of the bead so that only one makes contact with the toy at a time. And if you have a folk theory of contact causality, that is a pretty good way to isolate your variables. And not one that had occurred to, say, the PI on this investigation. But in fact, it occurred to about half the kids again, in that condition.

In the some beads condition, where there was uncertainty, the kids were more likely to design their own intervention to try to isolate the variables than in the condition where all the beads worked. So preschoolers are using information about the base rate of candidate causes to distinguish the ambiguity of the evidence, and they're selecting and designing potentially informative interventions to isolate these causal variables.

All right. I'm going to show you some new work now, kind of on the same theme. One way that investigations can be uninformative is because evidence is confounded. We're all familiar with that, right? We think we did the perfect experiment, then we're like, oh, well, it really could have been because of this really silly, boring reason. And that's disappointing. And we have to run it again.

But another reason that investigations can be uninformative is because they generate outcomes that are super hard to distinguish. So if I have a handkerchief in one pocket, and a candy cane in the other, then a child who wants that candy cane is going to have no trouble patting you down and finding the candy cane. But if I have a pen in one pocket and a candy cane in the other, that's going to be a harder problem.

And this might be more salient to you if I say, you're going to go in and have a lab test for a fatal disease, or potentially a benign disease. And you know what the results are going to look like? One is going to be reddish maroon, and the other is going to be maroonish red. OK, that's not the kind of test you want. You want yellow, blue, right?

So this is important. We care about how much uncertainty there is over interpreting the outcome as well. So if children are sensitive to how useful actions are for information gain, then they should prefer interventions that generate distinctive patterns of evidence.

So Max Siegel in my lab has been running some experiments like this. He started with a very simple one, basically the equivalent of the handkerchief and the candy cane. He said, OK, there's either a bean bag in this box that I'm going to put in here, or a pencil in this box. It is a shiny, cool, hologram, sparkly pencil. You'll want it. That's going to go in this box over here.

And in this box, either the really cool, shiny hologram pencil, or the really boring yellow pencil is going to go in this box. And guess what I'm going to do? I'm going to take each box, and I'm going to shake it. So he does that. And you know what you hear both times? Ka-thunk, ka-thunk, ka-thunk. Ka-thunk, ka-thunk, ka-thunk. Indistinguishable sounds. And now the question is, which box do you want to open? Which box you want to open? Right?

And if you're sensitive to the ambiguity, you should say, well, listen, if it were a beanbag, I would really know, so that must be the sparkly pencil, right? But I'm never going to know in this box, because both pencils are going to sound alike, so I really better choose this box.

And then he's going to do a harder problem. He's going to say, there are eight shiny, colorful marbles-- you really want them-- in this box, or two really boring white ones. Or there are eight colorful, shiny marbles in this box, or six really boring white ones.

In each case, they each get hidden, you're going to hear the box. In each case, it's going to make exactly the same sound, which is actually the sound of eight marbles in a box. And the question is, which box do you want to open? So let me show you how that works.

[VIDEO PLAYBACK]

- [INAUDIBLE] and I also have some marbles. Well, you see these marbles right here, the white ones? These are Bunny's. Oh, six of my marbles, yay. Oh, two of my marbles, yay. Guess what, Taylor? These marbles with lots of different colors, those are yours for right now. That's pretty cool, right? Those are your marbles. That's awesome.

And in this game, I'm going to hide either your marbles or Bunny's marbles inside of this box. And then I'm going to hide either your or Bunny's marbles inside of this box. Does that sound like fun? And then we're going to look for your marbles, OK? If you find them, you get another

sticker. All right, so I'm going to put Bunny's right here, and we're going to do the hide game.

So first, I'm going to choose either your marbles or Bunny's marbles and put them in here. I'm going to pour them in. Look, somebody's marbles are in here. Now I'm going to do the same thing with this box. Either your marbles or Bunny's marbles are going to go in this box. All right. Are you ready to begin shaking and listening?

So remember, over here, there's either your marbles or Bunny's marbles. OK, let's listen.

[CLUNKING NOISES]

All right. And over here, there's either your marbles or Bunny's marbles. Let's listen.

[CLUNKING NOISES]

Cool. Let's do it one more time.

[END PLAYBACK]

LAURA SCHULZ: We'll skip the one more time, but-- oops. Sorry. You can at the general principle. The children are overwhelmingly good at this kind of task, it turns out, in both of these cases and in many, many other iterations they went. They're very confident about which box they should pick, which means they're representing to themselves something about the ambiguity of the evidence, and their own ability to perceive these kinds of distinctions.

So with Max, we've been talking about this as a kind of intuitive psychophysics, where they can represent their own discrimination threshold to make these kinds of distinctions. And they prefer interventions that generate distinctive patterns of evidence and maximize the possibility of information gain. Is that clear? OK.

So there's a lot of ways in which children seem to be using intuitive theories, some kind of abstract, higher order of knowledge, to make inferences from data. But information is also costly to generate. And the costs themselves are informative. So I'm going to talk a little bit about that piece now.

It's costly both for the learner, who cannot learn everything. And in a cultural context, where you're not just learning and exploring on your own, but you're actually getting information from other people, it's costly also for the teacher or for the informant. And these kinds of costs, and

how we negotiate these kinds of costs, I think, are really fundamental to a lot of hard problems in communication and language.

Lots of the field of pragmatics deals with problems of underdetermination. We say these sentences, we understand each other. We understand each other in all kinds of ambiguous contexts. We use a lot of social cues and other information to disambiguate. But part of what we do is, we make inferences about how much this person is going to communicate in this context, given how much I need to understand. And we use this to resolve these kinds of ambiguities.

I'm going to give you a few examples of that here. Now, again, I'm going to start with the study we did a while ago then show you a little bit more recent work. I'm going to skip over some of this just to be able to cover all of this and say,

Because there's a cost of information for both teachers and learners, it predicts some trade-offs in the kinds of inferences you should make. So for instance, if a knowledgeable informant shows you a toy and says, here, this toy has a single function. Then if you, the learner, think that that teacher is trying to generate a true sample from the hypothesis based on what actually is going to get the right idea to your head, you should assume that there is only one function and not two, or three, or four, or five. Because if there were more, they should have shown them to you, right? Because if they know the true hypothesis, and they could just demonstrate that, they can rule out all of that for you.

But if you just stumble upon a single function of a toy, or a not knowledgeable teacher generates it accidentally, or if the teacher, as Liz Spelke pointed out, is interrupted in the middle of that demonstration, then you shouldn't assume that that evidence is exhaustive. It suspends that inference, right?

Now, OK, well, you showed me one, but maybe there are two, three, or four. So it's only in a condition where I think you are a fully informed, freely acting teacher that I should assume, well, look, if you are helpful, knowledgeable teacher, then the information you give me should not only be true of the hypothesis, it should help me distinguish that hypothesis from available alternatives. And what that means is, there's a specific trade-off between instruction and exploration. Because if I'm instructed that there's one function of the toy, I don't need to explore any further. But if I just happen to find one function of the toy, maybe I do.

So let me show you what we did to test this. We had a novel toy. It actually had four interesting

properties, a squeaker, a light, a mirror, and music. And we demonstrated a single function of the toy, the squeaker, in three conditions. And we also had a baseline.

In the pedagogical condition, we said, watch this, I'm going to show you-- sorry, the alignment's off-- but watch this, I'm going to show you my toy. They pulled the toy and then said, wow, see that. OK, the accidental condition was, look at this neat toy I found here, accidentally pulled this tube in the same way, wow see that. The baseline was, just look at this neat toy I have here, with no demonstration. And the interrupted condition was identical to the pedagogical condition, except the teacher was interrupted immediately after pulling the tube, and then she said, wow, see that.

So is that clear? I'm sorry for the slide misalignment. And the prediction is that in the first condition, children should constrain their exploration relative to all the other conditions. So let me show you what that looks like. Or not. Which is too bad, because this is a really super cute slide. But it's not going to work.

In this condition, what we found was this. We found a child in the children's' museum with a toy with all of these kind of wow properties. We say, wow, see this. We show them the property of the toy, and the child spends 90 seconds pulling only the squeaker. He then says, I'm very smart for a five-year-old. And when she asks for all of the other functions of the toy, he doesn't know any of the other functions, because he hasn't explored.

And what we think is, he is very smart for a five-year-old. Because it's a completely rational inference that, if there were more functions of the toy, then they should have been demonstrated. And so what we find overall is, in fact, that children do fewer actions, and they discover fewer functions of this toy. This isn't just true, it turns out, we now know, because we live in a hyperpedagogical culture. Because Laura Shneidman and Amanda Woodward just replicated this study with Yucatec Mayan toddlers and found the same kind of effect, constraints in the pedagogical condition, even though it's a culture that's pretty limited in their pedagogy.

So information is costly, and pedagogical contexts strengthen the inference that the absence of evidence-- a teacher's failure to go on and teach you more information-- is, in fact, evidence of its absence. Is that clear? And this is a very sensible inductive bias, but it predicts that instruction will, for better or worse, constrain expression. Because that's what it's supposed to do. It's supposed to constrain the hypotheses you consider. And indeed, it works quite well.

And that's good if you're right about the world, right? And it constrains it to efficient learning. But it's bad if you're wrong about the world. Because the unknown unknowns, the things you don't know are true that you failed to teach, are going to potentially mislead a learner.

How much is enough information? Well, there are lots of good reasons why teachers ought to provide very limited information. First of all, as I showed you in the first set of studies, evidence often supports generalization. Right? One dog toy squeaks, probably they all squeak, barring how generalizable that sample is. So I don't need to show you every single toy. I don't need to show a child, this is a cup, and that's a cup, and this a cup too, and that's a cup, and that's a cup. Once the child has a cup, I can assume that that child herself will be able to make the rational generalization.

Or sometimes I know you're not going to be able to make it, but the additional information is just too costly. I'm working on teaching you two plus two, I'm not going to teach you linear algebra right now. It's a waste of our time. So that's another reason why you might provide limited information.

So what are the contexts in which omitting information is a reasonable thing to do, and when is it misleading? When is this a real problem? And the answer turns out to be, if I'm the informant, and I know I'm providing information that is going to lead you to the wrong hypothesis, then we consider that a sin of omission. Right? If I'm omitting information, and I'm not doing that, then maybe that's not a problem.

So one of the questions is, can children distinguish these contexts? Can they tell when the teacher is providing too little information, and it is going to cost the learner something in terms of what they can gain, and when they're not. So to test this-- this is again, Hyowon Gweon's work-- we introduced a toy. And it had one function, this wind up mechanism. And the kids got to explore, and they found out the toy did one thing.

In the other condition, the toy looked the same. But in fact, the toy had lots of functions, and the children knew that. So the children always knew the ground truth. The toy either had one function or four. And then there was a teacher who taught Elmo. The teacher always did the same thing. The teacher always taught just one function.

And the first question was, the teacher's always doing the same thing with an identical looking toy. Do the kids penalize the teacher? Do they think he's a bad teacher if he only teaches one function when there are really four, compared to when he teaches one function and there's

only one. So the first thing we did was, ask kids to rate that teacher. And indeed, they think that this teacher is a terrible teacher when there's four functions and he only teaches one. They think he's a good teacher when he teaches one of one.

But the really interesting question was, what would the children do to compensate if they knew they had a bad teacher? So we ran exactly the same set up where the teacher shows Elmo the toy in the one function case and the four function case. And for reasons that will become clear, we also ran a control condition where there were four active functions, and the teacher taught all four. In all cases, the teacher then goes on and runs that experiment I just showed you. The teacher shows just the squeaker toy here of this single function.

The question is, what should the kids do? So it's a complicated set up, so I'll walk you through it a little bit. When you teach one of one function, you should infer the toy probably does one thing, it's a good teacher. And so when they show you one function of this, you should constrain your exploration, say that was a sensible inference. When they teach one of four, you should say, that's a bad teacher. The toy probably does more than one thing, the new toy probably does too. I'm going to explore more broadly.

But we don't know, in that case, if they're doing it because they just saw a toy with one function, and so they think this toy has one function. And they just saw a toy with four functions, so they think this toy has four functions. So we can disambiguate those with this condition. Now if they're just generalizing from the toy, this toy has four functions, well, then they should think this toy has four functions. But if they're generalizing from the teacher, this is a good teacher. So when the teacher now tells you about the new toy, that it has one function, the kids should constrain their exploration.

So does everyone understand the logic of the design here? And in fact, that's exactly what we find. The children compensated with additional exploration when they thought the teacher had provided insufficient information to the learner. Is that clear? OK. So information is costly to teachers and learners. If teachers minimize their own costs and provide too little information, children think they're poor teachers. They suspend the inference that that information is representative. They compensate with additional exploration.

I'm going to go ahead and show you just a couple more examples here. There's too little information. But because information is costly, you can also provide too much information. I might be doing that right now. Too much information is costly. It's taking a toll on the learner to

absorb all of that. And you have to know, well, are you providing me too much information, or just the right amount?

And at the risk of falling into this trap myself, I am going to show you quickly this study. Because how much information is too much information depends on a hard question which is, how much do you already know? Right? You've all been here all summer. You know a lot of things. I'm not a very good estimate of what you know or how much this information is going, so it's a little hard for me titrate what's the right amount of information to give you. And the question is, can children take these kinds of theory of mind problems into account in order to estimate what information they should be getting?

To test this, we give kids a 20-button toy. If I push a single button and it makes music, how many of you think that all the other buttons make music? Because you can generalize from data, and that is a really good inductive inference there. They look the same, it's a toy. One makes music, they probably all do.

But suppose I go on now to show you-- so that's your prior expectation, they all work. But that one doesn't work, and that one doesn't work, and that one doesn't work, and that one doesn't work, and that one doesn't work, and that-- oh, that one works. And that one doesn't work, and that one doesn't work, and that one doesn't work, and that one doesn't work. I'm doing this for a reason. I know this is really boring. It's partly to give you a break, but it is also-- oh, that one works.

OK, so now what you have learned about this toy is that actually, only three of these buttons work. And suppose I show you this across a couple of toys. You've just changed your expectation. Now if I show you that one button works, you don't think all the rest work. You think probably two others work, right?

So if I bring out a brand new toy, and I push this button, you'll probably be relieved if I just go ahead and push these three, right? And I don't go around and show you all of the inert buttons on the toy. Because information is costly. You have to sit there through all those demonstrations.

In this experiment, I'll show you Gweon's work. We gave kids a common ground condition where everybody shared prior knowledge, this abstract theory you can use to constrain your interpretation of this data. In this condition, there were two toy makers, who are the informants, and there's two naive learners, Ernie and Bert. And in the common ground

condition, Ernie, and Bert, and the toy makers are all there while the child explores and finds out that only three buttons work on these toys. OK.

And then one teacher shows of a brand new toy just like this every single button, the inert ones and the non-inert ones. And the other teacher shows just the three working buttons. Right? And then we say, hey, kids, guess what? We have a whole closet full of more of those toys. One of these teachers can show them to you. Which one do you want? Which one do you want, OK?

The other condition is almost the same, but guess what? The child explores on their own, right? And so there's no common ground about what these toys do. And then the teachers do the same thing. One teacher shows exhaustive information, and the other teacher pushes only the three working buttons. So in this case, that efficient information, that less costly simple demonstration could mislead the learners about what the true hypothesis is. They have a prior that all these buttons ought to work.

Which toy maker would you rather learn from depends on whether the learners share that prior background information or not. And that was true not only when the children were judging the informants, but when they were teaching themselves.

So again, these are four and five-year-olds. The children had a condition where Elmo got to see that only three buttons worked on the toy, and the condition where Elmo didn't get to see how many of those buttons worked on the toy. And then the children got to teach Elmo the toy. And the children were much more likely to press more buttons and provide exhaustive evidence in the no common ground condition than the common ground condition. So children themselves are adjusting the cost of the information they provide based on their prior expectation about what they think the learner is going to learn from the data.

What did we do is, lastly, tell you a little bit about how the costs of information are informative not just in figuring out what data to learn from and how you should communicate information, but in actually figuring out what people are doing, and actually grounding out ordinary, everyday theory of mind.

I'm going to start with an example you're familiar with. I know you've seen this before. This is an experiment by Gergely and Csibra, a rational action. This little ball jumps over the wall to get to the momma ball-- you've all seen this? I think Josh was presenting it maybe? OK. And when you took the wall away, babies expect that ball to take an efficient route. So we think that

rational agents should take the most efficient route to the goal, they should maximize their overall utility.

But there's a lot of reasons why that ball might have jumped over the wall, and they all have to do with costs of rewards of action. One might be, it was really hard to get over the wall, but really rewarding to do so. Another reason, though, is that it was really easy to get over the wall, so she might as well, but she didn't care that much about the reward. These could have the same net utility, but psychologically, you really care about the difference.

If we're talking about the internal structure of an agent's motivations, you want to decompose this simple argument about rational, goal-directed action into what the particular costs and rewards are. So Julian Jara-Ettinger in our lab has developed this account he calls the naive utility calculus, which is our way of reasoning about other peoples' actions. Which is, we assume other agents are acting to maximize utility, but we care about the internal structure also.

There are agent invariant rewards and costs. Two cookies is always more than one cookie. Higher hills are always higher than lower hills for all of us. But some of us are more motivated to get cookies than others of us, and some of us find hills more costly than others of us. So in addition to these agents invariant aspects of costs and rewards, there's also these internal subjective things that are harder to judge, right? How competent you are, what your values are, and your preferences.

And so understanding how all of these worked together lets you take a very, very simple analysis and make surprisingly powerful inferences about what other agents are doing. I'm going to show you a few examples and actually connect it back to how children are scientists in this regard.

So here is an example experiment. Here is Grover, and there is a cracker and a cookie down on this low shelf. And Grover goes ahead, and he chooses the cookie. And now there's a cracker and a cookie and this shelf, and Grover goes ahead and chooses the cracker. And the question is, what does Grover like better?

Well, if your read out of preferences, it's your actions you take. It's your goal-directed action. You should be at chance-- you chose a cracker once and a cookie once. But if even young children understand, no, it's not that simple, right? You're not just acting to maximize reward,

you're acting to maximize utility. You have to take the costs into account. Then clearly his preference is what he chose when the costs were matched, right? Not when the costs were mismatched. And the children should say, no, no, no, it's what he chose on the low box. Which treat does he like best? Is that clear?

And indeed, that is what children do. You can also introduce a couple of characters. Cookie Monster, who has a strong preference for cookies, and Grover, who is indifferent, who likes them both. So for Cookie Monster, the reward value of cookies is much higher. For Grover, they're equivalent. And now you can set up a situation where there's crackers in a low box and cookies at a high box. And you say, guys, go on, you can make a choice. And Grover chooses the cracker and Cookie Monster chooses the cracker. And you say to the kids, which puppet can't climb?

Now, no puppet has climbed. No puppet has failed to climb. No puppet has even thought about climbing. But the kids can know the answer, right? Because if Cookie Monster could climb, and he had a high reward, then who would do it. So you don't really know about Grover, but you can make an inference that the costs were not equivalent.

And by the way, in case you're worried that kids are just saying, well, Cookie Monster, I've been listening to Michelle Obama. You know, obesity and fitness-- cookies aren't good, I can't climb. We ran the same experiment with Grover and clover, and Grover really likes clovers, and Cookie Monster likes them both. And you now flip the inference around.

OK, so they can consider how the costs affect the inferences they make. All right, so let's bring this all back together. I've thrown a lot of information-- probably too much information-- at you about how kids can reason from sparse data, about how they use their theories to make these inferences, about how they use this in teaching and learning in social context. But if kids are sensitive to these utilities and the trade-offs, then the kinds of things that I showed you them doing with beads and with machines, they should also be able to do in social context.

We believe psychology to be something of a science, as well as all of these other sciences, and they should be able to apply some of the same principles-- holding some things constant, manipulating others-- in order to gain information. So we basically ask that question of the children here. Can they distinguish agents' different competencies and rewards by manipulating the contexts that they see and gaining information?

So here, we don't know if Cookie Monster can climb. So let's put one treat on each box. Where

should we put the treats to find out if Cookie Monster can climb? Now, only one of these interventions is informative. If the cookie is down low, and you know Cookie Monster prefers cookies, then you're not going to get any information. But if the cookie is up high, then you are going to get information, right? And in fact, the kids are overwhelmingly good at this.

In case you think, well, treats should be put up high-- and this is also true, by the way, for clovers, but it's all right. In case that you think they just have a heuristic like, oh, well, let's put treats up high, you can ask the question a different way. You can say, both of our friends can climb the short box. But only one of our friends can climb the tall box, and we don't know which one. So let's put the cookie up here and the cracker down here. And if we want to figure out which one of our friends can climb, which friend should we send in?

Well, Grover has no particular incentive to climb. He could just do the cracker. But Cookie Monster, he has an incentive to climb. You should probably send in Cookie Monster. And again, these are the kinds of inferences that young children can make. Is that clear? All right. So, end of a long winded talk. I want to return it to this.

So I framed it in terms of the way I've been increasingly thinking, and the projects that we're increasingly moving towards, which is thinking not just about the pure pursuit of knowledge and information, but how do you pursue information in a complex world where it's not just your own individual exploration. You get it in a social context, you get in interaction with others. The information is costly both to deliver and to process. But that, those costs themselves, are information, and you can use them to make sense of the world.

And so I want to sort of bring these together and come back to a problem that I posed a bit earlier, which is, I think, I hope I've made the case that what kids are doing is not reasoning about huge sets of data. What kids are doing is, they're taking some very abstract structure knowledge and using it to constrain their inferences about tiny amounts of data. Cookie Monster this, a couple of trials of evidence. And then they make good inductive guesses, which are sometimes wrong, but they are good.

And I said, these abstract representations, not all of them are innate, right? You've seen beautiful evidence of the many that are, but a lot of them aren't. A lot of them aren't. A lot of the things that govern your common sense knowledge every day, how do you get those? And how do you get those from tiny amounts of data?

This is a problem that bugged me deeply for a very long time, and I think there's been a real

leap and a very exciting ascent of how it is actually possible to use tiny amounts of data to make really rich abstract inferences, which then constrain your interpretation of subsequent data. I think that is a really important problem. And with that, I'm going to turn it over to Josh and Tomer, who maybe can tell you how that is going to actually work. So thanks to everyone here at Woods Hole.