Situated Cognition, Dynamicism, and Explanation in Cognitive Science

Christopher A. Greenlee

Thesis Submitted to the Faculty of the Virginia Polytechnic Institute and State University in Partial Fulfillment of the Requirements for the degree of

Master of Arts
in
Philosophy

Valerie Gray Hardcastle, Chair
Harlan B. Miller
John Roach

June 19, 1998
Blacksburg, Virginia

Keywords: Dynamical Systems, Situated Cognition, Cognitive Science

Situated Cognition, Dynamicism, and Explanation in Cognitive Science

Christopher A. Greenlee

(ABSTRACT)

The majority of cognitive scientists today view the mind as a computer, instantiating some function mapping the inputs it gets from the environment to the gross behaviors of the organism. As a result, the emphasis in most ongoing research programmes is on finding that function, or some part of that function. Moreover, the types of functions considered are limited somewhat by the preconception that the mind must be instantiating a function that can be expressed as a computer program.

I argue that research done in the last two decades suggests that we should approach cognition with as much consideration to the environment as to the inner workings of the mind. Our cognition is often shaped by the constraints the environment places on us, not just by the "inputs" we receive from it. I argue also that there is a new approach to cognitive science, viewing the mind not as a computer but as a dynamical system, which captures the shift in perspective while eliminating the requirement that cognitive functions be expressable as computer programs.

Unfortunately, some advocates of this dynamical perspective have argued that we should replace all of traditional psychology and neuroscience with their new approach. In response to these advocates, I argue that we cannot develop an adequate dynamical picture of the mind without engaging in precisely those sorts of research and hypothesizing that traditional neuroscience and psychology engage in. In short, I argue that we require certain types of explanations in order to get our dynamical (or computational) theories off the ground, and we cannot get those from other dynamical (or computational) theories.

Chapter 1

1.0 Embodied Cognition and Explanation

I wake up in the morning when the sun hits my face, or when my alarm goes off, or when I get too hot under my sheets. There is usually some external cue that brings me to wakefulness, often grudgingly. Most of the time I'm expecting it; I've set an alarm, and expect a loud, raucous noise to tear me from my dreaming. Other times, though, I don't expect it – the sun waking me when I'd intended to sleep in, or a nightmare startling me awake in the wee hours of the morning. All in all, there are quite a few ways to wake me up.

The most obvious question for anyone interested in studying my waking behavior is, "why did he wake up just then?" Asking why I woke up when my alarm went off without taking into consideration the alarm would be silly, though. The environment provides a context in which we can interpret behavior. A person who asks you for your money from a huddled position on the sidewalk is probably a beggar, while a person who asks you for your money while pointing a gun at you is probably a robber. A person holding their hands in front of them horizontally, moving their fingers up and down, will be typing if they have a keyboard under those hands and fingers. If there is a piano instead of a keyboard there, they will be playing a tune.

The environment is also important because it constantly gives us subtle cues that guide our thoughts and actions, often in directions we wouldn't voluntarily take them. A child walks by with an ice-cream cone, and I desire ice-cream. A bird lands on my windowsill, and suddenly I can't remember what I was typing, but can only look at the small colored feathers on its underbelly. I'm flipping through radio stations, and hear a song that I've always loathed – and I can't get it out of my head. These are all experiences that everyone can identify with; they are everyday examples of the ways the environment influences our cognition.

In none of these cases does it make sense to explain the behavior in question without referring to the environment. If someone asks why I desire ice cream, a perfectly suitable answer will be that a child just walked by holding an ice-cream cone, and it started an internal chain of events resulting in my craving. If, however, the question were answered without mentioning the appearance of ice-cream in my visual field, we would not know why that particular chain of internal events occurred, rather than another. And when we are studying cognition and behavior that is precisely what we want to know.

So the question is one of causes – what were the ultimate causes of a particular behavior? The answer to that question might end up appealing to factors both internal and external to the organism. If I see ice cream after I have just gorged myself on a bucket of the same, I might very well feel ill, whereas if I had not had ice cream in a month I might very well desire it. Both internal and external states combine to shape my behaviors. The task of this chapter is to provide an analysis of precisely how external and internal factors combine to shape cognition and behavior. It is my contention that no cognitive scientific model is complete unless it addresses the relationship between an organism and its environment.

I will attempt to show this by analyzing several different cases of cognitive behavior, and making clear why they are "cognitive", why they are "behaviors", what their causes are, in what sense they are "internal" or "external", and what precisely counts as "environmental" for them. My analysis of each case will conclude with a discussion of the integral nature of the organism and its environment, and the incompleteness of any explanation referring either only to internal or only to external causal factors. I will then try to cash out precisely what having an explanation that deals with both entails.

1.1 The Environment and Human Development: Learning to Walk

Thanks to developmental psychologists such as Esther Thelen and Linda B. Smith (e.g. Thelen and Smith 1994; Thelen 1995), we have a pretty good picture of some fairly subtle factors that shape the cognition of human infants. I'm not talking about such factors as the smile on a mother's face, or the sound of a familiar voice, nor am I talking about the toys the infant is given or the color of the room in which the infant lives. While these things undoubtedly play some role in the cognitive development of the infant, it is difficult to specify that role. Instead, Thelen and Smith (1994) give us a rather interesting account of how infants learn to walk.

Infants at all ages can perform coordinated stepping, given suitable conditions. Up until about 2 months, an infant can only perform those motions while held off the ground. At around the two month mark, infants lose the ability to step while suspended, but begin kinematically similar kicking when lying on their backs. At between 8 and 10 months, infants regain their coordinated stepping when held off the ground, at the same time as they become able to support their own weight standing. Then, at about 12 months, infants start to walk.[1]

Thelen and Smith noticed that while stepping disappears at 2 months and doesn't reappear until between 8 and 10 months, infants in the intermediate period, lying on their backs, still exhibit what would be stepping motions if they were upright. What Thelen and Smith discovered is that the crucial parameter that changes at the 2 month mark, causing the disappearance of reflex-stepping, is leg mass. It seems that at about 2 months, the muscle development in infants' legs is such that the mass of the leg overcomes the spring-like tendency of the leg muscles that causes stepping. So this change is not determined solely (if at all) by cognitive development, but rather by the physical development of infant bodies.

As part of their experiment, Thelen and Smith attached weights to the legs of stepping infants to see whether the increased leg-mass would have the predicted effect. It did.

[1] Thelen and Smith (1994).

They also suspended the legs of non-stepping infants in water, to reduce the effective mass of their legs, and observed that stepping resumed while the mass was thus reduced. Moreover, non-stepping infants between the ages of 1 and 7 months displayed coordinated stepping behavior when suspended with their feet on a treadmill. This last experiment suggests that the tension in the legs when pulled backwards causes a stepping behavior as well.

1.1.1 Analysis of the Experiment

Thelen (1995) recounts these experiments in support of the development of an embodied cognition; that is, in support of the notion that we must offer explanations of cognitive behavior that involve environmental factors. What, then, are the environmental factors in this case? If they are, as I will contend, the *external* factors, then what precisely counts as "external", and what as "internal"? (And just as important, what are these factors internal or external *to*?) Moreover, how is this a case of *cognitive* behavior, since walking is most typically thought of as sub-cognitive?

In answer to our first question, the factors Thelen considers "environmental" are the mass and tension of the infants' legs. As it turns out, the infant leg behaves almost precisely like a spring-and-pulley system, with the mass of the leg and the tension of the leg muscle determining how the leg responds to influxes of energy. Because the mechanics of the legs seem to dictate the shape of the behavior, Thelen considers them environmental. Much as a weight attached to the leg of an infant impedes stepping infants, increased muscle mass impedes infants from stepping. Whether the weight is added externally or internally doesn't make a difference to Thelen.

But I would argue that both the strap-on weight and the increased leg mass are *external* factors in a very relevant sense. Consider a distinction drawn by Fred Dretske (1988). Certain things we wish to say an organism *does*, and certain things we wish to say are *done to* the organism. To use his terminology, sometimes the movements of an organism are *merely* movements, and at other times they are behaviors. For instance, if a stepping infant with a bent leg, held upright, had a weight attached to its leg, the leg would become straight. But the infant did not *cause* the leg to straighten – the weight did.

So how do the infant's own leg and muscle count as external factors? It is easy to consider a strap-on weight an external cause of a movement, but it seems almost counterintuitive to consider the infants' own muscle as an external factor. Surely, if anything is internal to the infant, its own muscles are. But consider Dretske's analysis of reflexes:

> Despite [the] involuntary nature [of reflexes], and despite the fact that we sometimes classify a response as the behavior of some bodily part (the *leg*'s response to a light tap on the knee), we often classify reflexes as behavior. We do so because the reaction to a stimulus, although perfectly reliable, is quite unlike the body's Newtonian response to a shove (where acceleration is proportional to net impressed force). Obviously, internal processes, drawing on their own power supply, are at work, transforming the input into the output. *Dretske (1988), p. 26.*

While the muscles and the legs of the infants in Thelen and Smith's study are surely internal to the body, the causal roles they are playing in reflex stepping have more in common with the shove than with the stimulus-response pairing of a reflex. One might say that they are *causally external* to the movement in question.

So what is a factor that is causally internal to reflex stepping? Consider that the legs of an infant do not just constantly oscillate in stepping as a result of their physical structure. Even a weight bounding on the end of a spring will come to rest if not perturbed. For reflex-stepping to occur, the infant has to provide energy to the legs. In other words, there has to be some neural impetus that adds tension to the leg muscles (contraction), and then relaxes them such that the mass of the leg straightens it again. This neural impetus is the internal processing, the *internal cause*, of the movement of the leg.[2]

[2] Of course, there is a hint of arbitrariness lurking around somewhere in this analysis. If we trace the causal chain back far enough, we will discover that the infants leg mass and muscle tension are results of the internal processes that grew them. Going even further back along the same chain we will eventually run into the fact that the entire body of each infant is the result of some external cause, having something to do

So, for Thelen, something counts as an environmental feature if it is causally external to the movement being studied. Conversely, something counts as an internal feature of the organism if the role it plays in producing the behavior is internal. In terms of Dretske's distinction between a mere movement and a behavior, a movement generated solely by causally external factors is a mere movement, while a movement with some causally internal factors is a behavior. We also have an answer to another of our questions here: reflex stepping is cognitive because the causally internal factors are neural. This is somewhat of a low bar for entrance to the realm of cognition, but it is not prima facie unacceptable.

1.1.2 The Nature of the Explanation

Take a moment and ask what sort of explanation is required when we ask why an infant is stepping. Consider the different experiments Thelen and Smith conducted, involving infants with weights attached to their legs, with leg mass great enough to overcome the tension of their leg muscles, who stepped when suspended in water or over a treadmill. In each of these cases, asking why an infant was or was not stepping seems to require an explanation that adverts not only to the (causally internal) cognitive processing in the infant, but also to certain causally external factors.

For instance, if we ask why an infant at the age of 1 month steps in a coordinated manner when suspended above a treadmill, it would seem insufficient to explain that it does so because processing in its motor cortex caused such-and-such muscle contractions and expansions. Instead, we must discuss both the neural processing involved *and* the environmental factors that play a role in shaping the behavior. The reason an infant at 1 month can walk on a treadmill is that the backward pull of the treadmill stretches the leg

---

with their parents and an old story about the birds and the bees. But if I want to know precisely how H_2SO_4 and a solution of OH will react when mixed, I generally leave the origin of the chemicals involved out of the explanation of the resulting reaction. For the purposes of providing a scientific explanation of behavior, this arbitrariness is only to be expected.

and causes it to rebound forward in a stepping motion. So we have a combination of factors contributing to the behavior, some of which are external, and some of which are internal.

The overall developmental picture we get is one that isn't fully characterized by appeal only to internal or only to external causes, but to a combination of the two. Stepping in an infant at the age of 1 month on a treadmill seems mostly mechanical in nature, with minimal cortical involvement, while stepping in an infant walking at the age of 12 months is a coordinated, volitional behavior, driven primarily by cognitive processing. It looks like the reflex-stepping evidenced in younger infants is a process of mechanical patterning which prepares the infant for walking, by developing neural pathways that will control walking voluntarily at a later age.

The explanation Thelen and Smith give of the behavior is one that paints a rich picture of infant learning. It is also one that *requires* reference to both causally internal and causally external factors. As just suggested, leaving out either of these categories of causal influences leaves out a significant part of the causal picture of the infants' learning. This suggests that at least in some cases a complete cognitive scientific explanation of behavior will require an ineliminable reference to the environment as well as to the organism.

1.2 The Environment and Human Development: Kicking

Esther Thelen (1994) has also provided us with an interesting insight into how infants, even at an early age, can adapt rapidly to changes in the environment. It seems that infants, up until about 5 months, typically kick with one leg at a time, either alternating legs or exclusively kicking with a single leg repeatedly. After 5 months, the dominant pattern of kicking becomes one of simultaneous kicking; the infants kick with both legs at the same time. So Thelen devised an experiment to see if she could coax infants under the age of 5 months to shift into the later pattern of simultaneous kicking.[3]

In her experiment, she tethered the left legs of a group of 3-month-old infants to a mobile. Since the infants were rewarded by the movements and sounds of the mobile whenever they kicked, they learned to kick more often. However, for some of the infants, Thelen added a soft elastic yoke between their ankles. These babies could kick in their more usual patterns (single leg or alternating legs), but it was easier and more effective for them to move the mobile by kicking with both legs. (As controls, some infants received the yoke but no tether, and some received the tether and no yoke.)

All the tethered infants whose legs were yoked shifted into simultaneous kicking, while the infants without the yoke remained in their original pattern. Over the course of a 16-minute experiment, Thelen would take an infant with the asynchronous kicking schedule and yoke its legs. During that 16 minutes, the infant would slowly shift into simultaneous kicking. Thelen would then remove the yoke from the legs, and during the remainder of the 16 minutes, the infant would return to its previous asynchronous kicking. So during a period of 16 minutes, the infants adapted their behavior to minimize effort and maximize reinforcement not once, but twice, and the first time by adopting a behavior that doesn't normally occur until a later developmental stage.

1.2.1 Analysis of the Experiment and of the Explanation

So what are the relevant factors in this experiment? The infants in this experiment are clearly engaging in a cognitive task; they are adapting their coordination patterns to achieve maximum mobile jiggling – a very desirable thing from the standpoint of an infant. So there are causally internal factors involved both in the infants' decision to change patterns and in the shift to the new pattern. There are obvious causally external factors, too. Neither the yoke, tether, or mobile seem nearly as problematic *qua*

---

[3] For a nice description of her experiment, and her design goals, see Thelen (1995).

environmental features as "leg mass" and "muscle tension" were in the previous experiment. The behavior involved is also more obviously cognitive.

What makes this experiment difficult to analyze is the actual behavior being studied. We are not looking at the kicking of the infants, but at the *coordination* of the kicking. The "movement" we are in fact studying is not the movement of the legs, but of the infant's cognitive apparatus from one coordinative pattern (alternate or single-leg kicking) to another (simultaneous kicking). Developmentally speaking, the first two patterns are stable during that stage of infant development, and the third is very unstable (and so infrequent). But it seems that, for this experiment, the pre-existing control apparatus the infant uses to coordinate its kicking is causally external to the behavior. To put it bluntly, the pre-existing neural apparatus merely gets in the way of the infant getting what it wants.

We can clear this up by borrowing another distinction from Dretske, one that in the last experiment would have been identical to our causally internal / external distinction. The yoke, the tether, the mobile, and the pre-existing neural structures that initially control their kicking are all *structural causes* of the behavior. That is, they affect the behavior of the infants because they constrain their actions. As Dretske puts it, they structure the environment such that when a particular stimulus is presented, the given behavior will result. The neural processing responsible for re-shaping the neural control structures, even though it might reside within those very same structures[4], is the *triggering cause* for the shift in coordination patterns. Or, as Dretske puts it, it is the cause that actually produces the behavior, given the right environmental conditions.

But how can we consider neural structures both causally internal and external? And if neural processing is the requirement for cognition, can cognition then be causally external to the behavior of an organism? To answer the first question, the new neural control

[4] For example, Hebbian learning takes place merely through the activation of synapses. Thus, the acquisition of the new coordination pattern could well be taking place within the actual structure that produced the previous coordination pattern.

structure can be causally internal because of its participation in the change from one coordination pattern to another. Likewise the pre-existing control structures can be causally external because of their role in shaping the actual kicking behavior. Insofar as a neural structure constrains behavior, it is causally external. Insofar as a neural structure is altered so as to provide different constraints, it is causally internal.

To answer the second question, of course cognition can be causally external to a behavior. If I am thinking about Mozart and chewing bubble gum, one likely has little to do with the other. But asking whether cognition can be external to a behavior is a question that this example simply can't illuminate for us. The loose standard for cognition we arrived at in section 1.1 was neural *processing*, and the growth and atrophy of neural connections may be the *result* of neural processing, but it is certainly not processing in itself. So the causally external factors involved in the behavior are not cognitive. In Dretske's terminology, neural systems are triggering causes insofar as they are processing, and structuring causes insofar as they are constraining behavior or processing.

Clearly, this is another case where an explanation of behavior – indeed, an explanation of cognition – necessarily involves reference to the environment in order to be complete. We need to know that the current neural structures are what they are, that they change, that the infant is receiving incentive for that change (pleasure), that a mobile moving provides that pleasure, etc. Without referring to all the various causes of the behavior, both external and internal, we don't get a complete explanation of the behavior – we get an incomplete causal picture.

1.3 The Environment and Human Development: Reaching

Thelen (1995) gives us an account in similar terms describing how the environment affects infants learning to reach for objects. Infants begin reaching – targeted movement of the arms designed to grasp an object – at different times, although usually this is not a late development. In the study by Thelen, et. al. (1993) of four children, two first reached around 3 months, and two reached around 5 months. Hannah, who reached at 22 weeks, was a fairly passive child. Before the onset of reaching, most of her arm motions were very small, keeping her arms close to her body. Gabriel, on the other hand, was very energetic, arms flapping wide of the body on wild paths. Gabriel began to reach at 15 weeks. For both children, the first instance of reaching occurred spontaneously during a normal arm movement, and appeared to be an adjustment of that movement.

1.3.1 Analysis of the Experiment and the Explanation

Notice that these two infants had very different problems to solve in order to reach a toy placed in front of them. Hannah, whose motions were tiny and without force, needed to learn how to increase the energy of her movements to extend her arm. Gabriel, on the other hand, had to learn how to control his wild swinging motions in order to bring his hands to the toy – he had to learn how to reduce the energy of his swings. The gross behavior – reaching – was the same for each infant, but the actual solution to that problem was very different for each child.

What Thelen discovered was that the arm movements of the infants prior to reaching were excellent predictors of their reaching behaviors. More importantly, it turns out that the transition from non-reaching arm movements to reaching modeled a shift in the energy parameter in an equation describing an oscillating spring. Gabriel's arm motions in particular were very similar to the behavior of a forced-mass spring, and his reaches and his previous movements are modeled quite nicely by such a system. The task for each child was, in essence, to modulate that energy parameter from its normal values to achieve reaching.

In most respects, this example is much like the last. The motor cortex of the children, responsible for controlling their previous motions as well as their reaching behaviors, plays both causally external and causally internal roles. The structures developed by the patterning of previous movements are causally external insofar as they constrain the movements of the children's arms to certain patterns, but are causally internal insofar as they provide the basis for the reaching movement. Once again, the *alteration* of a pre-existing structure qualifies that structure as an internal cause. The example is different, though, insofar as Hannah and Gabriel both shifted permanently into a new developmental stage, whereas the infants in the previous experiment showed a temporary adaptation to their environment.

The structure of the infants' arms is another structural cause of the reaching behavior. The way the arms are structured for each infant varies the required amount of energy for reaching, and the way the infant has to modulate the energy input to the "system" in order to produce a reach. Their arms do in fact work like springs, and so the final reaching behavior had to resemble the action of a spring. In this experiment the task was not one of modifying the coordination of an already learned pattern, but of modifying the energy input to an already patterned system.

Any explanation of Hannah and Gabriel's reaching behavior that didn't explain the causal roles of the infants' arm structure, their neural structures, and their perceptual shift, would certainly be incomplete. To say that Gabriel reached the way he did because his arm operated like a spring ignores the influence of his previous arm movements and his desire to reach instead of flap. To say that Hannah reached the way she did because her previous movements were slow and controlled and because she had a perceptual shift doesn't offer the key factor that explains the *shape* of arm movements – the structure of the arm. Even with these three factors, unless we explain the perceptual shift each of the infants underwent, we have not completely explained why infants reach the way they do. But these three are essential, and they cross the boundary between the organism and its environment.

1.4 The Environment and Everyday Cognition: Speech

J. A. Scott Kelso has also provided us with some interesting data about the role the environment plays in cognition. In Kelso (1995), he reports two sets of experiments from Kelso et. al. (1982) and Kelso et. al. (1984), concerning the effects of the environment on speech. His experiments were designed to test a hypothesis about the nature of cognition, and we will go more deeply into his conclusions in that regard in Chapter 2. What should interest us most at this point is the nature of the experiments involved, and how they illustrate the interplay between agent and environment that produces cognition.

In these two studies, Kelso et. al. decided to study the ability of the human speech-production system to adapt to spontaneous and novel shifts in environmental constraints.[5] To do this, they placed a device on the experimental subjects designed to freeze the jaw in place, on command, for a period of a few milliseconds. They then attached various sorts of sensors to various parts of the speech system (major muscles, the lips, and the jaw) to see which parts of the system were doing what at various points in the experiment. The goal was to study how the various parts of the system responded to the perturbation of the behavior of one part: the jaw.

What Kelso et. al. found out was that the vocal system responded quite appropriately to this novel change in the environment. Instead of fumbling completely, other elements of the vocal system compensated to produce the desired sound.

> When we suddenly halted the jaw for a few milliseconds as it was raising toward the final [b] in [b ae b] (rhymes with lab), the upper and lower lips compensated immediately so as to produce the [b] but no compensation was observed in the tongue. Conversely, when we applied the same jaw perturbation during the final [z] in the utterance [b ae z], rapid and increased *tongue* muscle activity was observed exactly appropriate for achieving the tongue-palate configuration for a fricative sound, but no active lip compensation. *Kelso (1995), p. 40.*

Such rapid adaptation of a learned behavior, in response to changes in the environment, suggests that we are designed such that we are constantly adapting to our environment; we even accommodate small, sudden changes.

1.4.1 Analysis and Explanation

It looks as if this is a nice illustration of the interplay between environment and cognition. The neural solution we have found to speech is one that directs the muscles involved not towards some rigid goal-state – if it did, we could not adapt to novel perturbations – but rather directs them as a system towards the production of a sound. The muscles themselves, individually, could not adapt to sudden perturbations; the adaptation must take place at the neural level, unless we wish to allow for cognition at the level of musculature. This is real-time neural ⇔ environmental interaction, with the environment determining the solution our brain reaches to the task we have set it. This is the interplay between cognition and the environment I have been attempting to illustrate.

This study offers a perfect opportunity to introduce the notion of *soft assembly*. As Clark (1996) notes,

> A traditional robot arm, governed by a classical program, provides an example of "hard assembly". It commands a repertoire of moves, and its success depends on the precise placement, orientation, size, and other characteristics of the components it must manipulate. Human walking, in contrast, is soft-assembled in that it naturally compensates for quite major changes in the problem space. *Clark (1996), p. 42*

A soft-assembled system is organized around some central goal, like walking, and recruits whatever resources are available and useful to the task of achieving that goal. Walking a tightrope and walking on a sidewalk require very different sets of muscles, and walking with bare feet over rocky terrain requires a different gait and different patterns of muscle control than walking in shoes over a plush carpet. We use whatever resources are necessary to achieve our goal of walking in each of these cases, but the end result is still walking.

---

[5] For a nice but brief exposition of the experiments, see Kelso (1995), p. 39-41.

Similarly, in Kelso's example, we have several different components in the articulatory system, and each is co-opted for the gross goal of speech. But whatever cognitive solution we have to the problem of speech doesn't specify a rigid behavior for each of these components. Rather, as a component is prevented from fulfilling its individual goal, the other components of the system compensate to achieve the overall goal. This suggests that the articulatory system is soft-assembled, and that the solution to the task of speech is a decentralized one, not dependent on any one component, but instead on complex feedback loops between components.[6]

We can see this for, if it were that each followed its own explicit set of rules, the adaptations evidenced could not have occurred. In the first case Kelso mentions (freezing the jaw during production of [b ae b]), the lips responded to the shift *in the jaw*. In the second case (freezing the jaw during production of [b ae z]), the tongue responded, but not the lips. In both of these cases, the response was to a *system-wide* goal, not to the goal of the individual component. Each of the components of the system relies on feedback from the others; the lips need to know what the tongue and jaw are doing, the jaw what the lips and tongue are doing, and the tongue what the lips and jaw are doing, in order for each to act appropriately.

What we learn from identifying this system as soft-assembled is that a useful explanation of the production of speech cannot advert merely to the behavior of individual components, but must discuss their interactions. At this level of interconnection, there isn't really any use in distinguishing the systems controlling each of the elements from each other, since the *useful* processing – the processing that produces speech – arises as a result of the interconnection of the elements.

1.5 Analysis and Conclusions

[6] It should be noted that the components here include not just musculo-skeletal features, but the nerves, etc. that constitute sensors for the purpose of feedback.

In this Chapter we have looked at several instances of cognitive behavior, and seen that the causal story explaining each behavior was incomplete without some discussion of the environmental factors involved in cognition. A baby learning to walk is affected as much by leg mass (and hence gravity) as by motor cortex control. A baby kicking adapts itself to its environment on the fly, with the sole explanation for the change in behavior lying in the change in the environment. A baby learning to reach adapts pre-existing neural structures to handle a new task, and is constrained by those structures and by the mechanical nature of its arm. Human adults attempting to produce simple monosyllables use a soft-assembled system to reliably produce speech in the face of environmental interference.

In each of these cases I have made it clear that, without discussing both external and internal factors, we are missing an important part of the explanation of the behaviors involved. This suggests that any complete cognitive scientific model of cognition will have to address both internal and external factors. As Andy Clark (1996) puts it, we will require "an account of the gross behaviors of the well-functioning organism in the environment – an account that may invoke collective variables whose componential roots span brain, body, and world".[7] This sort of explanation would properly refer to the causal roles of both internal and external factors (in our sense).[8] Clark has a particular sort of explanation in mind, though, as is clear from his mention of "collective variables". A collective variable is a variable whose value represents some property of a system, not of an individual component.

Perhaps the easiest example of a collective variable is heat. When we discuss the heat of a glass of water, we are referring to the mean molecular kinetic energy (MMKE) of the water as a system. Whatever value we have for the heat of the water is not guaranteed to correspond to the actual molecular kinetic energy of any one molecule in that glass, but is rather a property only of the aggregate of water molecules in the glass. We refer to the heat of the glass, though, because *that is what we are interested in*; it is not useful to

[7] Clark (1996), p. 126.

know the molecular kinetic energy of each particular molecule. Similarly, when we explain the production of speech, we are not interested in the behavior of the discrete components of the articulatory system, but in the behavior of the system as a whole. Or, more precisely, the explanation of the behavior of each component relies on the behavior of every other component.

So now that we have seen that we need explanations that advert to both internal and external factors, we are able to ask what sort of approach – what method of modeling – we should be using in cognitive science. It should be clear that an approach that can't provide this sort of explanation will never be able to offer a complete explanation of cognitive behavior. In the next Chapter we will compare two types of approaches that could possibly provide this type of explanation, and see which of them best satisfies the explanatory demands of cognitive science. Then, in the final Chapter, I will discuss the limits of the approaches we saw in Chapter 2, and argue that they result from the ineliminable need for two other types of explanation that can only be provided by participation in the cognitive scientific endeavor of the biological sciences.

Chapter 2

2.0 Computationalism and Dynamicism

Now that we have established that a completed cognitive science will require explanations in terms of properties at multiple levels of abstraction and in multiple physical locations, it remains for us to examine the current contenders for an approach to do just that. In a sense, I wish to frame this comparison of computational and dynamicist models as conceptually opposite the analysis in the chapter 1. Instead of attempting to show that a certain type of explanation is required by cognitive science, I will be arguing that we are not justified in claiming that another type of explanation is not. Instead of arguing for a specific type of explanation, I will be arguing that we must not limit our explanations without good reason.

The root of this limitation lies in the assumptions implicit (and sometimes explicit) in computationalism. In their (1995), Horgan and Tienson present an analysis of computationalism that reveals 5 essential assumptions:

> (1) Intelligent cognition employs structurally complex mental representations.
> (2) Cognitive processing is sensitive to the structure of these representations (and thereby is sensitive to their content).
> (3) Cognitive processing conforms to precise, exceptionless rules, statable over the representations themselves and articulable in the format of a computer program.
> (4) Many mental representations have *syntactic* structure.
> (5) Human cognitive transition functions conform to a *tractably computable* cognitive-transition function.
>
> *Horgan and Tienson (1995), p. 24-5*

On their analysis, for dynamicism to be a valid alternative to (and not merely a special case of) computationalism, the dynamicist must reject at least one of these assumptions. In particular, they argue that the dynamicist should reject assumptions 3 and 5.

I should make it clear that Horgan and Tienson do not accurately describe the whole of the computationalist movement in either cognitive science or artificial intelligence. While most computationalists will agree with assumption (1) that representations are important, some (e.g., Simon 1969) claim that we require simple representations, not the complex representations that Horgan and Tienson themselves favor. Likewise, while most computationalists will agree with assumption (3) that cognitive processing conforms to rules statable over representations and articulable in the form of a computer program, the restriction to exceptionless rules would be considered by many to be too strict. The recognition that cognition seems to be filled with exceptions and novel behaviors drives much of the research into computational models, under such guises as “non-monotonic reasoning” and “circumscription”.

The arguments in this chapter will rest, though, on the parts of Horgan and Tienson’s description of computationalism that will be accepted by most computationalists. My first goal in this chapter is to make clear precisely why the computationalist must hold on to assumptions 3 and 5, with the caveats I have noted, and why it is that *cognitive scientists* – not just dynamicists – should not. The reasons lie in the nature of a dynamicist explanation of behavior, and more importantly, in the mathematical characterization of behaviors that dynamical systems theory can provide. Of course, this will not be a conclusive argument for dynamicism. That certain assumptions of computationalism are not warranted at this time, and that dynamicism does not have to make those assumptions, does not establish that dynamicism is the *correct* view. However, this argument does mean that employing dynamicism is preferable to sole reliance on computationalism.

My second goal, then, is to provide some reason to think that cognitive scientists should be dynamicists. Here the results of chapter 1 will come into play, as I argue that dynamicism is well equipped to provide explanations spanning the brain, body, and world. To do this I will return to experiments, and make clear precisely how the explanations of the behaviors in each case are *dynamical* explanations. I will then generalize the advantages of employing a dynamicist perspective in those cases to the task of providing cognitive scientific explanations in general.

2.1 Computationalism, Dynamicism, and Tractable Computability[9]

Tractable computability, as opposed to strict computability, is the subject of our debate for a simple reason. Consider for a moment what it means for cognition to be computable. Given the total cognitive state (TCS) s of an agent at some time t, there is some function F for which $F(t+1,s,X)$ yields the TCS of the system at time $t+1$, where X is the set of inputs the agent receives at time t. Horgan and Tienson label this function the *Cognitive Transition Function* (CTF). The computationalist claims that the CTF for every cognitive function, and indeed for describing the evolution of any agent's cognitive states over time, is computable. Minimally, this is the claim that there is some strictly computable function that is the CTF for a given agent.

Of course, humans are finite beings. Our brains have finite resources for computation, and finite time to compute in. So it must be the case that, if our CTF is computable, it is computable in the relatively short time it takes us to change our TCS, and it must be computable with the computational resources of our limited brains. So, minimally, the computationalist *must* have in mind a notion of *tractable computability*, since our brains cannot be instantiating CTFs that they cannot compute in the time that they do in fact compute them. For this reason, Horgan and Tienson are right in characterizing computationalism with assumption 5.

Typically, we are given one of two reasons to assume that human CTFs are tractably computable. The first is possibly the most annoying: What else could cognition be but computation? The second reason is, simply, that we have some evidence in favor of the

[9] My thanks to Horgan and Tienson for providing an excellent framework within which to discuss computationalism. Many of the arguments in this section are either offered by or inspired by their (1995), chapter 2.

computational view. To answer the former, let me digress for a moment into a discussion of the mathematics of dynamicism.

2.1.1 Modern Dynamical Systems Theory

Since much of the argumentation in this chapter depends on the mathematics of dynamical systems theory, I will spend some time now articulating its core concepts. The reader familiar with dynamical systems theory may wish to skip to the next section.

A dynamical system is any system that changes over time. For our purposes, a dynamical system is a mathematical entity, a description of some system in terms of a set of differential or difference equations that specify the state changes in the system over time. A dynamical system that evolves over discrete (incremental) time is governed by difference equations, while one that evolves over real (continuous) time is governed by differential equations. In cognitive science, the latter type is more important, since the dynamicist views the continuous evolution of cognizer and environment in real time as essential to the explanation of cognition. For our purposes, then, a dynamical system is described by a set of differential equations.

A system (in the loose sense) will be described by one or more state variables. That system will then be modeled by a system of differential equations, each specified in terms of the evolution of a single state variable over time. The *state space* of a dynamical system is the space with an axis for each state variable, upon which the value of that variable at any given time can be located. The state of the system at any time *t* is described by an ordered *n*-tuple locating the system in that state space. The *phase space* of a dynamical system is the space formed when an axis for time is added to the state space; it is the space in which we can chart the changes in the state of a dynamical system – the *trajectory* of the system – over time.

In general, the strength of dynamical systems theory is the analysis of unsolvable systems of differential equations. While all linear systems of differential equations are solvable, problems arise when we try to discuss nonlinear systems. As Norton (1995) says,

> If a system is not linear, all bets are off: typically nonlinear equations cannot be solved explicitly. Nonlinear equations are important because (1) most systems are modeled by nonlinear, rather than linear, equations, and (2) solutions can have complicated and interesting behaviors, requiring various qualitative techniques of dynamical systems to analyze. *Norton (1995), p. 29*

The qualitative techniques Norton mentions consist primarily in analyzing the state space of the dynamical system, to discover how a system is likely to evolve from its current status.

A dynamical system evolves – changes its location in its state space – in response to one or more attractors. *Point attractors* are single points in state space where the system will remain until perturbed. *Periodic attractors* are "orbits" in state space that repeat, so that a system that evolves to a periodic attractor will endlessly repeat the same sequence of states, unless perturbed. *Quasi-periodic attractors* cause the system to evolve along a trajectory asymptotic to a periodic orbit; a system evolving in response to a a quasi-periodic orbit will not be repeating the same sequence of states, however. *Chaotic* (or *strange*) attractors cause the system to evolve through an infinite, non-repeating series of states that are not asymptotic to a periodic orbit.[10]

Given a set of differential equations describing a dynamical system, we can attempt to identify from those equations where the attractors lie in the state space of the system, and what types of attractors they are. By identifying the basins of attractors – the sets of points from which the system tends to evolve towards particular attractors – we can construct an attractor landscape. A dynamical system at some point *X* will evolve in response to every attractor in whose basin it resides, and by identifying the attractors (and

[10] Horgan and Tienson (1995), p. 47-8.

their strengths, types, and their distance from the current state of the system) we gain insight into how the system will evolve. This analysis is qualitative because we have not arrived at a *solution* to the system by describing its evolution, but instead have a characterization of its tendencies.

Sensitivity to initial conditions is the hallmark of a dynamical system, and plays a large part in the uncertainty of this sort of analysis. Much like a ball rolling across an uneven surface, the state of the system will change depending not only on where it is currently in its state space, but where is *was*. So by analyzing the attractor landscape of a dynamical system, we can determine which trajectories through phase space the system is likely to take from its current location.

For a more comprehensive *introduction* to mathematical dynamical systems theory, see Norton (1995). A more advanced text, for those already familiar with the general principles of dynamical systems theory but wishing to enhance their knowledge, is Katok and Hasselblatt (1995).

2.1.2 What Cognition Could Be That Isn't Computation

The picture we get of cognition on a computational model is that of a fixed CTF that describes the changes in our TCS over time. There is no prima facie reason that this picture of cognition isn't precisely what we would get from dynamicism. After all, a dynamical system is a system whose movement from one total state (one point in its state space) over time is described by a set of differential equations. It isn't much of a stretch to view the position of the system in phase space as the TCS of the system at a given time, or the differential equations describing its evolution as a CTF. So how is it that dynamicism gives us something other than computation?

There are several differences, all of which relate back to the differing claims of computationalism and dynamicism. Modern dynamicism rests on what Van Gelder and

Port (1995) label the *Dynamical Hypothesis*: Natural cognitive systems are dynamical systems, and are best understood from the perspective of dynamics.[11] Van Gelder and Port (1995) suggest that computationalism is characterized best by the *Physical Symbol System Hypothesis*: Physical symbol systems (computers) are necessary and sufficient for intelligent behavior (Newell and Simon, 1976).[12] The dynamicist wants to claim that cognition actually is the operation of a dynamical system, and the computationalist wants to claim that cognition actually is the operation of a computational system.

In a reply to Horgan and Tienson, Litch (1996) argues that this difference isn't as important as it seems. Her point is that a system might be best described as a dynamical system at one level, but as a physical symbol system at yet another (lower) level – or vice versa. She thinks the real difference between the dynamicist and the computationalist is over the real-valued conception of time that the dynamicist employs and that the computationalist cannot (since digital computers can only have discrete, not continuous representations). I am not certain that this is really a problem for the computationalist, unless Litch wishes to make the claim that we require infinitely precise time and not just extremely precise, though finite, time (see Van Gelder and Port, 1995 for a view sympathetic to Litch).[13] I think that there is, nonetheless, a conceptual difference between dynamicism and computationalism.

A dynamical system can, after all, evolve in response to a chaotic attractor. A system evolving in response to such an attractor will have a transition function that is not articulable in the form of a computer program, and that may not be tractably computable. There will be no finite set of rules defined over a finite set of symbols that can represent all the state changes of the system. This is because a system evolving in response to a chaotic attractor will evolve through a non-finite, non-repeating sequence of states. The

[11] Van Gelder and Port, p. 5.
[12] Ibid.
[13] Essentially, to say that a computationalist cannot handle real-valued time requires Litch to show either (a) that cognition requires infinite precision, or (b) that there is some bound on computers we build (as opposed to brains) that makes it impossible for them to compute within the time scales the brain computes in

computationalist, then, has to claim that human cognitive systems do not *ever* evolve in response to chaotic attractors.

The conceptual difference I am aiming at between computationalism and dynamicism is, then, that dynamicists neither restrict nor commit themselves to non-chaotic models. The dynamicist does not claim that the operation of the brain conforms to the operation of rules over symbols, and so is not bothered by our inability to capture the CTF for a given cognitive system as a computer program. The dynamicist can instead suggest that the brain is instantiating a dynamical system evolving in response to a chaotic attractor, if in fact a given cognitive process appears to be chaotic. The description of the system consists in a set of differential equations, of which one or more (or the interaction of several) is chaotic.

This does not provide the computationalist with a back-door into assumption 3 – a set of differential equations describing the evolution of a system does not, in itself, constitute a computer program. To comply with assumption 3, we must provide our computer with a means to solve these differential equations. This is what we have just shown to be the problem: We can specify programs to solve many non-linear differential equations for special cases, but quite often we either lack the knowledge of methods for solving sets of non-linear differential equations, or the processing power to implement those methods. Moreover, with a chaotic system, we will have infinite special cases, and hence require a program of infinite length.

Likewise, this poses a problem for assumption 5 of computationalism. We may very well discover a method for solving a chaotic system in some special case, but that method need not be one computable within the constraints of human cognition. It is certainly possible that the only method for solving a chaotic system of equations in finite time requires computing resources beyond what we are provided with by nature. If this is the case, then a human brain instantiating this system would not be computing, since the computational solution to this system could not be instantiated in (or as) a human brain.

What we have shown here is merely that it is *possible* to articulate a conception of cognition as something besides discrete, tractable computation. We move now to a discussion of whether we have any reason to suppose that our CTF is in fact computable. We will consider arguments designed to show that the brain is actually computing, and arguments designed to show that it might not be, and evaluate their various strengths and weaknesses. We will then decide whether we have good reason to dispense with assumptions 3 and 5 of computationalism, and advocate a dynamical approach. From here on out, though, we are assuming a notion of tractable computability, not computability in principle.

2.1.3 Why Computationalism is Correct

There are two sorts of evidence offered in favor of a computational model of the mind, both based on our *experience* of cognition. The first is prima facie evidence: sometimes we really have an experience of computation, as when we perform arithmetic in our heads. The second sort of evidence involves an inference from the results of our cognition – specific sorts of behaviors we frequently engage in – to the functions underlying those behaviors. Let us focus our attention on each of these sorts of evidence, in turn.

Van Gelder (forthcoming) gives a useful example of and response to the first sort of evidence. Suppose you were to add 25 and 44 in your head, and carry out explicitly the column-wise addition necessary to arrive at the result of 69. It certainly looks like you are computing! After all, you imagine the numbers, you manipulate them symbolically according to a computable algorithm, and you reach a solution. Moreover, this is an experience everyone can have, with only the difficulty of the problem being adjusted. But is this good evidence that our cognitive functions are computable?

Ryle (See Ryle, 1959) motivates the response Van Gelder gives. Imagine you are looking at the Eiffel Tower, that you have a very detailed image of it in your head. Now, close your eyes and examine that image. It *looks* like the Eiffel Tower – but it does not follow from this that you have the Eiffel Tower in your head. In the same manner it looks like you have a particular algorithm (that of column-wise addition) in your head – but it does not necessarily follow that you have that algorithm in your head. It merely appears as if you do, regardless of the actual state of affairs in your head.[14]

So the first sort of evidence, relying entirely on introspection, runs up against a common problem with introspection: we are not guaranteed that our mental processes are anything like they appear to us. The second sort of evidence, though, proceeds from external behavior with a little abstraction. A good example is Chomsky's productivity argument: we are able to produce an incredible variety of sentences in our native tongues, and we could not do this if the process producing these sentences were not computable. More generally, we could not store all of these sentences individually, so we must generate them as needed – in a computable manner.

This type of evidence, as well as the previous type, fails to establish the computability of cognition for a simple reason. Even granting the productivity argument (or that we are computing when we perform mental arithmetic), it is not clear that *all* cognitive processes are computable. The productivity argument establishes at best only that *some* cognitive processes are! It is not even clear that it establishes this weak conclusion, though, since it isn't clear why a non-computable function could not suffice to produce the multitude of sentences that we can produce.

[14] Ericsson and Simon (1984) argue for the validity of the confirmation of computational models based on verbal reports. Their technique is to produce programs that produce output relevantly similar to humans, given certain data, and the similarity of outputs is offered as a proof of the accuracy of the description provided by the program. This is precisely the sort of case Ryle's argument attacks: We are given no evidence that humans *in fact* produce their output as a result of the same algorithm the computer uses, but only evidence that the algorithm the computer uses is an adequate description at a high level of abstraction.

So the evidence in favor of a computational model of the mind is weak, at best. The strongest conclusion warranted is that some cognitive processes are computable, and even this conclusion is not well-founded. The obvious next question to ask is whether there is evidence to support the use of an approach that can handle non-computable as well as computable cognitive functions.

2.1.4 Why Assumptions 3 and 5 Are Not Warranted

In his (1996), James Garson argues that the brain is not only capable of, but likely to employ, chaotic (non-computable) processes. While there is preliminary research data to indicate that there are chaotic processes in the brain, the research is just that – preliminary. It isn't clear whether the chaotic patterns observed in the brain belong to one process that is chaotic, or are the result of multiple non-chaotic processes interacting, or are limit cycles with noise. Nor is it clear that this chaotic behavior is necessary for cognition, rather than an incidental byproduct of it. However, Garson does offer some reasons independent of this research to believe that cognition is (at least in some instances) non-computable.

The first reason he offers is somewhat weak, though interesting. It seems that most systems exhibiting certain characteristics tend to be chaotic, and the human brain (in all its various forms) exhibits two of these characteristics: "(1) subunits (neurons) with non-linear dynamics; and (2) recurrent (loop-forming) connections between the subunits".[15] So there is a (weak) prima facie case for the brain's employment of non-computational processes, and hence for the rejection of a computational model in favor of a model that can handle computable and non-computable functions. [16] Dynamical systems models fit the bill exactly.

[15] Garson (1996), p. 308.
[16] While it is important to note that the behavior of any dynamical system can be described by an algorithm, it is also important to recognize that not all algorithms are tractably computable. An algorithm describing the evolution of a dynamical system evolving in response to a chaotic attractor would have infinitely many steps, and hence be non-computable.

Garson's second reason for preferring a dynamical systems approach to a computational approach is a bit stronger. While the set of computable functions and the set of non-computable functions are both infinite sets, the set of computable functions is vanishingly small compared to the set of non-computable functions. Since section (3.1) established that we don't have any real reason to suppose that all cognitive functions are computable, and we have just seen that we have some (weak) reason to suppose that some cognitive functions are non-computable, a mere comparison of the sizes of the sets of non-computable versus computable functions would suggest the brain would more likely arrive at non-computable than computable solutions to problems.

So in the absence of any strong evidence on either side, it seems reasonable to proceed with a model that can handle both computable and non-computable functions. Bear in mind that if we use a dynamical systems approach, we lose no explanatory power. A dynamical system can implement any computable function[17], so if it turns out after in-depth investigation that the brain does not employ non-computable functions, it would not render a dynamical systems approach invalid. However, if it turns out (as Garson suggests it will) that the brain employs non-computable functions, computational models of the mind will be shown invalid by the falsification of assumptions 3 and 5.

2.2 Dynamical Models and Situated Cognition

In chapter 1, the first example of a case of situated cognition we discussed was from Thelen and Smith (1994), concerning the development of walking in infants. The study analyzed a number of movements that infants produce that are kinematically similar to walking, and the different developmental stages at which those movements appeared. What they discovered was that movements kinematically similar to walking appear at almost any stage of development after 1 month, as long as certain factors in the infants' environment are altered. The factors that Thelen and Smith isolated as key to

[17] See Horgan and Tienson (1995), Chapter 4 for an argument to this effect.

determining whether an infant could produce those motions or not were leg mass and muscle tension, and their explanation of walking relied on treating the legs as a spring-and-pulley system whose systemic constants changed as the infant grew.

Leg mass and muscle tension are factors that are known more technically as *collective variables*. The tension of the leg muscles, and the mass of the legs, while being derivative of many components, are represented by a single value or a small set of values. The mass of the infants' legs changes over time and in response to environmental changes. The infants muscle tension changes according to signals the muscles receive from the central nervous system, and in response to changes in the environment of the infants. In certain instances, the muscle tension is dependent on factors like the backward pulling of a treadmill *and* the neural impulses telling the muscle to contract. So the collective variable "muscle tension" represents the influence of numerous factors in the brain, in the body, and in the world.

It's now time to let the cat out of the bag. Each of the four experiments detailed in Chapter 1 were attempts to develop dynamical systems explanations of behavior. In the case of the three experiments from Esther Thelen and her various colleagues, the goal was to provide a dynamical systems explanation of the behavior involved. (In fact, these were also tests for dynamicism – if no satisfying account could be constructed on dynamicist terms of these behaviors, then it would be a serious blow to dynamicism.) The experiment by Kelso, et al (1982) was designed to test the hypothesis that speech was a task showing characteristics indicative of *synergy* – a notion that later led Kelso to dynamical systems theory.

Our job in this section is to examine dynamicism as characterized by the Dynamical Hypothesis (see section 2.1.2, above) to determine what makes it such a natural candidate for providing explanations involving factors in multiple locations and at multiple levels of abstraction. I will be highlighting the key concepts of collective variables and *coupling*, by examining one more dynamical model of a cognitive task. It is my contention that each of these concepts contributes to the explanatory power (and coherence) of dynamicism, and that we should look for the ability to deal with these concepts within any explanatory framework we adopt for cognitive science.

2.2.1 The Nature of Dynamicist Explanation

Recall the experiments by Kelso (1982, 1984) we discussed in section 1.4. Kelso, et al discovered that perturbing the motion of the jaw when it was drawing towards closure in the production of [b ae b] or [b ae z] caused compensation by the other components in the articulatory system. As Kelso notes, these compensations were extremely fast, generally within 20-30 milliseconds of perturbation. According to Kelso, this suggests that the components of the articulatory system are highly interconnected – as Saltzman (1995, p. 157) describes the results, they suggest that "there is some sort of automatic "reflexive" organization established among the articulators with a relatively fast loop time". Kelso also concluded that the manner in which those components compensated for the perturbation suggests that the specific compensations were not hardwired.

Saltzman (1995) took the results of these two experiments, as well as other research (e.g. Saltzman, 1992; Kelso, et al, 1986), and developed a dynamical systems model of speech articulation. The data suggests that there is no hard-wired solution to a given perturbation, since the system compensates for novel perturbations with goal-dependent responses. So Saltzman developed a model of speech that assumes the relevant processing goes on at the level of the system, with an end-state defined in terms of a task goal. This sort of model is, usefully, labeled *task dynamical*.

Saltzman's model posits certain variables corresponding to the aperture of various portions of the vocal tract during the production of speech – *tract variables*. Each of these tract variables corresponds to the movement of different *sets* of articulators in the system. So, for instance, the tract variable LA (for lip aperture) "defines the degree of bilabial constriction, and is defined by the vertical distance between the upper and lower lips".[18] As you can see, the tract variables are collective variables – they represent the behavior of numerous parts of the articulatory system (and their subunits) with a single value.

On this model, the tract variables are each associated with a set of *model articulator* coordinates.[19] In the model articulatory system Saltzman uses, each tract variable specifies by its value coordinates for the model articulator it is associated with. As Saltzman says, "the model articulators are controlled by transforming the tract-variable dynamical system into model articulator coordinates".[20] What happens when this transformation occurs is that the different and discrete model articulators are implicitly *coupled*. This means that the behavior of each articulator becomes influenced by the behavior of every other articulator.

Tract variables and articulator coordinates form the heart of Saltzman's model. Briefly, his model behaves perfectly in accordance with behaviors observed in the experiments already mentioned and with others well-known throughout the literature on speech production. His model explains the behaviors in a very abstract manner, by glossing over the details of articulator movement to get at the causally relevant features of the movement (with respect to the task at hand). This is one strength of dynamicism; the explanation of articulation Saltzman offers us isolates the relevant aspects of the behaviors of components that combine to achieve the gross behavior.

This model also demonstrates another strength of dynamicism, the ability to *couple* discrete systems. For any pair of articulators in this model, there is (implicitly) a function describing the influences each has on the other. So while we can discuss the behavior of each articulator independently, we can then construct a full model of the system by employing these coupling functions. In essence, Saltzman has provided us with a model of each subsystem of the articulatory system in the form of a tract variable, and the

[18] Saltzman (1995), 160.
[19] Ibid, 161.
[20] Saltzman (1995), p. 161.

mapping of tract variables to model articulator coordinates applies a coupling function to describe the system as a whole. Interestingly enough, because Saltzman's model is a mathematical dynamical systems model, it can be readily coupled with Browman and Goldstein's *articulatory phonology*, which is also a mathematical dynamical systems model, to provide a broader mathematical model of speech production (Browman and Goldstein, 1995).

The ability to provide explanations at the level of collective variables, and the ability to couple dynamical systems models, provides the rather natural fit between dynamicism and explanations that advert to causally internal and external factors. The potential to offer models of cognition that can be coupled with models of the environment is certainly one reason dynamicism is currently *en vogue*.

2.3 Conclusions

So we have seen that dynamicism looks to be a natural candidate for providing the sorts of explanations we saw were necessary in Chapter 1. Through the use of collective variables, which allow us to abstract from the details of implementation, and the use of coupling, which allows us to combine models of independent systems to create a coherent model of the system they form together, dynamicism allows us to discuss causally relevant factors at multiple levels of abstraction simultaneously. In fact, if we want a mathematical characterization of behavior, it is difficult to conceive of another method for achieving this unity between levels of abstraction.

Likewise, we have seen that dynamicism allows us a measure of explanatory freedom. Dynamicism, unlike computationalism, does not restrict us to discussions of tractably computable CTFs articulable as programs. We saw that limiting ourselves to explanations without chaos seems shortsighted, or at least unwarranted at the moment. We also saw that we have no reason to assume that the brain is in fact always computing when it is engaged in cognition. While it is possible that we could provide useful and informative computational models of some cognitive processes, there is no guarantee that we can provide such models for all cognitive processes. Given our current lack of knowledge concerning the actual neurobiological realization of cognition, right now we simply do not have reason to limit the explanations we seek to computational explanations.

So, with dynamicism looking so good, it's time to evaluate its limits. In the next chapter, we will discuss precisely why dynamicism cannot offer us every sort of explanation we desire. I will be arguing, though, that this is not a flaw with dynamicism, or a reason to abandon it. Instead, I will argue that because of the nature of the explanations dynamicism provides, it cannot provide certain other types of explanations that are necessary within cognitive science. I will in fact be arguing that dynamical explanations should be complemented by traditional neurobiological explanations.

Chapter 3

3.0 Top-Down vs. Bottom-Up Cognitive Science

In chapter 1 we identified the need in cognitive science for explanations of behaviors in terms of factors at multiple levels of abstraction and in multiple physical locations. In chapter 2 we saw that, in addition to being well-suited to providing these sorts of explanations, dynamicism is less restrictive of the sorts of explanations we can offer than computationalism. The question we will be asking in this chapter is whether dynamicism alone can provide *all* the sorts of explanations we require in cognitive science. In particular, we will be asking whether a *pure* dynamical systems approach can operate without input from the traditional biological sciences.

Randall Beer follows what might be construed as a pure dynamical systems approach. In his work modeling autonomous agents, he begins by constructing a dynamical characterization of single-neuron systems. He then studies multiple-neuron systems as coupled single-neuron systems. In this way, he proceeds from a very low level of abstraction – an explanatorily primitive single-neuron dynamical system – to dynamical systems modeling ever more complex organisms (e.g., Beer, 1995).[21] An approach like Beer's (but applied to the study of actual human cognition) recognizes that remaining true to our neurobiology is important (in some sense), but seeks to replace traditional neurobiological explanations with dynamicist explanations.

Andy Clark, in his (1996), argues that such an approach loses sight of other important goals of cognitive science. Instead of seeking only to explain the gross behaviors of an organism, we might very well like to know what parts of the organism are responsible for what parts of the behavior. Saving a more detailed analysis for later, consider what sorts of explanations we seek from (and are often offered by) psychopathology. When a

[21] Interestingly enough, in his (1995) and elsewhere, Beer follows the modeling guidelines of W. Ross Ashby (1960), modeling the co-evolution of an agent and its environment as two coupled continuous-time dynamical systems. (See, e.g., Beer, 1995, p. 130.)

patient is diagnosed with a developing lesion in area V4 of the visual cortex, it is of some value to be able to tell the patient what functions will be impaired. We often find it useful, for a variety of reasons, to assign various information-processing roles to neural structures.

Recognizing the value of dynamicist explanations, though, Clark advocates an approach involving multiple types of explanations, including both dynamicist and information-processing explanations. He ties them together with a requirement for explanations that identify the components subserving higher-level cognitive processes. The result is this list of types of explanation he thinks are required by cognitive science:

> (1) An account of the gross behaviors of the well-functioning organism in the environment – an account that may invoke collective variables whose componential roots span brain, body, and world.
>
> (2) An account that identifies the various components whose collective properties are targeted by the explanations proper to (1). Two important subtasks here are to identify relevant neural components and to account for how these components interact.
>
> (3) An account of the various information-processing roles played by the components (both internal and external) identified in (2) – an account that may well assign specific computational roles and representational capacities to distinct neural subsystems.
>
> *Clark (1996), p. 126*[22]

There is clearly interplay between each of these types of explanation; (2) specifies the components of the explanations in (1), and (3) assigns information-processing roles to those components. Clark's goal is to provide a link between generalizable explanations like those a dynamical model can offer with specific explanations of actual neural (and physical) structures.

In chapter 1 we saw reasons why we need the type of explanation in (1), and in chapter 2 we saw that dynamicism can indeed provide explanations of that sort. In this chapter we will examine the reasons why the types of explanation in (2) and (3) can't be provided without the assistance of traditional neurobiology, and why an approach uniting neurobiology with dynamicism will be, ultimately, more fruitful than an approach employing only one or the other. To do this I will be re-introducing a distinction from computer science into the vocabulary of cognitive science. The distinction is one that cognitive scientists should be somewhat familiar with: the distinction between top-down and bottom-up design methodologies. My goal is to show how the distinction operates with respect to dynamicism and neurobiology, and then demonstrate why the co-evolution of dynamicist and neurobiological explanations will likely prove more fruitful than their independent development.

To this end, in section 3.1 I will develop the top-down / bottom-up design methodology distinction. In section 3.2 I will import the distinction into cognitive science, making clear the differences in its functions in each discipline, which may result from the difference between design and analysis. Then, in section 3.3 I will present arguments for the combination of neurobiological and dynamicist explanations on the basis of the relationship between neurobiology, dynamicism, and the top-down / bottom-up distinction in cognitive science.

3.1 Top-down vs. Bottom-up Design Methodologies

Suppose for a moment that you are given the task of designing a graphical user interface (or GUI). GUIs are in use throughout the software industry now, almost entirely replacing the old text-interface standard. Some of the key goals of the move towards GUIs are achieving an "intuitive" feel, increasing ease-of-use, and making computers seem less "technical". So when designing a GUI, you should consider each of these goals. If you are deciding how to label individual windows, then you should try and come up with the most obvious (while least insulting) way of letting the user know that the label is, in fact, a label.

---

[22] Kelso (1995), pp.43-44 argues for a similar set of explanatory types, and offers a similar list on p. 66.

When you begin to design this GUI, first off you have to decide what it needs to do. Do you need a multiple document interface? (One that allows the user to have multiple documents open and displaying information at the same time.) Will you need to handle a very restricted number of documents at one time (2 or 3), or a large number (20+)? What will the default view of your GUI be, when the user has no documents open? How will you allow the user access to documents? How will they open them from the GUI? These are all questions that will confront you immediately.

You can see the sense in asking these sorts of questions, as well. If you are designing a GUI, it makes sense to know what that GUI has to do before you start coding. After all, if you are designing a GUI for a game, you will likely only need to support a single view (document) at a time, so including support for more would be writing code that would never be used. On the other hand, it is quite useful for a word processor to handle multiple documents at once, and not coding that capability in will diminish the usefulness of your GUI. Knowing the answers to the above questions clarifies your task, and makes it clear exactly what code you have to write. The questions can get more and more specific, too, beginning with "Do I want this GUI to be able to handle multiple open documents?", to "Should this GUI be multithreaded, with each document having its own thread?", to "How should I handle updating each particular thread?".

This is a nice example of top-down design. The developer begins at the highest level of abstraction – the vague concept of a GUI – and proceeds to ask questions with more and more detail, conceivably until the answers to the questions consist of individual lines of code in the GUI. This is, for obvious reasons, the most frequently used (and useful) design methodology in computer science. There are many different design protocols based on this one fundamental notion, but they are all still top-down methodologies.

Applications are rarely designed from the bottom-up. Take our previous example, where you are asked to design a GUI. What is your first question, as a software developer intent on designing this from the bottom-up? At the extreme, there are no questions – you begin coding, and hope eventually that your code results in a GUI. Very few people are actually capable of ignoring the final goal in making plans – or writing code – however, so the usual characterization of a bottom-up design is more sensible, if almost as useless.

Instead of asking questions about whether your GUI needs to support multiple documents, you will start by asking what the program will be doing at the lowest level. Since it is a GUI, it will have to have graphics routines. So you might begin by writing graphics routines that allow you to display text and graphics. More specifically, you might write a routine to specify the color of a pixel on the screen. You then might design a routine that defines a region of the screen, and then one that can use your previous routines to define a region and color it. Then you might design a routine to define certain shapes – letters, for instance – the results of which can be fed to your previous routine to draw a letter on the screen.

As you can imagine, there is a long ladder to climb from these very low-level routines to the GUI you have as your desired result. Theoretically, the ladder in the other direction is just as long – but this ignores an important point. When employing bottom-up design, you don't have a clear picture of the end-state of your design. That is, you don't know precisely what you're building, so you don't know precisely what pieces you need. You might design those routines just mentioned, and at a later stage in development (after you have used those routines throughout your code) realize that those routines are not compatible with the multithreaded multiple-document interface you have just discovered you need.[23] At this point, you have to start over, or go back and revise – or end up with a program that doesn't resemble what you were told to develop.

So this is the distinction between top-down and bottom-up design. In the former case, you define your goal-state so explicitly that you know exactly what you need to do in order to reach it. In the latter case, you begin with little or no knowledge of the steps between the lowest and highest levels of abstraction, and begin building upwards from the bottom with no clear idea of the relations between those two levels. It should be easy to why computer scientists prefer top-down design to bottom-up design.

3.2 Importing the Distinction into Cognitive Science

Since in the last section we considered a task appropriate to computer science, consider now this task: you must model human cognition. While this task is perhaps several orders of magnitude more difficult than the last, it is still – in a sense – a design problem. You are now supposed to design a model of an ostensibly physical system. Now, taking a lesson from the last section, let's attempt to apply a top-down design strategy first. The questions you will need to ask are questions like, "What is cognition?" "What environmental concerns are relative to behavior?" "Is cognition time-dependent?" "How will I map environmental inputs into outputs (of any sort)?" "What counts as an input, and what as an output?"

Some examples of people I want to claim take a "top-down" approach to cognition are Noam Chomsky (e.g, 1975), Jerry Fodor (e.g., 1975), and Ned Block (e.g., 1995). In general, this approach moves from some analysis of human reasoning (or behavior) to conclusions about what the mind / brain must be like. Quite often the brain is seen as some kind of processor, and the "mind" is the program currently running. People taking a top-down approach to cognitive science generally try to infer the nature of the program, and then worry about the implementation.

The productivity argument, offered by Chomsky and others for over twenty years now, runs something like this:

1. Humans can produce theoretically infinite grammatically correct sentences in a language.

---

[23] Just as easily, you might realize you don't need the routines at all, and that they are irrelevant to your goals.

2. To do this, we must have some production system employing rules to produce that language.
3. Those rules must be represented in some language to be used.
4. That language – or the ultimate language in this chain of language -> representation language – is some innate mental language, for if there were an infinite regress, we could never learn a language.

This is clearly a top-down analysis of a cognitive task. The basic question is, "How do humans produce language?" The answers are increasingly detailed, specifying lower and lower levels of abstraction. They produce language by using rules. Those rules must be represented in some language. That language must be an innate mental language. This seems almost exactly like the sort of operation we would take to design a language processing system in computer science – except we would continue beyond step 4, and begin specifying that language. Most top-down cognitive scientists tend to think that a serious analysis of human language production will end up with a specification for our internal language.

Among the bottom-up cognitive scientists are researchers like Patricia Churchland and Terence Sejnowski (e.g., 1992), David LaBerge (e.g., 1995), and a host of others – most neurologists, neurobiologists, and the self-styled neuropsychologists, in fact. The goal of bottom-up cognitive science is isolating behaviors at the lowest level, presumably that of the neuron, and building a model of cognition based on knowledge of the actual processes in the brain. Hardcastle (Forthcoming) claims that one of the dominant characteristics of such approaches is the claim of *reduction*, the claim that high-level cognitive functions can be explained in terms of lower-level processes. Bottom-up cognitive science desires reduction insofar as it desires to explain how our neurobiology results in our gross organismic behaviors.

One question you might like to ask, as a cognitive scientist working from the bottom-up, is "What is the significance of this oscillation at this specific frequency in the prefrontal cortex?" The answer to this question will partially determine the next level of abstraction; at that next level whatever is going on is something that necessitates the oscillation in question. The difficulty in using an exclusively bottom-up approach is this: it isn’t clear what that next level of abstraction should look like, so it isn’t clear what role you should assign to the observed behavior. You could attribute this oscillation to a focusing of attention, or to the production of a phenomenal experience, or to the binding of a variable to some memory location – or to any one of a host of other processes. Which one of these characterizations of the behavior is accurate will not be immediately obvious from the behavior itself.

You will notice that the question just asked differs in kind from the questions proposed in the last section for a bottom-up design of a GUI. Instead of asking what processes need to go on at the lowest level – what behaviors we need to implement as routines – we are looking at behaviors already in place. In bottom-up cognitive science, we are given the building blocks, and must infer from them what the next highest level of abstraction is, to build a chain between the lowest and highest levels. In computer science, it is left up to us to determine (in a sense) what the building blocks are.[24]

3.2.1 Constraints at Two Levels

The difference between the top-down / bottom-up distinction as used in computer and cognitive science is perhaps best seen as a difference in the constraints on the problems in each domain. In both computer science and cognitive science there are constraints imposed from the “top”, and constraints imposed from the “bottom”, as it were. Constraints imposed from the top are imposed by the *goals* of the task at hand, constraints imposed as a result of our goal of designing a GUI or of modeling cognition. Constraints imposed from the bottom are imposed by the implementation architecture, constraints on the lowest level of abstraction.

[24] I think it important to recognize that the lowest level in computer science is somewhat arbitrary and task-relative. If you are programming, the ultimate building blocks are the sentences in the language being used. However, there are still a great many ways to combine those sentences into functional blocks of code for

Consider our two top-down tasks. The constraints on each are different, but of the same kind: there is some vague specification we are trying to analyze into discrete parts. In the case of cognition, we will be dealing with issues like consciousness, attention, language production, etc. In designing a GUI, we will be dealing with user requirements, multi- or single- document issues, and file-system / interface relationship issues. In each case, our constraints are open to interpretation – there are numerous ways to implement a multi-document interface, just as there are many ways that humans could handle the problem of attention. The constraints are loose: the GUI must handle multiple documents, while the model of cognition must include attentional processing. The task of a top-down approach is specifying constraints all the way down, until an implementation is determined.

Now, for comparison, consider our two tasks from the bottom-up. In designing a GUI, we will be concerned with developing subroutines that we think might be necessary. We might create subroutines for text-display and window-display, with the vague notion that our GUI will eventually need something like that. The constraints involved in bottom-up design, though, are not explicitly specified from above. Instead, they are determined by the architecture of the computer on which the GUI is intended to run, and by the language in which the GUI is written. These constraints will determine what the actual code looks like, and what the lowest-level subroutines actually *do*. However, we have freedom to specify what processes go on using that language; we are only constrained in that it must be written in that language.

In attempting to do cognitive science from the bottom-up, though, we run into a different type of constraint. The implementation is already determined – we do not need to decide what the *exact* code will be, for there is (in at least some sense) a matter of fact about what it is. The analogy would more appropriately be to a program that is already written. The constraints on our bottom-up model are imposed by the actual behaviors in human brains, and the actual cognition going on. Our task is not determining *a* relationship

---

very low-level routines. In cognitive science, I think the appropriate analogy is between those routines and between the lowest level of abstraction of cognition and the highest, but determining *the* relationship that in fact holds between them. We are constrained by what actually goes on in the brain, not merely by what sorts of processes the brain *could* implement.

This is one place cognitive science and AI diverge. In cognitive science, we are interested in discovering what the actual top-down and bottom-up constraints involved in cognition are. In AI it is perfectly reasonable to specify your constraints pragmatically – you can switch processors if your task cannot be performed on one, or change languages if the language you're working with proves unsuitable for the task at hand. On the other hand, in cognitive science the goal is to discover how *human* cognition (and hopefully, quite a lot of other cognition) works. To that end we are constrained by our actual cognitive processes. It is not enough that our models handle cognitive tasks, but that they handle them in the same way humans do. [25]

Note that this is not the claim that our top-down explanations can reduce to bottom-up explanations. As Hardcastle (1996) argues, the constraints imposed by our actual cognitive processes (at the lowest levels) only restrict the possible classes of higher-level explanations. They do not necessarily (and in certain cases do not look likely to) constrain us to one particular higher-level explanation. This is the requirement, though, that the higher-level explanations we have of cognitive behavior be such that they could be implemented as the actual neurobiological processes of humans. If they cannot, then explanations are not actually telling us anything about *human* behavior.

3.3 Dynamicism, Neurobiology, and the Top-Down / Bottom-Up Distinction

In section 2.2.1, we discussed some general features of dynamicist explanation. The ability of dynamical systems theory to deal with collective variables to reduce

---

the observed behaviors at the neural level.

[25] Gunderson (1971) argues for precisely this distinction. In his terms, it is the distinction between cognitive simulation (or the computer simulation of cognitive processes) and artificial intelligence.

complexity, and the ability of dynamical systems models to couple to form broader models were both illustrated nicely by the speech articulation model of Saltzman (1995). These features provide some of the more obvious reasons to employ dynamical systems theory, since it is in virtue of them that dynamicism can provide explanations in terms of factors at multiple levels of abstraction and in multiple physical locations. The very features that make dynamicism so successful at providing these explanations, though, are the features that will prove key in identifying dynamicism as inherently top-down.

Recall again the description of Saltzman's (1995) speech articulation model from section 2.2.1. Speech articulation is defined over a task space defined by tract variables, each of which represents with a single value the state of some set of articulators. Each point in the task space corresponds with some set of model articulator coordinates, such that at any point in time, the location of the model articulators will correspond to those coordinates determined by the state of the system in that task space. To go beyond our previous discussion of his model, it describes the movement of the model articulators over time in response to a goal state (the sounds to be produced), where that goal state is specified as a set of point attractors towards which each tract variable evolves.[26]

Imagine for a moment the process of constructing this model. We have a rough idea of the phenomena of speech; we have some basic information about how the vocal tract can move and we know that those movements affect the sounds we produce. Given this vague high-level characterization of the task to be analyzed, Saltzman then proposes that the best way to consider the task is in terms of a task space. The task space is in turn best defined in terms of variables representing certain constrictions in the vocal tract. Those constrictions are realized by the movements of various articulators, but the relationship between the constrictions and individual articulators is defined by a function mapping the values for tract variables onto coordinates for the model articulators. This is an excellent example of a (successful) top-down approach to explaining a cognitive task.[27]

[26] Saltzman (1995), p. 161.
[27] Saltzman has offered us an explanation in terms of a set of inputs (goal states characterized as point attractors) and a set of outputs (model articulator coordinates).[27] This is a functionalist explanation. Functionalist explanations are often contrasted with *structural* explanations (see, e.g., Fodor, 1975;

This is somewhat of an unfair reconstruction of Saltzman's design goals, though. He does in fact account for certain anatomical and neurobiological constraints in the construction of his model. These constraints are not imposed *by* his model, though, nor are they *part* of the model, per se. Instead, these constraints serve to explain why the model considers certain tract variables to map to certain sets of model articulator coordinates. These constraints are not *prima facie* dynamical constraints, though. Likewise they are not top-down constraints, but are instead constraints imposed by the actual instantiation of the model. Salztman is actually employing elements of top-down and bottom-up design in employing elements of dynamicist and neurobiological explanation. We will return to an analysis of combined approaches like his in section 3.4.

Consider what precisely about Saltzman's model causes us to identify it as (primarily) top-down. At each level of abstraction his explanation comes closer to the actual physical implementation. There is no reason why the model provided by Saltzman could not be instantiated by any number of physical systems; there is additional work needed to demonstrate that his model is in fact instantiated by the human articulatory system. To make clear exactly how the vocal tract instantiates his model, we must explain how the goal states are specified in the physical system, and how the individual components of the system that are implicitly coupled by the conversion of tract variables to model articulator coordinates are instantiated via physical processes. This is not Saltzman's concern, though – he is offering an explanation of the behavior of the system, not of how that behavior is instantiated.

What this means for the dynamicists, if we are correct in claiming their explanations are top-down, is that their explanations have no link to the structures that implement their models. Or, as Clark (1996) puts it, the dynamicist offering an explanation of cognition

---

Pylyshyn, 1984). P. S. Churchland (1986), pp. 153-4 notes that while most who grant this contrast as a strict dichotomy do not hold that explanations in terms of function require *no* information about structure (and hence need provide no explanation in terms of structure), there is generally a sense among those who offer functionalist explanations that no reduction is possible between functional explanations and structural explanations.

cannot tell us "how to build it".[28] This is not the demand that the dynamicist tell us precisely how to construct their model – that would be a tough requirement to fulfil – but a recognition that when we learn that, say, the behavior of the articulatory system is best defined over some highly abstract task space, we want to know how *humans* implement that. We want to know how the individual components of the articulatory system are hooked up to each other *such that they behave in the manner specified by the model*.

3.3.1 Dynamical Models and Implementation

To make this a bit clearer, consider this example from Dewan (1976), as related by Hooker (1981).

> Consider a set of electrical generators G, each of which produces alternating current electrical power at 60Hz but with fluctuations in frequency of 10% around some average value. Taken singly, the frequency variability of the generators is 10%. Taken joined together in a suitable network, their collective frequency variability is only a fraction of that figure because, statistically, generators momentarily fluctuating behind the average output in phase are compensated for by the remaining generators, and conversely, generators momentarily ahead in phase have their energy absorbed by the remainder. The entire system functions, from an input/output point of view, as a single generator with a greatly increased frequency reliability, or, as control engineers express it, with a single, more powerful, "virtual governor". The property "has a virtual governor of reliability f" is a property of the system as a whole, but of none of its components." *Hooker (1981), p. 509*[29]

As Hooker notes in the following discussion, while there are properties attributable to the system as a whole and not to the individual components of the system, those properties arise from the structure of the system and the laws governing the individual generators. More broadly, the properties of the system are the result of the interaction of its components – the structure of the system – according to laws. In terms of our discussion

[28] Clark (1996), pp. 120-2.
[29] This example, while from Hooker – who got it from Dewan, 1976 -- is taken verbatim from P. S. Churchland (1986), pp. 365-6, where it is used to make a similar point.

of Saltzman's model, while it is useful to have a mathematical model telling us that the components of the articulatory system are behaving in a certain way, we want to know what it is about the structure of the system that causes it to behave that way.

We should note here that, as in section 3.2, this is not specifically a requirement that our high-level explanations, such as Saltzman's model, should reduce explicitly to lower level explanations involving the structure of the system. Instead, there are questions that we can ask of Saltzman's model – namely, how it is implemented in the human body – that are questions answered only by a different kind of explanation. These are the types of explanations traditionally provided by neuroscience, rather than psychology. These are in fact explanations falling into Clark's second and third categories (see section 3.0 above). They involve an appeal to the neurological (or, more generally, physical) basis of the behaviors of which Saltzman is providing such a high-level characterization, and an identification of the roles the different physical components play in producing those behaviors.

Neurobiological explanations look more like bottom-up than top-down explanations. Instead of hypothesizing some abstract model based on some high-level characterization of the behavior of neurological components, models are constructed of the behavior of those primitive (ex hypothesi) components. The neurobiologist then attempts to explain what that behavior is doing, by assigning it some role in higher-level function. This chain of behaviors and explanations works upwards, with the eventual goal of explaining high-level cognitive functions in terms of lower-level neural behaviors. The interesting thing about Saltzman's model is that it is not, per se, strictly a dynamical model. In addition to providing Clark's first type of explanation – the ultimate goal of Saltzman's model – Saltzman has constructed his model based on explanations from categories two and three, both of which are produced not by dynamicism, but by traditional neurobiology (and biology in general).

3.4 The Co-Evolution of Top-Down and Bottom-Up Explanations

It is not surprising that Saltzman's model, while offering a top-down explanation of speech articulation, is constructed with input from a bottom-up explanation. We saw in section 3.2.1 that there are constraints on our explanations imposed from above and from below; from the high-level characterizations of phenomena and from the low-level behaviors. A top-down approach, by definition, will conform to those high-level design constraints. A bottom-up approach will, likewise, conform to the low-level design constraints. What Saltzman's model of articulation shows us is that a top-down approach can be brought into line with low-level constraints by input from bottom-up approaches, providing a more complete and consistent explanation.

This is a very similar point to that made by Flanagan (1992) (e.g., pp. 11-20) when he argues for his Natural Method. While he does not make the top-down / bottom-up distinction as I have it here, he acknowledges that different constraints have been placed on different approaches to cognitive science. Psychological theories are generally expected to remain true to the high-level phenomena, while neuroscientific theories are expected to remain true to the neurobiology. Essentially, his point is that while each of these is held *primarily* to a different standard, we still would like to know that our psychological theories can in fact be realized in the human physical structure, and that our neuroscientific theories don't conflict with the observed gross behaviors of the organism studied by psychology.

Arguing for the possibility of a reduction between psychological and neurobiological explanations, P. S. Churchland (1986) comes to a similar conclusion about the interaction of the disciplines. She argues (e.g., pp. 362-76) that by developing our psychological theories with input from neurobiology, and our neurobiological theories with input from psychology, we will eventually achieve a reduction between psychology and neurobiology. Simply put, she argues that by developing increasingly more detailed pictures of the constraints imposed from each level, we will develop a clearer picture of what is actually going on.

Hardcastle (1996) also recognizes the utility of combining top-down and bottom-up explanations, arguing that while we may not necessarily achieve reduction between functionalist explanations and neurobiological explanations, we can achieve explanatory extension.[30] As she notes, it is not necessarily the case that our neurobiological explanations will constrain us to a single psychological explanation of a behavior. Her point is well taken. Merely because we develop a clearer picture of the constraints at each level, we are not guaranteed to completely determine the relationship between the levels of psychological and neurobiological explanations. She argues, and I think rightly, that we should instead expect that there are questions we can ask at one level of abstraction that can only be answered by reference to theories at the other level. To put it more simply, explanations at one level of abstraction can be *extended* by reference to explanations at another level.

Salztman's model of articulation provides us with a perfect example of neurobiology extending the explanations of dynamicism. When we ask why the lips compensate for the obstruction of jaw movement during the production of [b ae b], Saltzman can offer an explanation in dynamicist terms adverting to the attractor dynamics of the system. If we want to know how the actual system responds to this perturbation, though, we have to appeal to the neurobiological basis of his model. Key assumptions relating tract variable values to model articulator coordinates are based on information about the actual structure and low-level function of the articulatory system. So in order to explain the behavior of the system in terms of the model, Saltzman must refer to the biological and anatomical information used to construct those assumptions (and hence the model). So in this case, neuroscientific explanations extend the explanations that the dynamical model can provide.

It is perhaps easiest to see the utility of employing a combined approach, involving top-down and bottom-up approaches, as a search problem. There is in fact some relationship between the gross behaviors of humans and their neurobiological behaviors, and we are, through the process of providing explanations of these behaviors, attempting to determine what that relationship is. By discerning the constraints arising from our neurobiology, we will be eliminating all psychological theories that conflict with those constraints. Likewise, by discerning the constraints arising from our psychological theories, we will be eliminating all bridging theories that conflict with those constraints.[31] In this way we can narrow down the search space of possible bridging theories between our gross behaviors and our neurobiology.

3.4 Conclusions

What we have seen in this chapter is that dynamical explanations of behavior are top-down explanations, and neurobiological explanations of behavior are bottom-up explanations. Elliot Saltzman's (1995) dynamical explanation of speech articulation is an excellent example of the utility of combining a top-down dynamical explanation with a bottom-up neurobiological approach. We saw that Saltzman's model, to be considered to actually model human speech articulation, must be in some sense constrained by our neurobiology. Likewise, our theories of how the neurobiology relates to the task of speech articulation should take into account the actual process of articulation that Saltzman's model so accurately describes.

In combining top-down and bottom-up explanations, we provide all three types of explanation Clark (1996) argues that we need in cognitive science. In the case of Saltzman's model, by informing it with neurobiological explanations of articulation, we can answer questions not only about how the vocal tract produces certain sounds (and how it compensates for perturbations), but how the structure of the vocal tract contributes

---

[30] Hardcastle (1996), pp. 102-3 finds the characterization of neurobiological explanations as distinct from functionalist explanations somewhat problematic.

[31] Following Hardcastle, a bridging theory is a theory explaining the relationship between two other theories – in this case, between our "complete" psychological and neurobiological theories.

to the overall behavior. As Hardcastle argued, we achieve explanatory extension by the combination of approaches.

So, in light of what we learned in chapters 1 and 2, a clearer picture emerges of the approach we should take in cognitive science. Since there is a need for each of the types of explanation on Clark's list, we need an approach that can provide each of those types. Likewise, since dynamicism is the best available approach for providing the first type of explanation, our "complete" approach should incorporate dynamicism. And, since dynamicism provides top-down explanations of behaviors, we should – for instrumental reasons – employ traditional neurobiological methods as well as dynamical methods for explaining behaviors. With a combined approach, using dynamical (top-down) and neurobiological (bottom-up) explanations, we are better able to fulfil the explanatory demands of cognitive science.

Bibliography

Ashby, W. R. (1965). *Design for a brain: The origin of adaptive behavior*. London: Chapman & Hall.

Block, N. (1995). The mind as the software of the brain. In Smith and Osherson, ed., *An invitation to cognitive science*, III, 377-425. Cambridge, MA: MIT Press.

Beer, R. (1995). Computational and dynamical languages for autonomous agents. In Port and Van Gelder, eds., *Mind as Motion*, 121-148. Cambridge, MA: MIT Press.

Browman, C. and Goldstein, L. (1995). Dynamics and articulatory phonology. In Port and Van Gelder, eds., *Mind as Motion*, 175-194. Cambridge, MA: MIT Press.

Churchland, P.S. and Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Churchland, P.S. (1986). *Neurophilosophy: Toward a unified science of the mind / brain*. Cambridge, MA: MIT Press.

Chomsky, N. (1975). *Reflections on language*. New York: Pantheon.

Clark, A. (1996). *Being there*. Cambridge, MA: MIT Press.

Dewan, M. (1976). Consciousness as an emergent causal agent in the context of control system theory. In *Consciousness and the brain*, ed. G. Globus, G. Maxwell, and I. Savodnik, 181-198. New York: Plenum.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

Elman, J. (1995). Language as a dynamical system. In Port and Van Gelder, eds., *Mind as Motion*, 196-225. Cambridge, MA: MIT Press.

Ericsson, K. and Simon, H. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Flanagan, O. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Fodor, J. (1975). *The language of thought*. New York: Crowell.

Garson, J. (1996). Cognition poised at the edge of chaos. *Philosophical Psychology* 9, 3, 301-322.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Grossberg, S. (1995). Motion perception, recognition learning, and spatial attention. In Port and Van Gelder, ed., *Mind as Motion*, 449-489. Cambridge, MA: MIT Press.

Grossberg, S. (1969). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics*, 1, 319-350.

Hardcastle, V. (forthcoming). *A pain by any other name*.

Hardcastle, V. (1996). *How to build a theory in cognitive science*. Albany, NY: SUNY Press.

Hooker, C. A. (1981). Towards a general theory of reduction. Part I: Historical and scientific setting. Part II: Identity and reduction. Part III: Cross-categorial reduction. *Dialogues* 20:38-59, 201-236, 496-529.

Horgan, T. and Tienson, J. (1995). *Connectionism and the philosophy of psychology*. Cambridge, MA: Bradford Books / MIT Press.

Katok, H. and Hasselblatt, B. (1995). *Introduction to the modern theory of dynamical systems*. New York, NY: Cambridge University Press.

Kelso, J. A. S. (1995). *Dynamic patterns*. Cambridge, MA: MIT Press.

Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (1986). The dynamical perspective on speech production : Data and theory. *Journal of Phonetics*, 14, 29-59.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (1984). Functionally specific articulatory cooperations following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 812-832.

Kelso, J. A .S., Tuller, B., and Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. *Journal of the Acoustical Society of America*, 72, S103.

Kelso, J. A. S. (1981). On the oscillatory basis of movement. *Bulletin of Psychonomic Society*, 18, 63.

LaBerge, D. (1995). *Attentional processing: The brain's art of mindfulness*. Cambridge, MA: Harvard University Press.

Litch, M. (1997). Computation, Connectionism, and Modeling the Mind. *Philosophical Psychology*, 10, Number 3.

Newell, A. and Simon, H. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the Association for Computing Machinery,* 19, 113-126.

Norton, A. (1995). Dynamics: An introduction. In Port and Van Gelder, ed., *Mind as Motion,* 45-68. Cambridge, MA: MIT Press.

Port, R. and Van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Pylyshyn, Z. (1984). *Computation and cognition: Issues in the foundations of cognitive science*. Cambridge, MA: MIT Press.

Ryle, G. (1959). *The Concept of Mind*. Chicago: University of Chicago Press.

Saltzman, E. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In Port and Van Gelder, eds., *Mind as Motion*, 149-174. Cambridge, MA: MIT Press.

Simon, H. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Thelen, E. (1995). Time-scale dynamics and the development of an embodied cognition. In Port and Van Gelder, ed., *Mind as Motion*, 69-100. Cambridge, MA: MIT Press.

Thelen, E. and Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: Bradford Books / MIT Press.

Thelen, E. (1994). Three-month old infants can learn task-specific patterns of interlimb coordination. *Psychological Science,* 5, 280-285.

Thelen, E., Corbetta, D., Kamm, K., et. al. (1993). The transition to reaching: matching intention and intrinsic dynamics. *Child Development*, 64, 1058-1098.

Townsend, J. and Busemeyer, J. (1995). Dynamic representations of decision-making. In Port and Van Gelder, ed., *Mind as Motion,* 101-120. Cambridge, MA: MIT Press.

Van Gelder, T. (forthcoming, 1997). The dynamical hypothesis in cognitive science. *Brain and Behavioral Sciences.*

Van Gelder, T. and Port, R. (1995). It's about time: An overview of the dynamical approach to cognition. In Port and Van Gelder, eds., *Mind as Motion,* 1-43. Cambridge, MA: MIT Press.