Book Review: Behavior – The Control Of Perception

Posted on March 6, 2017 by Scott Alexander

Epistemic status: *I only partly understood this book and am trying to review it anyway as best I can.*

I

People complain that psychology is paradigmless; it never got its Darwin or Newton to tie everything together. Nowadays people are pretty relaxed about that; who needs paradigms when you can do n = 50 studies on a mildly interesting effect? But historically, there were all of these larger-than-life figures who were sure they'd found the paradigm, geniuses who founded schools which flourished for a while, made big promises, then either fizzled out or toned down their claims enough to be accepted as slightly kooky parts of the mainstream. Sigmund Freud. BF Skinner. Carl Rogers. And those are just the big ones close to the mainstream. Everyone from Ayn Rand to Scientology tried their hand at the paradigm-inventing business for a while.

Will Powers (whose name turns out to be pretty appropriate) lands somewhere in the middle of this pack. He was an engineer/inven-

tor who specialized in cybernetic systems but wandered into psychology sometime in the sixties. He argued that everything in the brain made perfect sense if you understood cybernetic principles, and came up with a very complicated but all-encompassing idea called Perceptual Control Theory which explained thought, sensation and behavior. A few people paid attention, and his work was described as paradigm-shifting by no less of an expert on paradigm shifts than Thomas Kuhn. But in the end it never really went anywhere, psychology moved on, and nowadays only a handful of people continue research in his tradition.

Somehow I kept running into this handful, and they kept telling me to read Powers' book *Behavior: The Control Of Perception*, and I keep avoiding it. A few weeks ago I was driving down the road and I had a moment of introspection where I realized everything I was doing exactly fit Powers' theory, so I decided to give it a chance.

Powers specializes in *control systems*. The classic control system is a thermostat, which controls temperature. It has a reference point, let's say 70 degrees. If it gets much below 70 degrees, it turns on the heater until it's 70 again; if it gets much above 70 degrees, it turns on the air conditioner until it's 70 again. This is more complicated than it sounds, and there are other control systems that are even more complicated, but that's the principle. Perceptual Control Theory says that this kind of system is the basic unit of the human brain.

While I was driving on the highway a few weeks ago, I realized how much of what I do *is* perceptual control. For example, I was effort-

lessly maintaining the right distance from the car in front of me. If the car sped up a tiny bit, I would speed up a tiny bit. If the car slowed down a little bit, I would slow down a little bit. Likewise, I was maintaining the right angle relative to the road: if I found myself veering right, I would turn slightly to the left; if I found myself veering left, I would turn slightly to the right.

The theory goes further: while I'm in the car, I'm also operating as my own thermostat. I have a desired temperature: if I go below it, I'll turn on the heat, and if I go above it, I'll turn on the AC. I have a desired level of satiety: if I'm hungry, I'll stop and get something to eat; if I'm too full, there's maybe not a *huge* amount I can do but I'll at least stop eating. I have a desired level of light: if it's too dark, I'll turn on the lights; if it's too bright I'll put down the sun visor. I even have a desired angle to be sitting at: if I'm too far forward, I'll relax and lean back a little bit; if I'm too far back, I'll move forwards. All of this is so easy and automatic that I never think about it.

Powers' theories go further. He agrees that my brain sets up a control system to keep my car the proper distance from the car in front of it. But how do I determine "the proper distance"? That quantity must be fed to the system by other parts of my brain. For example, suppose that the roads are icy and I know my brakes don't work very well in the ice; I might keep a much further distance than usual. I'll still be controlling the distance, I'll just be controlling it *differently*. If the brain is control systems all the way down, we can imagine a higher-tier system controlling "accident risk" at some level (presumably low, or zero) feeding a distance level into a lower-tier

system controlling car distance at whatever level it receives. We can even imagine higher systems than this. Suppose I'm depressed, I've become suicidal, I want to die in a car accident, but in order not to scandalize my family I have to let the accident happen sort of naturally. I have a top-level system controlling "desire to die" which tells a middle-level system controlling "accident risk" what level it should go at (high), which in turn tells a lower-tier system controlling "car distance" what level *it* should go at (very close).

It doesn't even end there. My system controlling "car distance" is sending signals to a lower-tier system controlling muscle tension on my foot on the accelerator, giving it a new reference level (contracted muscles that push down on the accelerator really hard). Except this is an oversimplification, because everything that has to do with muscles is a million times more complicated than any reasonable person would think (at least until they play qwop) and so there's actually a big hierarchy of control systems just going from "want to go faster" to "successfully tense accelerator-related muscles".

II

Actually, Powers is at his most convincing when he talks about these lower-level functions. At this point I think it's pretty mainstream to say that muscle tension is set by a control system, with the Golgi tendon organs giving feedback and the spinal cord doing the calculations. Powers goes further (and I don't know how main-

stream this next part is, but I'm guessing at least somewhat), saying that this is a first-tier control system, which is itself controlled by a second-tier "direction" control system centered in the nuclei of the brainstem, which is itself controlled by a third-tier "position" control system centered in the cerebellum/thalamus/midbrain (a friendly amendment might add the basal ganglia, which Powers doesn't seem to know much about).

If you stimulate certain parts of a cat's midbrain, it will go into specific positions – for example, a position like it's ready to pounce. So it seems like those areas "code for" position. But in order to have a neuron/area/whatever that codes for position, it needs to have hierarchical control over lots of lower-level things. For example, it needs to make sure the leg muscles are however tense they're supposed to be in a pouncing position. So the third-tier position control system controls the second-tier direction control system at whatever level is necessary to make the second-tier direction control system control the first-tier muscle control system at whatever level is necessary to get the muscles in the right position.

The fourth- and fifth-tier systems, now well into the cortex (and maybe basal ganglia again) deal with sequences, eg "walking" or "playing a certain tune on the piano". Once again, activating a fourth/fifth-tier system will activate this higher-level concept ("walking"), which alters the reference levels for a third-tier system ("getting into a certain position"), which alters a second-tier system ("moving in a certain direction"), which alterns a first-tier system ("tensing/relaxing muscles").

Why do I like this theory so much? First, it correctly notes that (almost) the only thing the brain can actually *do* is change muscle tension. Yet we never think in terms of muscle tension. We don't think "I am going to tense my thigh muscle, now untense it, now tense my ankle muscle, now…", we just think "I'm going to walk". Heck, half the time we don't even think *that*, we think "I'm just going to go to the fridge" and the walking happens automatically. On the other hand, if we really want, we *can* consciously change our position, the level of tension in a certain muscle, etc. It's just that usually we deal in higher-level abstractions that automatically carry all the lower ones along with them.

Second, it explains the structure of the brain in a way I haven't seen other things do. I always hear neuroscientists talk about "this nucleus relays signals to that nucleus" or "this structure is a way station for this other structure". Spend too much time reading that kind of stuff, and you start to think of the brain as a giant relay race, where the medulla passes signals onto the thalamus which passes it to the basal ganglia which passes it to the frontal lobe and then, suddenly, thought! The obvious question there is "why do you have so many structures that just relay things to other structures?" Sometimes neuroscientists will say "Well, some processing gets done here", or even better "Well, this system modulates that system", but they're always very vague on what exactly that means. Powers' hierarchy of fifth-tier systems passing their calculations on to fourth-tier systems and so on is exactly the sort of thing that would make sense of all this relaying. My guess is every theory of neuroscience has something at least this smart, but I'd never heard it explained this well before.

Third, it's the clearest explanation of tremors I've ever heard. Consider the thermostat above. When the temperature gets below 65, it turns on the heat until the temperature gets above 70, then stops, then waits as the hot air leaks out through the window or whatever and it's 65 again, then turns on the heat again. If we chart temperature in a room with a thermostat, it will look sort of like a sine wave or zigzag with regular up/down motions. This is a basic principle of anything being controlled by a less-than-perfect control system. Our body has microtremors all the time, but when we get brain damage or some other problem, a very common symptom is noticeable tremors. These come in many different varieties that give clues to the level of brain damage and which doctors are just told to memorize. Powers actually explains them:

> When first-order systems become unstable, as when muscles exert too much effort), clonus oscillations are seen, at roughly ten cycles per second. Second-order instability, as in the tremors of Parkinsonism, involves groups of muscles and is of lower frequency, around three cycles per second or so. Third-order instability is slower stilll, slow enough that it can be characterized as "purpose tremor" or "over-correction". Certain cerebellar damage due to injury or disease can result in over- and under-shooting the mark during actions such as reaching out to grasp something, either in a continuous self-sustained oscillation or a slowly decrasing series of alternating movements.

This isn't perfect – for example, Parkinsonian tremor is usually caused by damage to the basal ganglia and the cortex, which is re-

ally hard to square with Powers' claim that it's caused by damage to second-tier systems in the medulla. But after reading this, it's really hard not to think of tremors as failures in control systems, or of the different types of tremor as failures in different levels of control system. For example, athetoid tremors are weird, seemingly purposeful, constant twisting movements caused by problems in the thalamus or some related system; after reading Powers, it's impossible for me not to think of them as failures in third-order control systems. This becomes especially clear if we compare to Powers' constant foil/nemesis, the Behaviorists. Stick to a stimulus-response paradigm, and there's no reason damaged brains should make weird twisting movements all the time. On a control-systems paradigm, it's *obvious* that that would happen.

There are occasional claims that perceptual control theory can predict certain things about muscles and coordination better than other theories, sometimes with absurdly high accuracy of like r = 0.9 or something. Powers makes some of these claims in the book, but I can't check them because I don't have the original data he worked with and I don't know how to calculate cybernetic control system outputs. But the last time I saw someone bring up one of these supposed experiments it was thoroughly shot down by people who knew more statistics. And I found a blog post where somebody who knows a lot about intricacies of muscle movement says PCT can predict some things but not much better than competing theories. In terms of predicting very specific things about human muscular movement its record seems to be kind of so-so.

III

And I start to get very skeptical when Powers moves to higher-tier control systems. His sixth tier is “relationships”, seventh is “programs”, eighth is “principles”, and ninth is “systems”. Although these tiers receive just as many pages as the earlier ones, they start sounding very abstract and they correlate a lot less well with anatomy. I understand the urge to postulate them – if you’ve already decided that the fundamental unit of the brain is the control system, why not try to explain things with control systems all the way up? – but it becomes kind of a stretch. It’s easy to see what it means to control the distance between me and the car in front of me; it’s harder to see what it means to control for “communism” or “honesty” or things like that.

I *think* the way things are supposed to work is like this. A ninth-tier system controls a very abstract concept like “communism”. So suppose you are a communist; that means your internal communism-thermostat is set to maintain your communism at a high level. That propagates down to eighth-tier principles, which are slightly less abstract concepts like “greed”; maybe your ninth-tier communism-thermostat sets your eighth-tier greed thermostat to a very low temperature because communists aren’t supposed to be greedy. Your eighth-tier greed thermostat affects levels of seventh-tier logical programs like “going to work and earning money” and “giving to charity”. I’m not really sure how the sixth-tier fits into this example, but let’s suppose that your work is hammering things. Then the fifth-tier system moves your muscles in the right

sequence to hammer things, and so on with all the lower tiers as above.

Sometimes these control systems come into contact with each other. For example, suppose that along with my ninth-tier system controlling "communism", I also have a ninth-tier system controlling "family values"; I am both an avowed communist and a family man. My family values system thinks that it's important that I earn enough to provide for my family, so while my communism-system is trying to input a low reference level for my greed-thermostat, my family-values-system is trying to input a high one. Powers gets into some really interesting examples of what happens in real industrial cybernetic systems when two opposing high-level control systems get in a fight, and thinks this is the source of all human neurosis and akrasia. I think he later wrote a self-help book based around this (hence the nominative determinism). I am not very convinced.

Am I strawmanning this picture? I'm not sure. I think one testable consequence of it is supposed to be that if we're really controlling for communism, in the cybernetic control system sense, then we should be able to test for that. For example, hide Lenin's pen and paper so that he can't write communist pamphlets, and he should start doing some other communist thing more in order to make up for it and keep his level of communism constant. I think some perceptual control theory people believe this is literally true, and propose experimental tests (or at least thought experiment tests) of perceptual control theory along these lines. This seems sketchy to me, on the grounds that if Lenin didn't start doing other stuff, we

could just say that communism wasn't truly what he was controlling.

That is, suppose I notice Lenin eating lots of chocolate every day. I theorize that he's controlling for chocolate, and so if I disturb the control system by eg shutting down his local chocolate store, he'll find a way to restore equilibrium, eg by walking further to a different store. But actually, when I shut down his local chocolate store, he just eats less chocolate. In reality, he was controlling his food intake (as we all do; that's what an obesity set point is) and when he lost access to chocolate, maybe he ate cupcakes instead and did fine.

In the same way, maybe we only think Lenin is controlling for communism, but he's actually controlling for social status, and being a communist revolutionary is a good way to gain social status. So if we make it too hard for him to be a communist revolutionary, eg by taking away his pen and paper, maybe he'll become a rock star instead and end up with the same level of social status.

This sort of thing seems so universal that as far as I can tell it makes these ideas of higher-tier control systems unproveable and unfalsifiable.

If there's any point to them at all, I think it's the way they express the same interesting phenomenological truth as the muscle movement tiers: we switch effortlessly between concentrating on low-level concepts and high-level concepts that make the low-level ones automatic. For example, I think "driving" is a good example of Pow-

ers' seventh tier, "programs" – it involves a predictable flowchart-like set of actions to achieve a simple goal. "The distance between me and the car in front of me" is a sixth-tier system, a "relationship". When I'm driving (focusing on my seventh-tier system), I don't consciously think at all about maintaining the right distance with the car in front of me. It just happens. This is really interesting in a philosophy of consciousness sense, and Powers actually gets into qualia a bit and says some things that seem a lot wiser and more moving-part-ful than most people on the subject.

It does seem like there's something going on where my decision to drive activates a lot of carefully-trained subsystems that handle the rest of it automatically, and that there's probably some neural correlate to it. But I don't know whether control systems are the right way to think about this, and I definitely don't know whether there's a sense in which "communism" is a control system.

IV

There are also some sections about things like learning and memory, which looks suspiciously like flowcharts of control systems with boxes marked "LEARNING" and "MEMORY" in them.

But I realized halfway through that I was being too harsh. Perceptual control theory wasn't quite a proposal for a new paradigm out of nowhere. It was a reaction to Behaviorism, which was still the dominant paradigm when Powers was writing. His "everything is a con-

trol system" is an attempt to improve on "everything is stimulus-response", and it really does.

For example, his theory of learning involves reward and punishment, where reward is reducing the error in a control system and punishment is increasing it. That is, suppose that you're controlling temperature, and it's too hot out. A refreshing cool glass of water would be an effective reward (since it brings you closer to your temperature reference level), and setting your hand on fire would be an effective punishment (since it brings you further from your temperature reference level). Powers notes that this explains many things Behaviorism can't. For example, they like to talk about how sugar water is a reward. But eventually rats get tired of sugar water and stop drinking it. So it seems that sugar water isn't a reward per se; it's more like reducing error in your how-much-sugar-water-should-I-have-and-did-I-already-have-the-right-amount system is the reward. If your optimal level of sugar water per day is 10 ml, then anything up to 10 ml will be a reward, and after that it will stop being attractive / start being a punishment.

As a "theory of learning", this is sort of crappy, in that I was expecting stuff about Hebb and connectionism and how memories are stored in the brain. But if you're living in an era where everybody thinks "The response to a stimulus is predictable through patterns of reward and punishment" is an A+++ Nobel-Prize-worthy learning theory, then perceptual control-based theories of learning start sounding pretty good.

So I guess it's important to see this as a product of its times. And I don't understand those times – why Behaviorism ever seemed attractive is a mystery to me, maybe requiring more backwards-reading than I can manage right now.

How useful is this book? I guess that depends on how metaphorical you want to be. Is the brain a control system? I don't know. Are police a control system trying to control crime? Are police a "response" to the "stimulus" of crime? Is a stimulus-response pairing a control system controlling for the quantity of always making sure the stimulus has the response? I think it's interesting and helpful to think of some psychological functions with these metaphors. But I'm not sure where to go from there. I think maybe there are some obvious parallels, maybe even parallels that bear fruit in empirical results, in lower level systems like motor control. Once you get to high-level systems like communism or social desirability, I'm not sure we're doing much better than the police-as-control-system metaphor. Still, I think that it's potentially a useful concept to have.