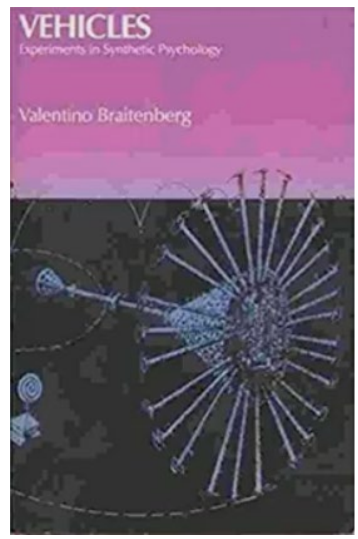
5.5: Artificial Animals

We've been talking about this sort of biological turn in cognitive science and up to now our attention has, quite plausibly, been focused on human intelligence, and I think human intelligence in biological systems, that is people, and also representing human intelligence or modeling human intelligence in computational terms. What I'd like to do today is talk about a kind of different strain of work. It still has, in fact as you'll see, a heavy computation element. This is representative of a branch of cognitive science that, among other things, tries to model not human intelligence, and you might call even not animal intelligence but animal behavior primarily. We're going to use as our focus a kind of classic book by Valentino Braitenberg that was published in the 1980s called *Vehicles*, and we're going to use that as a source of examples.



[1:28] But this is part of a sort of larger community that might expand on the computational side into robotics and also on the modeling side into simulations that often go by the name of artificial life. So, the idea here is rather than focus on particular aspects of human intelligence, what we'd like to do is try to make simple models of, in general, simple animals, simple animal behavior and see what they can tell us about cognition in general. As I say, the book that we're going to use as a focus of examples at least for the first part of this discussion, is a truly wonderful book. It came out in the 1980s, by a physiologist I guess a neurophysiologist, physiological biologist, Valentino Braitenberg. The book was called *Vehicles*. It's a short book, in some ways it's a very easy to read book. I'd certainly recommend, it's still in print. It's still a great favorite among computer science students, especially robotics students. I'd definitely recommend your reading it.

[2:58] The one thing I have to tell you about the book, and you'll see this as you go along, is that it's format is rather special. It seems to be telling a very simple story, but the story gains complexity as you go through the book. And what Braitenberg does in this book is that he begins by asking you to imagine modeling very, very simple animals. And as you're reading along, you think well that's trivial, obviously I could do that, I could make a model of that simple animal. And then chapter by chapter, in the book, he builds on the animals from the previous chapter until he elaborates on those animals so that in each successive chapter you're getting more complex or more interesting behavior and by the end of the book you've somehow built a complete intelligent system, kind of like a human intelligent system, and it's not at all clear where stepwise in the book the tasks went from being obviously doable to conceivably impossible. I mean, it's a little bit like, I've described it in the past as a little bit like the fable, I guess, of the frog that you put into water and then as you slowly heat the water, the frog doesn't notice that the water is getting hotter because it's getting hotter very slowly, until at the very end you could boil the frog and it wouldn't never have jumped out of the water because it never sensed that the temperature had changed by a large enough amount that it was able to leave. I hope nobody's ever tried that. But in any event, reading this book is kind of like being that frog. That is, you're reading along in the book and you're thinking oh that's easy, and the next chapter is only a little harder than the previous chapter, and the chapter after that is only a little harder than the previous chapter to that, and by the end you feel like you may be doing the impossible. It's really a quite strange experience.

[5:24] But again for that reason I think it's a wonderful sort of inspiring book to read. I'm just going to take you through the animals of the early chapters of the book, which are the easy ones to implement. Just so you get a sense for how Braitenberg thinks about modeling animal intelligence. Now the animals that I'm going to show you, you could think of them as being artificial animals on a computer screen, or you could think of them as being simple robots, physical robots that you could make with certain kinds of behavior. And depending on how you prefer to think about it, either one is ok. The problems that arise in creating physical robots are somewhat different, naturally, because you're operating in the physical world than the problems that arise when you're operating on a computer screen.

[6:22] Let's begin with Braitenberg's first vehicle. He calls this vehicle one. And hopefully the slide is large enough so that you can kind of squint in and see what the caption is. But let me just explain what's going on here. The way in which these vehicles are drawn, this is a sort of typical drawing, a little box that represents the animal, that antenna at the top is a sensor of some kind, some kind of sensory apparatus. And for our purposes, you could vary what that sensor is, you could leave it unspecified. It could be a temperature sensor, or a light sensor, or a proximity sensor, or a sound sensor. However you want to think about it. Could be a sensor for chemical gradient. So, it is some kind of sensor. The sort of wheel in the back is a motor. You could think of it a wheel to roll around on the ground or fins to navigate through water or something like that. So this vehicle is just a connection between a sensor and a motor. That's all it is. The sensor as you see is linked to the motor by a line. And the basic idea here is that the more input, the more of whatever sensory input the sensor is looking for, the greater the signal that's being sent to the motor.

Figure 1
Vehicle 1, the simplest vehicle. The speed of the motor (rectangular box at the tail end) is controlled by a sensor (half circle on a stalk, at the front end). Motion is always forward, in the direction of the arrow, except for perturbations.

[8:08] And we'll assume for this first example that the relationship is linear, if there is zero stimulation on the sensor then there is zero stimulation to the motor. If there's a little bit of stimulation to the sensor, then there's a little bit of stimulation to the motor, and so forth. So the sensor is just a direct connection to the motor. Now, just for the sake of specificity, imagine that the sensor were a temperature sensor and the motor were fins in water, ok? The motor manipulates fins in water. Then this would be a vehicle which if you put it into water, if the water is warm, it may start moving along in the preferred direction. Notice that the motor is such that it's going to push the vehicle in a particular direction. If you heat up the water, the vehicle will move faster. If there's a temperature gradient in the water, the vehicle will move slowly in the colder part of the water and faster in the warmer part of the water. If you were to walk into a room, and just see a creature, suppose you didn't know how we built this, sensor linked to motor, and it was just disguised as a little fish or something like that. If you were to walk into a room and see a creature like this in water, behaving like this, your first natural assumption would be that this is alive. It's not an inanimate object, and it doesn't move like an inanimate object in water. Inanimate objects may move in water like little dust motes moving according to Brownian motion and so forth. But Brownian motion is statistical random motion, and this, on the other hand, is motion with a preferred direction. So the vehicle is moving in the direction in which it's pointing. It's a very simple animal. It's an extraordinarily, there's not much behavior to this creature, but you would still kind of think it's a creature. According to the format of Braitenberg's book, in the next step he makes these vehicles, these creatures a little more complex.

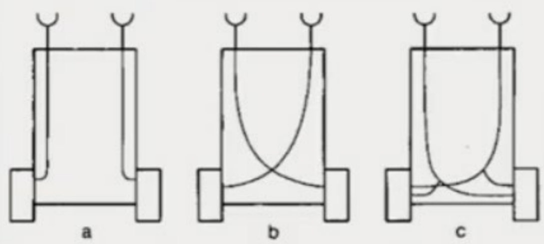


Figure 2
Vehicle 2, with two motors and two sensors; otherwise like Vehicle 1. The connections differ in a, b, and c.

[10:46] So here are possibilities now. What happens if we have two sensors and two motors? Ok. In vehicle A we have sensors connected to motors that are on the same side. So, the left sensor's connected to the left motor, the right sensor's connected to the right motor. In vehicle B, the sensor motor connections are crossed. In vehicle C the sensor motor connections are just doubled. Each sensor is connecting to two motors, ok? But what would happen with these vehicles now? I mean, what would the behaviors of these vehicles be? Well, it turns out that vehicle C, even though it looks the most complex, is not that interesting, it's just a kind of strengthened version of vehicle 1. A and B, however, have interesting behavior.

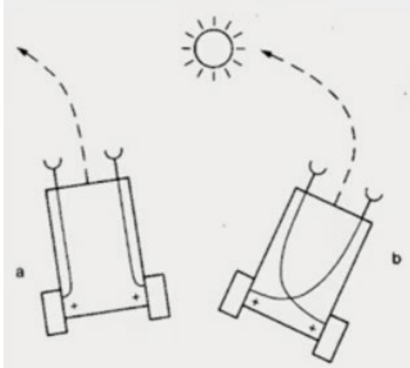


Figure 3
Vehicles 2a and 2b in the vicinity of a source (circle with rays emanating from it). Vehicle 2b orients toward the source, 2a away from it.

[11:48] Here's the diagram from Braitenberg's book. Now the little sun icon you could think of as the source of sensory input. So, for example, if you think of it as a lightbulb, suppose these vehicles were now ground vehicles rolling around on wheels on a floor and the little sun is a lightbulb and the sensors are light sensors. Ok? So in this example we'll switch to ground vehicles and light sensors. Now let's look at vehicle B first, because that's a little easier to talk about. Here's the one where the vehicle motor connections are crossed. Now what happens if a vehicle of that description gets close to a light source? In that case, the left sensor in this drawing, the left sensor of vehicle B is getting a little more stimulation than the right sensor. The left sensor is a little closer to the light source. The right sensor is getting a little less stimulation. What is the result of that? It means the right wheel turns at a slightly higher rate than the left wheel. So the vehicle turns toward the light source until the input to both sensors is equal, when the vehicle is directly pointed to the light source. So what would vehicle B do in a room where it's rolling around on a floor and it's near a light? It would roll until it's, and then turn so that it's facing the light and notice that it's getting more and more stimulation so it's going faster and faster, and perhaps it would eventually sort of just run full tilt into the light and smash the lightbulb.

[13:56] What would you say about this kind of creature? If you came into a room and saw what looked like some kind of creature behaving this way, it rolls along on the floor, when it gets near a light it turns toward the light and runs into the it and smashes it. What you might say about this kind of creature is that it hates lightbulbs, it likes to smash them. But notice that we're attributing that sort of complex concept to the vehicle. We're saying, oh this guy hates lightbulbs. There's no representation of lightbulb inside this vehicle, and there's certainly no representation of hatred. There's just two sensors and two motors. That's all it took. Once you know the engineering of this thing, somehow it demystifies even this very simple creature's behavior.

[14:55] What about vehicle A, how would it behave? Well, it would roll along the floor, it would get faster in the neighborhood of a light, but notice that in this case the, as it gets close to the light if it's not dead on which it very likely would not be, if it's not sort of dead on facing the light, any sort of variation from that as in the drawing here, the right motor will get more stimulation than the left motor and the vehicle will turn away from the light. It'll go fast and then turn away from the light and then run fast and slow down once it's at a distance from the light. So, you might say that this is a vehicle that's kind of afraid of lightbulbs. It gets agitated when it's near them, it turns and runs away, and then it calms down once it's at a distance from them. So it's vehicle A is a vehicle that fears lightbulbs, vehicle B is a vehicle that is angry at or hates lightbulbs. Ok?

[16:04] So next step. This is the next kind of vehicle. Instead of a positive connection between the sensor and the motor, let's now stipulate that we can make an inhibitory connection between the sensor and the motor. So the more stimulation the sensor gets, the slower the motor goes. Ok? Now in that case let's take a look at vehicle A here with inhibitory sensor motor connections and uncrossed wires. That is, the left sensor is connected to the left motor and the right sensor is connected to the right motor. Here the vehicle will be rolling along, it's not getting any stimulation, so it's rolling along reasonably fast on the ground, but as it approaches the light, the stimulation to the motor causes them to slow down. In this case, if you look at vehicle A, the left sensor will be getting more stimulation than the right sensor, which causes the left wheel to slow a little more than the right wheel, so the whole vehicle slows down, the right wheel is going just a little bit faster again until the input is equal, and the vehicle is just facing the lightbulb. And it gets close enough that we can assume that it just

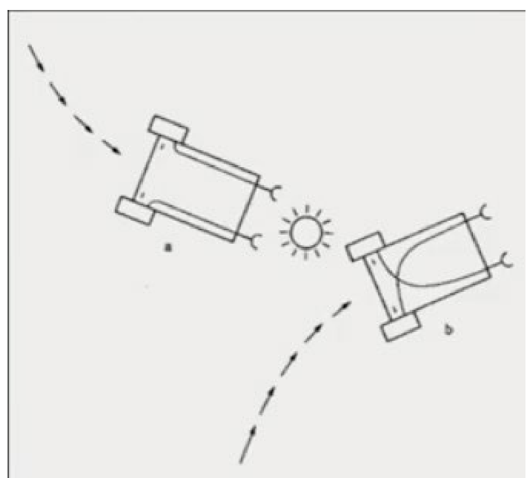


Figure 4
Vehicle 3, with inhibitory influence of the sensors on the motors.

stops. So, think about how this vehicle moves. It’s rolling along, it gets near the light, both motors slow down, but the left motor slows down little more in this case, and the right motor slows a little less. So, it’s slowing down but turning toward the light source and then as it gets close it stops and hovers in front of the light source. One might say of this vehicle that it loves, or it worships the light. Or just wants to bask in the light. Something like that.

[18:14] So what about vehicle B in this case, with inhibitory connections but they’re crossed. Well, in this case, the vehicle slows down as it gets near the light, but the right wheel slows down a little more as you see from the drawing. The left wheel, because it’s attached the right sensor, is going a little faster. So, the vehicle slows down, but it turns away from the light and then as it moves even a little further and further from the light, it goes faster. So, this vehicle sort of slows down, turns, and then starts speeding up again. One might say of this vehicle that it’s fickle. It slows down near the light, but then turns away to look at other things. Or perhaps it’s a curious vehicle, it wants to go explore other… It stops to hover around the light, but it speeds up as it moves away from the light, and perhaps looks for other light sources. So again, you could kind of give personalities to these vehicles. A loves light or venerates light or however you want to put it, and vehicle B, in this case, is curious about light and hangs out near it, but then moves away to look at other things. Ok? So now we’ve gotten up through vehicles with crossed or uncrossed connections and inhibitory or excitatory sensors.

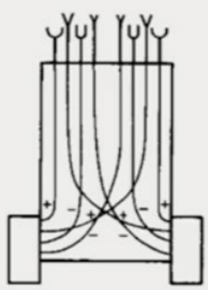


Figure 5
A multisensorial vehicle of brand 3c.

[19:58] Now, of course, we can mix and match these things and start building them into more complex vehicles. So, we could have sound sensors and light sensors and temperature sensors, so you could make a vehicle that loves, you know, loves light sources but attacks sources of sound or things like that. Just by sort of layering on more and more of these sensor motor connections, you could have vehicles that exhibit more than one of these characteristics.

Figure 6
A nonlinear dependence of the speed of the motor V on the intensity of stimulation I, with a maximum for a certain intensity.

[20:37] The next step, we’re heating the frog a little bit, we’re going to make the vehicles a little more complex. The next step would be to abandon the assumption that the connection between sensor and motor is linear. So, for example in this graph on the left, here the horizontal axis is intensity of stimulation and the v, the y axis is response from the motor. So as the intensity goes up from zero, the response from the motor goes up at a sort of steady rate. It reaches a maximum, and as the intensity gets very high, the stimulation to the motor goes down. So, this is neither a purely excitatory or a purely inhibitory sensor motor connection. In fact, it’s in many ways a sort of more biologically

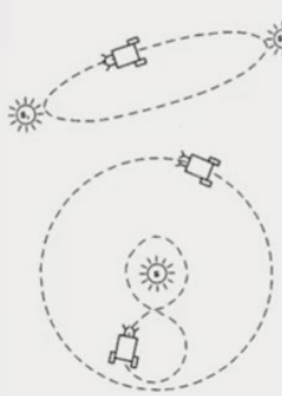


Figure 7
Trajectories of vehicles of brand 4a around or between sources.

realistic sensor motor connection in that it implies a sort of threshold beyond which the stimulation to the motor can't go, and there's a kind of, you might call it a fatigue, when the intensity is too high that sort of diminishes input to the motor. So that's what this… But you could make any kind, all kinds of interestingly designed, non-linear connections between sensors and motors, and depending on the connection between sensors and motors that you get. You could now create vehicles that, for example, oscillate between two lights or circle a light or make a figure 8 around a light. You could play with of a lot of different ways of linking sensors to motors.

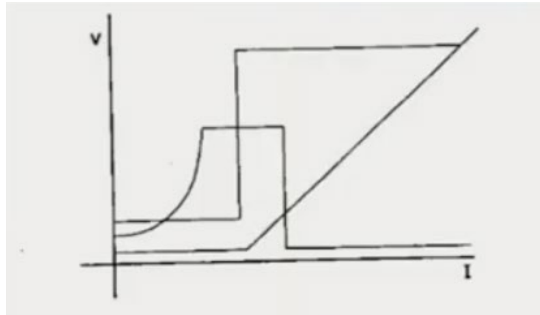
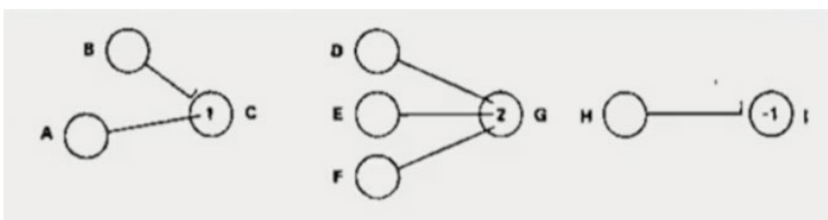


Figure 8
Various bizarre kinds of dependence of the speed of the motor (ordinate) on the intensity of stimulation (abscissa) in Vehicle 4b.

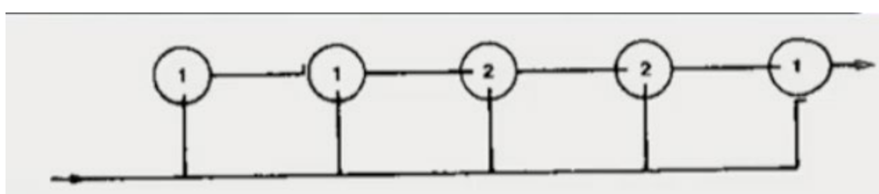
[22:42] Braitenberg, in his book, shows a number of them. So, if you look at these several sample intensity and motor velocity relations, you see that, take a look at the one that goes like this and then up at a diagonal. That connection seems to have a certain threshold. Up to a certain threshold, the sensor doesn't send any, or very little, it sends very little stimulation to the motor. But beyond that threshold, it acts more or less like a linear excitatory connection. So that's one possibility. The kind of step function there is just sort of low to high input depending on a threshold. And then that third one with the curve is sort of a complex sensor motor relationship. But again, you could play with a lot different versions of these, and get much more complex vehicle behavior. As you're reading about this, all the way along in the book, again you're thinking, well, I'd have to experiment, I'm not sure I could predict what's going to happen with all these kinds of connections, but I could certainly experiment with it, I could program it or build a little robot with these sort of connections and see how the animal behaves.



[24:22] the next step that Braitenberg includes is to build things that are similar in spirit to the neural nets that we saw before, but in fact actually a little simpler. I'll step over to the side here. This is Braitenberg's notation, and I want to describe this carefully so that you can understand what he's getting at here. Look at the left sketch here. You're seeing three of these units. When a unit has a number in it, that indicates a threshold. So, let's look at unit C here in this left drawing. What that means, what the threshold of 1 means, is that if unit C receives an input of 1 or greater at time N, you think of this as being controlled by a discrete clock. So, if at time N, unit C receives an input of 1 or greater, then at time N+1, unit C would fire a 1 input on any of its outgoing wires. In this case, C is drawn without any outgoing wires, but if it had any, it would then send a 1 out at the next time tick. So again, if at time 16, C gets an input of 1 or greater, then at time 17, it will send out a 1 along any of its output wires. A here is a unit which could send a 1 or a 0, depending on what we want it to do so it's like a neural net or you could, in effect you could have seen it as something with a threshold of 0. So, if you decide to put a 1 in at unit A, it'll send a 1 out to unit C at the next time tick. B is likewise a threshold 0, if you send a 1 into unit B, it'll send a signal out on its output wire, but that little curve on the input from B to C indicates that this is an inhibitory connection. So, a 1 at input to B at time 16 will send a -1 input to C at time 17.

[27:08] Let's just see what this whole little machine does, ok? It's a little three-unit machine. If at time 16, we put a 1 into A and a 0 into B, then at time 17, C will fire. That is, if A is firing at time 16 and sending a 1 into unit C and B is not firing at time 16, then C will fire at time 17. If A is firing at time 16 and B is also firing at time 16, then C will not fire at time 17, and the reason for that is because A is sending a 1 and B is sending a -1 and together those add to 0, so C does not meet its threshold and does not fire at time 17. What about, suppose B is firing but A is not. Then again, C will not fire at time 17, because the net input to C is -1. With that as introduction, you should see that for example in that middle machine, what have we got here? It's like one of our OR gate perceptrons, if any two or all three of D, E, and F fire at time 16, then G will fire at time 17. If fewer than 2 of D, E, and F fire at time 16, G will not fire because it has a threshold of 2. And finally, the little machine at the left, well, I'll just explain it in the sort of most natural terms. The unit I has a threshold of -1. If it's getting no input, then it's firing at every time tick. If it gets an input from H at time 16, that reduces the input to -1, and I will not fire at time 17. I guess to be really accurate here, I mean you know when reading this you know what Braitenberg is getting at. But to be really accurate here, you would want the threshold for I to be -0.99, something like that. So that when H fires at time 16, it sends a -1 to I, which is below its threshold, and then at time 17, I would not fire at all. On the other hand, if H is not firing at time 16, then I is above threshold and will fire. So, these are three little sample arrangements of these simplified neural nets. Again, they operate off a time stamp, so you could think of a clock ticking and depending on what each little unit is doing at time N, that will determine what the entire machine is doing at time N+1.

[30:42] With that, I'll give you a little puzzle for you to figure out. I'll tell you its behavior, but now you should go back and try this machine and



see how it works. The claim is that if you send on the left input wire, notice that left input wire goes to each and every one of the units here, five units. The first unit has a threshold of 1, as does the 2nd and the 5th. The 3rd and 4th unit have thresholds of 2, and that first unit has an inhibitory connection to the 2nd and the overall wire has an inhibitory connection to the 5th. I claim that if you send this machine a pattern of 01110, so 0 followed by three 1s followed by a 0, then this machine will fire at the next time tick. This, according to, you could then say, according to Braitenberg, this is a machine, you could fit this into a vehicle. Now place this kind of apparatus into a vehicle, so that if perhaps if there are three beats of red light, then the vehicle moves. If there are 2 or 4 beats of red light, the vehicle will not move. In fact, the only time this little machine fires is if there's a sequence of three beats put into the machine. So, if you put this sequence of units inside of a vehicle, it will respond to three successive inputs, but not 2 or not 4. If you saw an animal behaving this way, you would say that it is responsive to the number 3, it likes to get 3 inputs of some kind. But again, think of this from, you know, once we know what's inside the vehicle, where's the number 3 in there? How is this exactly representing the number 3? It's just connections between rather simple units. So, we would attribute a numerical understanding to the animal, when you open up the animal, you see that it just has a very simple arrangement inside.

[33:28] We'll stop with the sort of future elaborations of Braitenberg's vehicles here. Again, I would really recommend that you pick up the book and read it if you're interested in this line of thinking. But you can play directly with these sorts of things, there are vehicle simulations on the web, and the original, actually what you're looking at here are Lego vehicles, little simple robots. LEGO Mindstorms, the robotics Lego kit, is tailor made for making these kinds of working vehicles of the kind I just described to you. It's a perfect medium for representing the relationships between sensors and motors. The upper picture here was a very early prototype of the Lego Mindstorms system and the developer of that prototype, Mitchel Resnick, when he first created this set, this Lego set, he called that yellow processor in the middle, a Braitenberg Brick, after Lego bricks. This is a Braitenberg Brick that you could put into Lego constructions so it would behave like one of Braitenberg's vehicles. So, there was a direct connection, a direct intellectual connection between the little sort of robotics figures that you can construct, and the ideas first put forward in Braitenberg's book. Now this is, of course, just a very beginning of developing robots that respond to their environment and move around and so forth, but you can begin to see how an elaboration, how a continuation of this line of thinking could get robots for example that follow a black line on a smooth surface. So, if you have a white surface with a black path painted on it, you can use Braitenberg-like ideas to get a robot that will follow that black line or that will bump into a wall and then turn around and move in the other direction from that wall. Or will follow along the side of a wall perhaps a little bit like the robotic devices like Roomba, the robotic vacuum cleaner. You can get these robots to exhibit what you might of as kind of interesting, simplified animal like behavior.

[36:27] At its most inspiring, I mean that, how far can we take this as a representation of intelligence? Well the roboticist Rod Brooks, historically, really felt that the way to implement artificial intelligence was through building complete creatures like Braitenberg's vehicles and to layer one creature on top of the other to make them more and more complex over time. But his ideas resonate quite a bit with the ideas of Braitenberg. These are passages from a famous essay that Brooks wrote in the early 90s called *Intelligence Without Representation*. And this was something of a challenge to the sort of mainstream artificial intelligence community. So here are a couple of quotes: "We must incrementally build up the capabilities of intelligent systems, having complete systems at each step of the way and thus automatically ensure that the pieces and their interfaces are valid." And again: "At each step we should build complete intelligent systems that we let loose in the real world with real sensing and real action. Anything less provides a candidate with which we can delude ourselves."

[38:10] In particular here, Brooks was arguing against the style of artificial intelligence which tries to carve out particular cognitive problems and solve them completely in the absence of a surrounding animal in which those cognitive abilities are imbedded. So, solving the vision problem, for example, to Brooks is not terribly meaningful unless you solve it in the context of a creature, you know, a sensible somewhat intelligent creature that can make use of the vision in ways that are beneficial to it. His idea was that you have to build animals rather than to solve particular subproblems of AI independently. Here's some more quotes from Brooks: "When we examine very simple level intelligence, we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model." Think about that in the context, for example, of the vehicle that hates lightbulbs. It has no concept of lightbulb; it just attacks them. We don't need a representation of lightbulb to get a, to build a vehicle that "hates" lightbulbs. We don't need a representation of the number 3 to get a vehicle that will respond to three beats of a red light. And then Brooks goes on: "Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems."

[40:08] Now I should say that these are strong, he calls it a radical hypothesis, strong and highly controversial claims. A lot of people would say, that, for example, human level intelligence, and quite likely for the intelligence of mammals and large brained creatures, that it would be impossible to get the full intelligence without some elements of representation. The problem seems to be particularly pointed when we include the ability of human beings to use language. After all, language seems to be a medium of symbolic representation. So, not everyone agrees with Brooks, and in fact many, many people disagree. But there's value to this kind of strong interpretation and strong hypothesis to see how far you could go with it.

[41:14] Here are kind of slogans of the Brooks style of AI design, just to sort of give you this flavor of a kind of AI focused on the building of robots. And I should say, robots in the spirit of artificial animals. First, intelligence without representation. See how far you can get by letting, by having the creature respond directly to the world, directly to the input of the world, rather than make some intermediate abstract representation of the world. The second slogan is the world is its own best representation. Those three beats of light are the best representation of 3, you don't need a concept of 3 to respond to them. So, you can get input directly from the world and use that to make decisions, rather than get input from the world, make an abstract representation, and then use the representation to make decisions That's Brooks's view.

[42:27] Work on, what he's really describing here when he says horizontal microworlds as opposed to vertical microworlds, this is kind of like what I was saying before. It's better to get a simple entirely working creature with limited vision than an all-purpose perfect solution to the vision problem that is de-contextualized and removed from any sort of implementation in a working creature. It's better to make, that's what he means by horizontal, it's better to make animals than to solve vision and hearing and numerical processing and things like that. Better to make a working animal and then build greater and greater and more complex working animals on top of that. In that spirit, Brooks recommends this kind of layered architecture where you take an existing animal, and you add more things to it that sort of build upon and communicate with the layers of animal behavior below, if you want to think of that, in that spirit. A complete animal can be the lower layer of a more complex animal which has abilities that sit on top of the more simple animal. Again, there's no grand processor here, the intelligence could be distributed in many ways. In a number of robotics projects, for example, there are small processors that control little movements of limbs or legs or eyes or something like that. There's no one grand processor that communicates with each and every thing, the processing can be distributed in different parts throughout the animal. And the behaviors of those animals are therefore emergent, they emerge from the interaction of these simpler processors scattered throughout the entire animal.

[44:53] And finally, Brooks would always recommend that the test cases should be in the real world, as much as possible. That is, the test cases should be based on working animals that run on the floor, that run in the tub of water, that actually perform tasks, because that keeps you honest, keeps you from finessing away certain hard problems that you might be tempted to do if the vehicle or animal in question is purely simulated. Now is all this the key to AI? Well Brooks would argue that it is, or at least he did at this time argue that it is. Personally, I'm not convinced, but there's a great value to this argument because it provides alternatives, it provides a different lens through which to look at intelligence, and it provides a different lens through which to look at the design of intelligent creatures. And in this case, you might focus, for example, on rather than designing a human, you know, human level intelligence, something to pass the Turing test. You might design an artificial fish or moth or caterpillar or all kinds of different creatures to get a sense of how to represent the behavior of those creatures in rather simple, mechanical terms.