Concept Acquisition: How to get something from nothing

Dan Ryder, UNC – Chapel Hill

[Blank overhead] The mind at birth! This is kind of a shocking thing to suggest, because Fodor's given a powerful argument for concept nativism, the conclusion that it's *impossible* for the mind to start out as a blank slate. In fact, Fodor thinks that *lots* of concepts are innate, including (infamously) the concept of a computer and the concept of a platypus. I'm sure you'll all agree that thisis a conclusion we should work hard to avoid! So I'm going to present a theory of concept acquisition that escapes Fodor's argument. The usual response to Fodor is to defend some version of *minimal* nativism, whereby we start off with a few concepts and build new concepts from this innate base. (This is the response favoured by Fiona Cowie, for example.) Fodor has a number of independent arguments for why minimal nativism fails. The strategy that I will be presenting *could* be used to defend minimal nativism. But I'm going to go for the gusto. It could very well be, I claim, that *no* concepts are innate. I dream of being a Radical Empiricist.

First I should clarify my thesis. When I say the mind starts off as a blank slate, I'm saying that it's devoid of *substantive concepts* or *ideas*, that is non-logical concepts or ideas. Some examples of substantive concepts are: the concept of a cat, the concept of a quark, the concept of being square, and the concept of heaviness.

Also, by "concept" I *don't* mean an abstract entity, a constituent of a proposition. I mean what Fodor means: a mental particular in here, a stored mental representation that can be deployed in judgement and other occurrent propositional attitudes. I'll be assuming that that grasp of an *abstract* concept must occur via some mental representation that has that abstract concept as its content.

Also, with Fodor, I'm going to ignore the perceptual/conceptual distinction. In fact, I'll just use "concept" and "mental representation" interchangeably. For my

purposes, nothing hinges on any distinction there. Note how radical my thesis is, then – I'm really suggesting that you can get something from *absolutely nothing*, that you can start from no substantive mental representations *at all*, and acquire some.

So… Fodor's argument for concept nativism. This isn't quite Fodor's argument, rather it's an argument constructed from materials Fodor provides. Fodor's argument is actually a long and convoluted series, but I'm going to give what I think is the *essence* of the argument. And in responding to it, my strategy will be to grant Fodor as much as possible.

Either you get concepts from other concepts, or you get concepts from something that's not concepts. Suppose we get concepts from other concepts, that is, suppose we derive new concepts from concepts we've already got. One way to do this, for example, is by concept *construction*. You take concepts you already have, and put them together to produce *complex* concepts. That sort of thing is pretty popular with empiricists. But if you can only ever derive new concepts from concepts you've already got, then some concepts have to be innate. That much is clear.

Given the goal of a blank slate, we need to show that it's possible to derive concepts from something that's *not* concepts. That's what it means to get concepts from *nothing*. And this is where Fodor raises his roadblock. He raises a constraint that concept acquisition must satisfy – and then argues that this constraint *can't* be satisfied if you suppose that you get concepts from nothing. The only way to satisfy the constraint is to get concepts from other concepts.

He calls this constraint "the doorknob/DOORKNOB" problem. Any account of concept acquisition, he says, has to solve the doorknob/DOORKNOB problem: it has to explain why it is that we acquire the concept of a doorknob through

experiences with *doorknobs* rather than through experiences of whipped cream, or Darth Vader, or anything else.

Here's one venerable solution to the doorknob/DOORKNOB problem: the reason you acquire the concept of a doorknob from doorknobs, and not Darth Vader, is that acquiring the concept of a doorknob involves acquiring beliefs that are somehow connected with doorknobhood. Maybe it involves acquiring beliefs about *what a doorknob is*, or beliefs about *what the term "doorknob" applies to*, or beliefs about *what the essence of doorknobs is*, or beliefs about *what doorknobs look like*. Acquiring the concept of a doorknob involves learning some facts about doorknobs. And since doorknobs are a good source of evidence for facts about doorknobs, but Darth Vader *isn't* (even if you ask him nicely), you acquire the concept of a doorknob from doorknobs, and not from Darth Vader. It's a pretty obvious story, really. *It exhibits concept acquisition as a rational process*.

The problem, for our purposes, is that this way of solving the doorknob/DOORKNOB problem commits us to concept nativism. Or so Fodor argues, through what he modestly calls "The Standard Argument." Any rational process of concept acquisition like this will involve the application of *reason* to representations, *representations that you must already have*. There's nothing *else* you can apply reason to. For example, take hypothesis testing and confirmation – which is *the* rational process of empirical belief acquisition, according to Fodor. It requires that you already have the concepts needed *to frame your hypotheses*, to later be confirmed. Reasoning always takes representations as its input, so if concept acquisition always occurs *via* some sort of reasoning, at some point in the process there will have to be some *innate* representations to reason *from*. So this way of solving the doorknob/DOORKNOB problem – that is *via* a rational process – commits you to nativism.

So we need to find an alternative, *non*-rational process to explain concept acquisition, that is, concept acquisition by a *brute causal* process, or as it's often called, concept *triggering*. Acquisition by concept triggering works like this: When you're born, you don't have the concept of X. But your brain gets wired

(either at birth or through development) so that you'll *acquire* the concept if you get the right stimulus, the concept's trigger. It might take a long time for you to get wired so that the trigger will make you acquire the concept. (Maybe it takes until puberty!) But then you get exposed to the trigger, and BOOM, you've got the concept.

The problem is, Fodor argues, triggering leaves the doorknob/DOORKNOB problem unsolved. Here, I've made the trigger a doorknob. But if the process by which a concept is acquired is brute causal, that is, *not* a rational process, there's no reason why one stimulus rather than another should be the one to cause the concept to suddenly be acquired. Saying that a concept is triggered is sort of like saying that it's *grown*, that it's a product of development. (Things that are the product of development can still depend upon stimuli from the environment.) You give a plant water and sunlight, and it grows taller. The doorknob/DOORKNOB problem amounts to this: Why isn't it water and sunlight that make you grow the concept of a doorknob? Or, for that matter, why isn't it Darth Vader that makes you grow the concept of a doorknob? Brute causation isn't an acceptable account of concept acquisition, says Fodor, because it doesn't solve the doorknob/DOORKNOB problem.

So here's our situation. It looks as though if we want to avoid concept nativism, we have to satisfy two mutually unsatisfiable constraints. On the one hand, we have to solve the doorknob/DOORKNOB problem. But the only way we can see to solve that is for concept acquisition to be a rational process. But on the other hand, in order to avoid the Standard Argument, concept acquisition *can't* be a rational process, because that requires antecedent representations, which violates the other constraint. Concept acquisition has to be a brute causal process… but that won't solve the doorknob/DOORKNOB problem, and back and forth. So it looks like we can't avoid concept nativism. *That's* the powerful argument.

OK. Obviously I think there's a way out of this. The first thing to ask is: are these two constraints *really* mutually unsatisfiable? Couldn't we somehow *combine* the rational and the brute causal, and get the best of both? It's been done before!

To combine the rational and the brute causal would be to *mechanize* a rational process. Turing showed us one way to do this, in the computational explanation of deductive inference. Computation mechanizes deductive or mathematical inference in the following way. A deductive inference at the semantic level can be "realized" by a dumb process of symbol manipulation at the syntactic level. By this dumb process, the symbolic manipulations mirror rational relations amongst the meanings of the symbols. Then the syntactic operation can, in turn, be "realized" in some physical process. This was Turing's insight, which led to the development of the computer. And it's why computationalism offers much hope to the physicalist – mechanizing inference shows a way in which the mysterious faculty of reason *might not be so mysterious after all*.

This sort of inference operates on representations though; it isn't a process of the rational *acquisition* of representations. So it's not going to help with our problem. I'm just using it to show that there's no obstacle *in principle* to mechanizing rational processes, of getting a mechanical operation to have an input/output profile that accords with rational principles when characterized semantically. The idea is, then, to take a rational process of *concept acquisition*, and mechanize *that*. The hope is that the *rational* aspect of this process will solve the doorknob/DOORKNOB problem, while the brute causal aspect will allow it to operate even in the absence of representations.

First, I'll describe a particular kind of rational process of concept acquisition, one that we haven't seen yet. Then I'll show how to mechanize this process of concept acquisition. And finally I'll show how it can work in the absence of representations, while solving the doorknob/DOORKNOB problem.

OK. I call the model of rational concept acquisition that ultimately provides a way to avoid nativism "acquisition by abduction" (JOKE); "abduction" in the sense of inference to the best explanation. It's how Mendel, for instance, acquired the concept of a gene. Mendel observed a bunch of pea-plant crossings, and he observed some complicated correlations across generations of related pea plants. And he supposed that there was some hidden source for those correlations – and in supposing that, he thereby acquired the concept of the *gene*.

Now, acquisition by abduction is a rational process, and so, according to the Standard Argument, we should expect it to presuppose the prior availability of some representations. And so it does. In fact, it's a variety of hypothesis testing and confirmation, so Mendel, for instance, needed to already have some representations available – he needed them in order to frame his hypotheses.

But there's something different about acquisition by abduction, there's something unusual about one of the hypotheses that gets confirmed. There are two kinds of hypotheses. First, there are the hypotheses that genes explain a bunch of observations. Those clearly require antecedent concepts, of peas and greenness and stuff like that. But look at the hypothesis that genes exist. I'd like to propose the following: that when Mendel came up with the hypothesis that genes exist, this didn't presuppose his prior possession of the concept of a gene. His hypothesis that genes exist and his mental representation of genes *came into existence at the same time*. Mendel's observations of pea plants did not lead to the exercise of a concept *he already had*, it led him *to create the concept*.

Now, imagine some traditional empiricist adopting this account of concept formation. You start with a blooming, buzzing confusion of sensory states, and a primitive set of sensory concepts. Noting correlations amongst sensory states, you hypothesize the existence of *sources* of those correlations. First, maybe, you hypothesize the existence of shapes. Then you note correlations in how shapes are presented to you over time, and you hypothesize the existence of physical objects. Building on this, you could even, eventually, hypothesize the existence of natural kinds. For example, you notice the correlation between clarity, liquidity,

freezing at 0 degrees, etc., and hypothesize the existence of some underlying explanation for the correlation – namely the natural kind water, unified by some essence, which turns out to be its molecular structure. Maybe you'd hypothesize the existence of genes eventually.

So if abduction is our rational process of concept acquisition, it could be that our concepts of shapes, objects, kinds, etc. are *acquired* rather than being innate. The concepts wielded in the existence hypotheses don't have to be innate. But the Standard Argument commits us to *some* concepts being innate. In particular, we have to start off with some sensory representations – these are analogous to the representations of the properties of peas and pea plants that Mendel started off with. We can't start off with just *sensations*, if by that you mean some non-representational states or objects in the mind. If there are such non-representational states or objects, in order for our abductive process to proceed, we'd have to *appreciate* that they fall into certain patterns, which presupposes that we can *represent* them. Acquisition by abduction is an instance of reasoning, and we can only reason from representations. (This is just an application of a point that Sellars stressed a long time ago.) So these sensory representations, it seems, must be in the innate base.

The second set of concepts that will have to be in the innate base are connected with the *explanatory* hypotheses that one entertains when engaging in inference to the best explanation. Mendel hypothesized that genes *causally explain* the correlations in pea plant data, and here we hypothesize that shapes and physical objects explain sensory correlations. Whence these causal/explanatory concepts? They must be in the innate base.

So the rational process of acquisition by abduction shows how we can acquire some concepts, but it still commits us to two sorts of concepts being in the innate base – sensory representations and causal/explanatory concepts. I think that the *mechanization* of acquisition by abduction can eliminate the need for these representations to be in the innate base. Now I'll explain the mechanism.

The most likely place for this mechanism to be instantiated is in the main kind of cell in the cerebral cortex. So I'll explain the mechanism with respect to this kind of cell, the pyramidal cell.

A neuron receives inputs on its dendrites, which are these elaborate tree-like structures. There are thousands of connections from other neurons on this cell. The ones you can see are excitatory, which increase "activity", but it also has inhibitory connections, which decrease activity. You can think of activity as a signal level. Each principal dendrite – an entire tree-like structure attached to the cell body (show) – produces an activity determined by all of the excitatory and inhibitory inputs that it receives. This activity is that dendrite's *output*, which it passes onto the cell body (show). Each of these principal dendrites produces its own output, which it passes onto the cell body. The output of the whole cell (which it delivers elsewhere via its axon) is determined in turn by the outputs of its principal dendrites.

The input/output profile of a dendrite, and thus its contribution to the whole cell's output, can be modified by adjusting the strengths of these synaptic connections, and possibly by modifying other properties of the dendrites as well, like their shapes. An important question in neuroscience is: What principles underlie the adjustments a cell makes to its synaptic connections in order to settle on some input to output causal profile? Why do certain connections here become highly influential, while others get ignored or even dropped? And what determines the nature of the influence they come to exert? There's one proposal for a synaptic adjustment principle that, if it's right, would mean these cells are little abduction machines.

The proposal is this: that each principal dendrite will adjust its connections so that it will tend to contribute *the same amount of activity* to the cell's output as the other principal dendrites do. So if there are 5 principal dendrites, like on this cell, they'll each tend to adjust their connections over time so that they'll

consistently contribute $1/5^{th}$ of the cell's total output. I'll put this by saying "They try to match each other's activities." "Trying", of course, is just a convenient metaphor. It's just a brute causal tendency they have. Cells that have this particular brute causal connection adjustment tendency are called "SINBAD" cells. (That stands for a Set of Interacting Backpropagating Dendrites, which refers to the mechanism by which the dendrites try to match each other's activities.)

For simplicity, consider a SINBAD cell that has only two principal dendrites. They're trying to contribute an equal amount, 50%, to the cell's output, that is they're trying to match each other's activities. And they're trying to do that *consistently*, no matter what inputs they happen to get. Suppose they're connected to the same detector, or sensory receptor. It'll be easy for them to match. They both just pass that input on to the cell body, and they'll always match.

But dendrites *don't* get the same inputs. Then they're matching task isn't going to be trivial. Suppose they get two totally unrelated inputs. To use a fanciful example, suppose this one gets an input from a green ball detector, and this one gets an input from a whistle detector. Suppose both detectors go off at the same time, there's a green ball, and there's also a whistle. So both dendrites become active at the same level, let's say 40 units, and they pass that on to the cell body, which will become active at 80 units. They've passed the *same* amount of activity onto the cell body, so the adjustment principle says Great! Stay the way you are! You matched!

But the thing is, it was a *coincidence* that there was a green ball and a whistle at the same time. Next time, maybe there's just a green ball. This dendrite will account for 100% of the cell's output, and this dendrite zero. The output of the cell is then 40 units, and the adjustment principle says to the green ball dendrite: Whoah, too much, so that connection weakens. And to the whistle dendrite: Pick it up, mate, so any active connections will be strengthened. But it's a hopeless case; the two dendrites will never consistently match activities, because they're

getting totally unrelated inputs. *The only way they can match is if their inputs are in some way mutually predictable.*

The most basic form of mutual predictability is *simple correlation*. If green balls and whistles were consistently correlated, then the two dendrites would be able to match their activities consistently. So, for instance, if in addition you had a beak detector and a feather detector, the dendrites could learn to match. These two connections would strengthen, and these two to weaken to nothing. The learning rule would make this dendrite respond strongly to beaks, and this one to feathers. Because beaks and feathers are consistently correlated in the environment, the dendrites will consistently match.

But there are more complex forms of mutual predictability than simple correlation. Remember that these dendrites get *lots* of inputs - thousands. And they're capable of integrating these inputs in complex ways. So the dendrites can find not just simple correlations between beak and feather, but also what I call "complex correlations" between *functions* of *multiple* inputs.

Take another cell. Suppose that amongst the detectors this first dendrite is connected to is a bird detector and a Queen Elizabeth II detector, and for the second dendrite, an undecagon detector and a bronziness detector. (Clearly detectors that no well-equipped organism should be without!) Now, there's no consistent simple correlation between any two of these, but there is a consistent *complex* correlation – bird XOR Queen Elizabeth is correlated with undecagon AND bronzy. So in order to consistently match, the dendrites will have to adjust their input/output profiles to satisfy a couple of truth tables. The first dendrite will learn to contribute 50% when this function is satisfied, and the second one will learn to contribute 50% only when this function is satisfied; otherwise they'll both be inactive (output = 0). Since these two functions are correlated in the environment, the two dendrites will now always match their activities, and adjustment in this cell will cease.

Now, consistent environmental correlations aren't accidental: there is virtually always a *reason* behind the correlations. For example, the correlation between beaks and feathers in the first example isn't accidental – they're correlated because there's this natural kind, birds, whose essence explains why they tend to have both beaks and feathers. What's going to happen to a cell that has one dendrite that comes to respond to beaks, while the other comes to respond to feathers? The cell is going to respond to *birds* – the thing that *explains* the correlations in its inputs.

Or consider this cell, where the two dendrites discovered, not a simple correlation, but a complex correlation between *functions* of their inputs. Again, the correlation between the *xor* and *and* functions has an explanation in the environment, something that the cell as a whole will come to respond to – *loonies*. In a very simple way, the cell has "inferred" that the complex correlation amongst these properties has an explanatory source. It has performed, mechanistically, a very simple sort of inference to the best explanation. Like Mendel, it has "observed" a regularity in the environment, and "postulated" the existence of a hidden explanation for that regularity. *But it has done this entirely mechanistically, without the help of representations.* It's just brute causation. However, the cell acts in accordance with the rational principle that you should postulate *sources* for complex correlations that you observe; in this case, that source is in fact the loony. The cell makes its "postulation" despite the complex relation that exists between the data (the inputs to the cell's dendrites) and the "theoretical posit", the thing the cell comes to respond to. Whenever there are complex correlations to be found in a SINBAD cell's inputs, it will "postulate" the existence of, that is, come to respond to, the source that explains those correlations.

Mendel postulated the existence of genes in order to explain a set of complex correlations he observed in pea plant crossings. The existence of genes was implicitly suggested by the data that Mendel collected, since it stood to reason that the correlations he observed had some source. What Mendel did was make explicit what was implicitly suggested by his data. Although the "explanatory

inference" that a cell mechanizes is significantly less complex, it follows exactly the same pattern. The existence of some source of correlation is suggested by the regularities among the inputs a cell's dendrites receive, and the adjustment principle, *which is purely mechanical*, makes a cell come to respond to that source, thus making it explicit.

So I'm saying that a network of SINBAD cells (possibly a brain) starts off as a blank slate, no representations. You give it a bunch of detector inputs, and by mechanical acquisition by abduction, new representations get created. Now, when we looked at *non*-mechanical acquisition by abduction, it seemed that there were two sorts of representations that had to be in the innate base in order to apply abductive reasoning to them. How is it, exactly, that we eliminate the need for these representations in *mechanical* acquisition by abduction? How does mechanization get us a blank slate?

One set of representations we needed were causal/explanatory concepts. An essential part of Mendel's hypothesis that ultimately led him to conclude that genes existed was that they *causally explained* correlations in properties of pea plants across generations. In order to hypothesize this, he needed causal/explanatory concepts. But when a SINBAD cell does its thing, it doesn't mobilize any such representations. A cell is just brute causally structured so that it tends to come to respond to sources of correlation. It's like in the standard computationalist story, where an AND gate doesn't need a *representation* of conjunction in order to realize its truth table. Causal/explanatory concepts in mechanical abduction are analogous to logical or mathematical concepts in standard computation. (There's a price to pay for this though: mechanical abduction is only a very *simple* form of abduction, it's just the simplest postulation of a source of correlation. It's not what Mendel did!)

The other set of concepts that it seemed needed to be innate in non-mechanical abduction were concepts in the primitive base. Mendel needed concepts of the properties of peas and pea plants, and the empiricist account of concept acquisition layer by later, from shapes to objects to kinds, needed a *first* layer of

sensory concepts to start out from. I need to show that mechanical abduction eliminates the need for this first layer of sensory concepts.

The idea is that, sure, mechanical abduction is capable of operating upon inputs that are genuine representations, but it is equally capable of operating upon *non*-representational inputs that come directly from the environment. For example, you can get a representation of the natural kind bird abductively derived from *within* the network from representations of beaks, feathers, and what have you. But you can also get representations derived by the same mechanical process from non-representationally mediated interaction with the environment. So what I need to show is that these detector responses *aren't* mental representations, while the responses that develop in the abductive network *are* mental representations. Or, for short: the detectors are a *mere interface* between the environment and a representational system.

What I need to do is give an account of what mental representation is that delivers the result that the detectors *aren't* representational, but the SINBAD cells in the abductive network *are*. I'll give a variety of *locking* theory – locking theories are non-cognitive theories of concept possession, or mental representation possession. That is, they don't say that having a concept is a matter of having beliefs. Locking theories say that concept possession, or mental representation, is a matter of having some mental particular that "locks" to (or sometimes they say "resonates" with) some item in the environment, which is thereby the content of that concept. "Locking" is just a placeholder for some kind of naturalistic relation. There are a lot of different stories for what this relation is, what "locking" amounts to – like carrying information, or having the function of carrying information, or having a causal-historical link to, or whatever. Here, very briefly, is what I think is the right story – and it will get us what we want, namely mental representations, or "locking" here, but nothing here. Representations from nothing.

Mental representation, I claim, is just like representation in a *model*. Take, for example, an abstract model of the Space Shuttle re-entering the atmosphere, a

model an aeronautical engineer might use. It'll be isomorphic to a system consisting of the Space Shuttle and a bit of the atmosphere. It'll have parts, or "nodes", that stand in for the temperature of the hull, the angle of descent, the Shuttle's position and velocity, the atmospheric density, etc. etc. Also, it's a *dynamical* model, which makes it more useful than a toy model airplane. It will do "filling in." For example, if the Shuttle's coming down, and you know its position and velocity and stuff, but you don't know the temperature of its hull (maybe the thermometer's broken), you can plug the things that you know into your model, and then just look at the node that stands in for temperature, and read off the temperature of the hull. That's filling in missing information, one function of models. Or, if you want to know how fast to go in order keep the temperature below 600 degrees at a particular atmospheric density and angle of descent, you plug those values into the model and read off the maximum velocity from the velocity stand-in. So you can use a model, not only to fill in missing information, but also in order to figure out how to act. You know you need to keep the velocity down to a certain level in order to get what you need – a 450 degree hull temperature.

Now, the model might be isomorphic to all sorts of different environmental structures; that is to say, its nodes might correspond to all sorts of different things. (Correspondence is the analog of isomorphism at the level of the nodes in a structure. When two structures are isomorphic, then their nodes are said to *correspond*.) So, I was saying: a model can be isomorphic to lots of structures. But there's some subset of them (or sometimes even an individual structure) that we say the model represents. Why? Because that's how we use the model, that's how we designed it. We've assigned the model of the Space Shuttle the function of mirroring *the Space Shuttle*, and the model's nodes have the function of corresponding to hull temperature, atmospheric density, etc. That's why we say the model's nodes stand in for, or *represent* hull temperature and the rest.

Well, I think that the function of the abductive network in our brains is to *produce* models. It's the function of these networks **point to head** to *come to be* isomorphic to structure that's in the world, and for the SINBAD cells in the

networks to correspond to particular things in the world. SINBAD networks, as they do their mechanical abduction thing, gradually acquire a dynamical structure – a structure that's isomorphic to the environment. Why? Because of *lateral connections*, something I left out before. Note that if these two cells are in the same network, we don't need our bird detector any more. This cell can send its output, via a lateral connection to this other cell. There are lots of these connections – *most* of the inputs a cell gets aren't from sensory systems, but rather from other areas of the cortex.

There will be lots of cells all linked up together, in lots of different ways (remember they have thousands of connections). They each pick out their own source of correlation, and in trying to get their dendrites to match, they'll use whatever inputs they can find that are signs of that source of correlation. So the loonie cell might use inputs about size, weight, the words "one dollar", cash registers (because they typically contain loonies), and other things – either from detectors, *or* (very often) from other cells in the network. Because of these lateral connections, the relations in the world between birds and loonies and cash registers will come to be *mirrored* in the network. The whole network will gradually become *isomorphic* to the environment, with cells that *correspond* to things in the world. And when, for instance, the bronzy and undecagon and Queen Elizabeth detectors fail, other inputs to the quarter cell will *fill in* the missing information, and the network will say "loonie" anyway. Or if the network is in the "desire" functional mode, and the loonie cell is active (that is, you want a loonie), the model will fill in and tell you to go look in a cash register.

The detectors, on the other hand, *aren't* part of a model. It might be their function to detect, or indicate. But it isn't their function to *correspond* in an isomorphic structure, in order to facilitate filling in missing information and figuring out how to act. The detectors don't even have the *capacity* to do filling in.

Notice how different detectors are from models. Detectors have the function of *responding* to the environment. Whereas models have the function of

corresponding to the environment. Sure, a model can be used as a responder – that's like *belief*. Beliefs are supposed to correspond to how things *are*. But we saw that you can also use a model in order to figure out *how to act*. What velocity does the shuttle *need* to be in order to keep the temperature down. That's not responding. That's like *desire* – desires are supposed to correspond to how things *ought* to be, given your needs. Or you can just run the model offline, *exploring* it. That, roughly speaking, is what *thinking* is. Mental representations, I submit, have the function of *corresponding*, not merely responding.

Now, there's a lot more to say, for instance about why particular cells in the network have the function of corresponding to *particular* sources of correlation, that they represent *those* things. But all I need for my purposes today is to begin to convince you that the cells in here have an important characteristic of mental representations that the detectors lack – the function of *corresponding*. That's why I think these are genuine mental representations, while the detectors aren't. Which means that it's possible to acquire representations from non-representations, it's possible to acquire concepts *from nothing*.

I hasten to add that my argument against *Fodor* doesn't depend on my theory of mental representation. It's a good theory for our anti-nativist purposes because it delivers the result that these detectors aren't mental representations, although these things are. But another theory of mental representation, coupled with mechanical acquisition by abduction, could possibly deliver the same result. Or you could even mechanize some *different* rational process of concept acquisition in order to show that you can get concepts from nothing. My general strategy for avoiding concept nativism would still apply. That is: take a rational process of concept acquisition, and mechanize it. Because it's a rational process, it will solve the doorknob/DOORKNOB problem – it will explain why we acquire the concept of a doorknob from doorknobs and not Darth Vader. And because it's mechanical, it may be capable of operating upon non-representational inputs. Even when the inputs are raw detector readings, or "purely informational", or (going historical) raw, non-representational sensations, the mechanism by which a representation is created works just *the same* as when its inputs are genuine

representations. This, I think, is as close as you can get to acquiring concepts from nothing. It's a way of combining a rational process and brute causation, thus navigating between the horns of Fodor's doorknob/DOORKNOB and Standard Argument dilemma.