To appear in In R. Stainton (Ed.), *Contemporary debates in cognitive science*. Oxford: Basil Blackwell.

Phenomena and Mechanisms: Putting the Symbolic, Connectionist, and Dynamical Systems Debate in Broader Perspective

William Bechtel and Adele Abrahamsen
University of California, San Diego

Cognitive science is, more than anything else, a pursuit of cognitive mechanisms. To make headway towards a mechanistic account of any particular cognitive phenomenon, a researcher must choose among the many architectures available to guide and constrain the account. It is thus fitting that this volume on contemporary debates in cognitive science includes two issues of architecture, each articulated in the 1980s but still unresolved:

- Just how modular is the mind? (section 1) – a debate initially pitting encapsulated mechanisms (Fodorian modules that feed their ultimate outputs to a nonmodular central cognition) against highly interactive ones (e.g., connectionist networks that continuously feed streams of output to one another).
- Does the mind process language-like representations according to formal rules? (this section) – a debate initially pitting symbolic architectures (such as Chomsky's generative grammar or Fodor's language of thought) against less language-like architectures (such as connectionist or dynamical ones).

Our project here is to consider the second issue within the broader context of where cognitive science has been and where it is headed. The notion that cognition in general—not just language processing—involves rules operating on language-like representations actually predates cognitive science. In traditional philosophy of mind, mental life is construed as involving propositional attitudes—that is, such attitudes towards propositions as believing, fearing, and desiring that they be true—and logical inferences from them. On this view, if a person desires that a proposition be true and believes that if she performs a certain action it will become true, she will make the inference and (absent any overriding consideration) perform the action.

This is a prime exemplar of a symbolic architecture, and it has been claimed that all such architectures exhibit *systematicity* and other crucial properties. What gets debated is whether architectures with certain other design features (e.g., weighted connections between units) can, in their own ways, exhibit these properties. Or more fundamentally: What counts as systematicity, or as a rule, or as a representation? Are any of these essential? Horgan and Tienson offer their own definition of systematicity and also of syntax, arguing that syntax in their sense is required for cognition, but not necessarily part-whole constituent structures or exceptionless rules. They leave to others the task of discovering what architecture might meet their criteria at the scale needed to seriously model human capabilities. McLaughlin argues for a more classical position in which systematicity, more tightly defined, is a capacity that points to the rules and representations of a traditional symbolic architecture.

Our own goal is to open up the debate about rules and representations by situating it within a framework taken from contemporary philosophy of science rather than philosophy of mind. First,

we emphasize the benefits of clearly distinguishing phenomena from the mechanisms proposed to account for them. One might, for example, take a symbolic approach to describing certain linguistic and cognitive phenomena but a connectionist approach to specifying the mechanism that explains them. Thus, the mechanisms may perform symbolic activities but by means of parts that are not symbolic and operations that are not rules. Second, we point out that the mechanisms proposed to account for phenomena in cognitive science often do not fit the pure types debated by philosophers, but rather synthesize them in ways that give the field much of its energy and creativity. Third, we bring to the table a different range of phenomena highly relevant to psychologists and many other cognitive scientists that have received little attention from philosophers of mind or even of cognitive science—those that are best characterized in equations that relate variables. Fourth, we offer an inclusive discussion of the impact of dynamical systems theory on cognitive science. It offers ways to characterize phenomena (in terms of one or a few equations), explain them mechanistically (using certain kinds of lower-level models involving lattices or interactive connectionist networks), and obtain new understandings of development.

Phenomena and two ways of explaining them

Characterizing phenomena and explaining them are key tasks of any science. The nature of the characterization depends on the domain but, at least initially, tends to stay close to what can be observed or directly inferred. In the domain of language, many phenomena can be efficiently characterized in terms of rules and representations. For example, the phenomenon of past-tense formation can be expressed roughly as follows: for regular verbs, *V* → *V* + *ed* (the verb stem is represented categorically as *V* and linked to its past tense form by a general rule); for irregular verbs, *eat* → *ate*, *fly* → *flew*, *give* → *gave,* and so forth (specific pairs of representations are linked individually). In physics, many classic phenomena are characterized in *empirical laws* that express idealized regularities in the relations between variables over a set of observations. According to the Boyle-Charles' law, for example, the pressure, volume, and temperature of a gas are in the relation $pV = kT$. In psychology, there is a similar focus on relations between variables, but these relations are less likely to be quantified and designated as *laws*. Instead, psychologists probe for evidence that one variable causes an effect on another variable. Cummins (2000) noted that psychologists tend to call such relations *effects*, offering as an example the Garcia effect: animals tend to avoid distinctive foods which they ate prior to experiencing nausea, even if actual cause of the nausea was something other than the food (Garcia, McGowan, Ervin, & Koelling, 1968).

In the traditional deductive-nomological (D-N) model (Hempel, 1965), characterizations of phenomena are regarded as explanatory. For example, determining that an individual case of nausea is "due to" the Garcia effect explains that case. On a more contemporary view, such as that of Cummins, identifying the relevant phenomenon is just a preliminary step towards explanation. The actual explanation typically involves one of two approaches:

- recharacterizing the phenomenon in terms of such abstractions as **theoretical laws** or underlying principles. This approach to explanation originated in logical positivism and its hypothetico-deductive method (Hempel, 1965; Suppe, 1974) and produced an influential position holding that less basic theories can be reduced to more basic, fundamental ones (Nagel, 1961). On this view, a given science aims towards a parsimonious, interconnected system of laws (axioms) from which other laws or

> predictions can be deduced. Nagel's prime example is the derivation of the phenomena of thermodynamics (as expressed, for example, in the Boyle-Charles' law) from the more fundamental theory of statistical mechanics (in which theoretical laws incorporate such abstract constructs as *mean kinetic energy*). Though the full apparatus lost its elegance and general acceptance due to a succession of criticisms and modifications, the impulse to find fundamental explanatory laws or principles survives. Scientists still propose laws, and philosophers of science still ask how they cohere across different sciences.(For a contemporary approach emphasizing unification, see Kitcher, 1999.)

- uncovering and describing the ***mechanism*** responsible for the phenomenon, as emphasized in the mechanistic approach to explanation advocated since the 1980s by an emerging school of philosophers of science focusing on biology rather than physics. Examples include biologists' detailed accounts of such diverse phenomena as metabolism, blood circulation, and protein synthesis

The phenomena of cognitive science are so varied that every kind of characterization is encountered: symbolic rules and representations, descriptive equations comparable to the empirical laws of classical physics, statements that certain variables are in a cause-effect relation, and perhaps more. Individual cognitive scientists tend to have a preferred mode of characterization and also a preferred mode of explaining the phenomena they have characterized. Those who propose theoretical laws or principles typically work in different circles than the somewhat larger number who propose mechanisms. We will discuss both, but begin by introducing mechanistic explanation. Historically, the conception of mechanism was drawn from inanimate machines, but it was extended by Descartes to apply to all physical phenomena, including those of organic life. It was further developed by biologists in the 19th and 20th centuries who, in opposition to vitalists, viewed the pursuit of mechanism as the route to developing biology as a science. Explanatory accounts in modern biology predominantly involve mechanisms (rather than theoretical laws), and many parts of cognitive science have the same character. Here is a brief definition of mechanism that was inspired by its use in biology but is equally relevant to proposals about mental architecture:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena (Bechtel & Abrahamsen, in press; for related accounts, see Bechtel & Richardson, 1993; Glennan, 2002, 1996; Machamer, Darden, & Craver, 2000).

In an illustration from biology, the overall phenomenon of carbohydrate metabolism can be characterized as the harvesting of energy in the process of breaking down carbohydrates to carbon dioxide and water. This is explained by decomposing the responsible mechanism into various enzymes (parts) that catalyze intracellular biochemical reactions (operations) in molecular substrates (another kind of parts). For example, the enzyme succinate dehydrogenase oxidizes succinate to fumarate. But it is not sufficient to identify each reaction and the molecules involved; organization is equally important. For example, succinate → fumarate is followed by other reactions that include (omitting some intermediates, side reactions, and the ongoing provision of acetyl CoA at the step producing citrate): fumarate → malate → oxaloacetate → citrate → isocitrate → α-ketoglutarate → succinate. The complete set of reactions is known as

the Krebs cycle (or citric acid cycle)—a submechanism of metabolism that is a well-known exemplar of cyclic organization. The account can be completed by describing the spaciotemporal orchestration of the organized components in real time, that is, their dynamics. (Note to low-carb diet fans: fats and proteins have their own routes into the Krebs cycle. Any diet relies on the metabolic system's ability to dynamically adjust to the ever-changing mix of incoming food molecules.)

For philosophers of science, it was increased attention to biology that provided the impetus towards an emphasis on explanation in terms of mechanisms rather than laws. This turn towards mechanism has brought new perspectives on fundamental issues in philosophy of science (Bechtel & Abrahamsen, in press), but no new solutions to a problem known as *underdetermination*. Especially at the leading edge of any science, explanations tend to be underdetermined by available data. In the biochemistry of the 1930s, for example, there were a variety of partially-correct proposals before Hans Krebs nailed the essentials of the mechanism named for him. Underdetermination is particularly pervasive in our era for cognitive science. Dramatically different mechanistic explanations are routinely championed for a given cognitive phenomenon, and consensus is elusive. In the next section we discuss past-tense formation to illustrate the conflict between the symbolic and connectionist architectural frameworks for mechanistic explanation in cognitive science. More importantly, we then make the case that these architectures are not polar opposites for which one winner must emerge. After conceptually reframing the symbolic-connectionist debate in this way, we finish by pointing to two research programs that achieve a more integrative approach: optimality theory in linguistics and statistical learning of rules in psycholinguistics.

Getting beyond the symbolic-connectionist debate

The symbolic-connectionist debate over mechanisms of past-tense formation

If you are fluent in English, on a typical day you produce the correct past tense form for hundreds of regular and irregular verbs without giving it a moment's thought. How? This capability may seem trivial, but proponents of two different mechanistic approaches have been using it as a battleground for more than 20 years. One approach—symbol processing—keeps the mechanistic explanation very close to the characterization of the phenomenon by positing two different mechanisms. Basically, the language production system performs one of two different operations, depending on the type of verb involved:

- apply the rule *V* → *V* + *ed* if the verb is regular (e.g., *need* → *needed*), or
- get the appropriate past-tense form from the mental lexicon if the verb is irregular (e.g., the lexical entry for the stem *give* specifies its past-tense form as *gave*).

There are further details, such as distinguishing among three allomorphs of the past-tense affix *ed*, but the key point is that the mechanisms are at the same scale as the phenomonon. Operations like rule application and lexical look-up are assumed to directly modify symbolic representations.

The other approach is to explain past-tense formation by means of a single mechanism situated at a finer-grained level that is sometimes called *subsymbolic*. The best-known subsymbolic models of cognition and language are feedforward connectionist networks. Architecturally, these

originated in networks of *formal neurons* that were proposed in the 1940s and, in the guise of Frank Rosenblatt's (1961) *perceptrons*, shown to be capable of learning. Overall phenomena of pattern recognition were seen to emerge from the statistics of activity across numerous identical fine-grained units that influenced each other across weighted connections. Today these are sometimes called *artificial neural networks (ANNs)*. The standard story is that network and symbolic architectures competed during the 1960s, ignored each other during the 1970s (the symbolic having won dominance), and began a new round of competition in the 1980s. We suggest, though, that the symbolic and network approaches both were at their best when contributing to new blended accounts. Notably, in the 1980s they came together in connectionism when a few cognitive scientists pursued the idea that a simple ANN architecture could (a) provide an alternative explanation for well-known human symbolic capabilities such as past-tense formation while also (b) explaining additional phenomena of graceful degradation, constraint satisfaction, and learning that had been neglected (Rumelhart & McClelland, 1986b) Connectionist networks generally are construed as having representations across units, but no rules. The very idea that these representations are *sub*symbolic signals the ongoing relevance of the symbolic approach. Connectionists, unlike many other ANN designers, are grappling with the problem of humans' internal networks function in a sea of external symbols—words, numbers, emoticons, and so forth. In fact, at least one of the pioneers of connectionism had pursued a different blend in the 1970s that leaned more towards the symbolic side—semantic networks—but became critical of their brittleness (see Norman and Rumelhart, 1975.)

The first connectionist model of past-tense formation (Rumelhart & McClelland, 1986) performed impressively, though not perfectly, and received such intense scrutiny that its limitations have long been known. It explored some intriguing ideas about representation (e.g., coarse-coding on context-dependent units), but for some years has been superseded by a sleeker model using a familiar network design. As illustrated in Figure 1, Plunkett & Marchman's (1991; 1993) feedforward network represents verb stems subsymbolically as activation patterns across the binary units of its *input layer*. It propagates activation across weighted connections first from the input to *hidden layer* and then from the hidden to *output layer*. Each unit in one layer is connected to each unit in the next layer, as illustrated for two of the hidden units, and every such connection has its own weight as a result of repeated adaptive adjustments during learning (via back-propagation). An *activation function*, which typically is nonlinear, determines how the various weighted activations coming into a unit will be combined to determine its own activation. In this way, the network transforms the input representation twice—once for each pair of layers—to arrive at a subsymbolic representation of the past-tense form on the output layer. Although all three layers offer subsymbolic representations, it is the encoding scheme on the input layer that most readily illustrates this concept. The verb stems, which would be treated as morphemes by a rule appending *ed* in a symbolic account, are replaced here with a lower-level encoding in terms of the distinctive features of each constituent phoneme in three-phoneme stems. For example, the representation of "dez" (for convenience, they used artificial stems) would begin with an encoding of "d" as 011100 (0=consonant, 1=voiced, 11=manner: stop, 10=place: alveolar). With "e" encoded on the next six units and "z" on the last six units, "pez" is represented as a binary pattern across 18 subsymbols rather than symbolically as a single morpheme. Moreover, as the pattern gets transformed on the hidden and output layers, it is no longer binary but rather a vector of 20 real numbers, making the mapping of stem to past tense a statistical tendency. Connectionist networks are mechanisms—they have organized parts and

operations—but the homogeneity, fine grain, and statistical functioning of their components make them quite distinct from traditional symbolic mechanisms. (See Bechtel & Abrahamsen, 2002, for an introduction to connectionist networks in chapters 2 and 3 and discussion of past-tense networks and their challenge to rules in chapter 5.)

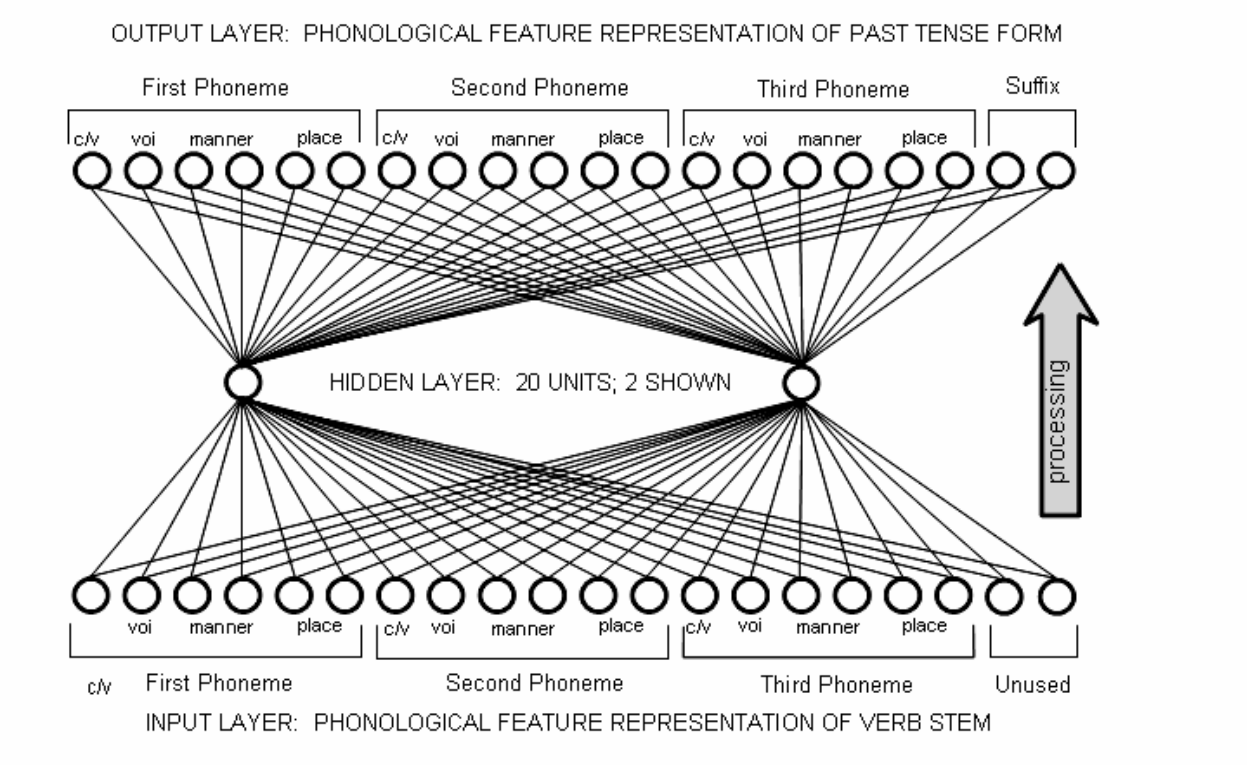

Figure 1. Plunkett and Marchman's (1991) feedforward network for past-tense formation. Each artificial verb stem gets a subsymbolic encoding on the input units, based on the phonological features of each of its phonemes Propagation of activation across weighted connections (shown for two of the 20 hidden layer units) transforms the input pattern into a past-tense form on the output units.

The trained network in Figure 1 can stand on its own as a mechanistic model accounting for past-tense formation by adult speakers. However, it is the network's behavior during training that has captured the greatest attention, due to claims that it provides a mechanistic explanation of an intriguing developmental phenomenon, U-shaped acquisition. It has long been known that in acquiring the regular past-tense, children overgeneralize it to some of their irregular verbs—even some that had previously been correct (Ervin, 1964; Brown, 1973). For example a child might correctly produce *went* and *sat* at age 2, switch to *goed* and *sitted* at age 3, and gradually return to *went* and *sat* at ages 4-5. Graphing the percentage of opportunities on which the correct form was used against age, a U-shaped acquisition function is obtained. This contrasts with typical learning curves, which tend to be sigmoidal or exponential. The phenomenon of interest is that acquisition of irregulars cannot be described by any of the usual functions but rather is U-shaped, as illustrated in Figure 2. Advocates of the symbolic approach have interpreted this as favoring the two-mechanism account (applying a rule for regulars and looking up the verb in a mental lexicon for irregulars); specifically, they attribute the decline in performance on irregulars to the replacement of lexical look-up (which gives the correct form) with overgeneralization of the rule

(yielding the regular past-tense that is inappropriate for these verbs). The alternative proposal, initially advanced by Rumelhart and McClelland (1986a), acknowledged competition but relocated it within a single mechanism—their connectionist network in which the same units and connection weights were responsible for representing and forming the past tense of all verbs. Especially when an irregular verb was presented to the network, activation patterns appropriate to different past-tense forms would compete for dominance. Like children, their network showed a U-shaped acquisition curve for irregular verbs across its training epochs. Pinker and Prince (1988) and others objected to discontinuities in the input and to shortcomings in the performance of Rumelhart and McClelland's network relative to that of human children (based not only on linguistic analyses but also on detailed data gathered by Kuczaj, 1977, and later by Pinker, Marcus and others). Subsequent modeling efforts (Plunkett & Marchman, 1991, 1993) addressed some of the criticisms, but the critics responded with their own fairly specific, though unimplemented, model. A readable, fairly current exchange of views is available in a series of papers in *Trends in Cognitive Sciences* (McClelland & Patterson, 2002a, 2002b; Pinker & Ullman, 2002a, 2002b).

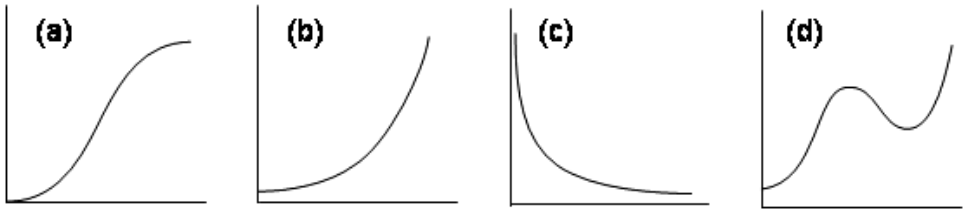

Figure 2. Some nonlinear curves: (a) sigmoidal (e.g., skill as a function of practice); (b) positively accelerated exponential (e.g., early vocabulary size as a function of time); (c) negatively accelerated exponential (e.g., number of items learned on a list as a function of list repetitions); (d) U-shaped acquisition (e.g., irregular past tense as a function of age).

When competing groups of researchers are working within architectural frameworks as distinct as the symbolic and connectionist alternatives, data alone rarely generate consensus. Advocates tend to adjust specific aspects of their account to accommodate new findings rather than abandon their architecture. In the longterm, an architectural framework may fade away because necessary accommodations make it increasingly inelegant, or it may be absorbed into a more powerful framework when new phenomena are identified that it cannot handle, or some other fate may await it. In the shorter term, though, a notable byproduct of the competition is that the phenomenon of interest becomes much more precisely characterized as each group makes increasingly detailed predictions and obtains data to test them. This raises the bar not only for the competing explanations but also for any future ones. On occasion, either the additional data or a consequent revised understanding of the mechanism lead to a substantial reconstrual of the original phenomenon.(Bechtel & Richardson, 1993, refer to this as "reconstituting the phenomenon"). More relevant to the case of past tense acquisition is that the explanations themselves or the relation between them may get reconstrued in a way that reframes the debate. The next section pursues this possibility.

Reframing the debate

Symbolic and connectionist approaches are treated in exchanges like this as competitors, but there are at least two ways of reframing the discussion that make it less contentious and perhaps more satisfactory. One way, as we have noted previously, is to consider the implications of the fact that connectionist networks

> …repeatedly find themselves in the same grooves. That is, they behave in ways that can be closely approximated by symbolic models, and for many purposes it is the symbolic models that are most convenient to use. . . . The real challenge for connectionists will not be to defeat symbolic theorists, but rather to come to terms with the ongoing relevance of the symbolic level of analysis. That is, the ultimate new alliance may be as simple, and as difficult, as forming a new relationship with the long-time opponent. (Bechtel and Abrahamsen, 2002, p. 16).

That is, the two competing approaches to past-tense formation might be given complementary roles. One way of construing this is to appreciate linguistic rules as well-suited to characterizing the **phenomenon** of past-tense formation but to prefer feedforward networks as a plausible **mechanism** for producing the phenomenon. Alternatively, both architectures might be viewed as suitable for mechanistic accounts, but at different levels—one course-grained and symbolic, the other fine-grained and statistical. Whatever the exact roles, providing a place for more than one approach can move inquiry towards how they complement each other rather than seeking a winner. In particular, both approaches need not directly satisfy every evaluative criterion. For example, considerations of systematicity proceed most simply (though not necessarily exclusively) with respect to symbolic accounts, and graceful degradation is one of the advantages offered by a fine-grained statistical account.

Looking beyond the Chomskian-connectionist axis of the past-tense debate, an alternative linguistic theory exists that has been very amenable to—even inspired by—the idea that symbolic and subsymbolic approaches each have a role to play in an integrated account. *Optimality theory* emerged from the collaboration between two cognitive scientists who were opponents in the 1980s: connectionist Paul Smolensky and linguist Alan Prince. They showed that the constraint-satisfaction capabilities of networks could be realized (somewhat differently) at the linguistic level as well. Specifically, they succeeded in describing various phonological phenomena using a single universal set of soft rule-like constraints to select the optimal output among a large number of candidate outputs (Prince & Smolensky, 1993, 2004). A given language has its own rigid rank ordering of these constraints, which is used to settle conflicts between them. For example (see Tesar, Grimshaw, & Prince, 1999 for the full five-constraint version): the constraint NoCoda is violated by any syllable ending in a consonant (the coda), and the constraint NoInsV is violated if a vowel is inserted in the process of forming syllables (the output) from a phoneme string (the input). The input string /apot/ would be syllabified as *.a.pot.* in a language that ranks NoInsV higher (e.g., English), but in a vowel-final form like *.a.po.to.* in a language that ranks NoCoda higher (e.g., Japanese).

Optimality theory (OT) offers such an elegant explanation of diverse phenomena that a substantial number of phonologists have adopted it over classic rule-based theories. (Uptake in

syntax has been slower.) For reasons difficult to explain without presenting OT in more detail, it is implausible as an explicit mechanistic account. Those with the ambition of integrating OT with an underlying mechanistic account have tended to assume a connectionist architecture. Prince and Smolensky (1997) posed, but did not solve, the most tantalizing question invited by such a marriage: Why would the computational power and flexibility offered by a statistical mechanism like a network be funneled into solutions (languages) that all exhibit the rigid rank-ordering phenomenon that makes OT a compelling theory?

A similar dilemma has been raised by a team of cognitive scientists whose integrative inclinations have operated on different commitments (their linguistic roots are Chomskian, and they regard learning as induction rather than adjustments to weights in a network). They have offered provocative evidence that the language acquisition mechanism is highly sensitive to distributional statistics in the available language input (Saffran, Aslin, & Newport, 1996), but seek to reconcile this with their view that the product of learning is akin to the rules and representations of linguistic theory. That is, a statistical learning mechanism is credited with somehow producing a nonstatistical mental grammar. This brings them into disagreement with symbolic theorists on one side, who deny that the learning mechanism operates statistically (see Marcus, 2001) and with connectionists on the other side, who deny that the product of learning is nonstatistical. In support of their nuanced position, Newport and Aslin (2000) cited the finding that children acquiring a signed language like ASL from non-native signers get past the inconsistent input to achieve a more native-like morphological system. On their interpretation "the strongest consistencies are sharpened and systematized: statistics are turned into 'rules'" (p. 13). They and their collaborators have also contributed a growing body of ingenious studies of artificial language learning by infants, adults, and primates, from which they argue that the statistical learning mechanism has selectivities in its computations that bias it towards the phenomena of natural language (see Newport & Aslin, 2000, 2004).

Attempts like these to integrate connectionist or other statistical approaches with symbolic ones offer promising alternatives to polarization. We mentioned, though, that there is a second way of reframing the discussion. Looking at the rise of connectionism in the early 1980s, it is seen to involve the confluence of a number of research streams. Among these are mathematical models, information processing models, artificial neural networks, and symbolic approaches to the representation of knowledge—especially semantic networks but extending even to the presumed foe, generative grammar. Some of these themselves originated in interactions between previously distinct approaches; for example, ANNs were a joint product of neural and computational perspectives in the 1940s, and semantic network models arose when an early artificial intelligence researcher (Quillian, 1968) put symbols at the nodes of networks rather than in rules. Some of the research streams leading to connectionism also had pairwise interactions, as when mathematical modeling was used to describe operations within components of information processing models. Finally, some these research streams contributed not only to connectionism but also, when combined with other influences, to quite different alternatives. Most important here is that dynamical systems theory (DST) took shape in a quirky corner of mathematical modeling focused on nonlinear physical state changes. It found a place in cognitive science when combined with other influences, such as an emphasis on embodiment, and some of the bends in DST's path even intersected with connectionism when it was realized that such concepts as attractor states shed light on interactive networks. Another example (not discussed in this

chapter) is that information processing models, neuroimaging, and other research streams in the cognitive and neural sciences came together in the 1990s, making cognitive neuroscience a fast-moving field both on its own and within cognitive science. As well, the idea that cognition is distributed not only within a single mind, but also on a social scale, gave rise to socially distributed cognition as a distinct approach in cognitive science (see Bechtel, Abrahamsen, & Graham, 1998, Part 3). Thus, an exclusive focus on polar points of contention would give a very distorted picture of cognitive science. This interdisciplinary research cluster in fact is remarkable for its protean nature across both short and long timeframes.

If one is seeking maximal contrast to symbolic rules and representations, it is to be found not in the pastiche of connectionism but rather within the tighter confines of one of its tributaries, mathematical modeling. Yet, except for DST, this approach has been mostly neglected in philosophical treatments of psychology and of the cognitive sciences more generally. In the next section we consider how mainstream mathematical psychology exemplifies the quantitative approach to characterizing and explaining phenomena in cognitive science. This provides conceptual context for then discussing DST and its merger with other commitments in the dynamical approach to perception, action, and cognition.

Mathematical psychology and its contributions to cognitive science

Mathematical psychology

In its simplest form, mathematical psychology offers the means of characterizing phenomena involving quantitative relations between variables. This is a very different class of phenomena than those characterized in terms of symbolic relations between representations, and accordingly, mathematical psychology has a distinctive look and feel. Although the term *mathematical psychology* only gained currency about the same time as the cognitive revolution,[1] its antecedents extend as far back as the 19th century.

The first area to be pioneered in what is now called mathematical psychology was psychophysics. A well-known example is Ernst Weber's (1834) investigation of the relation between physical dimensions such as an object's weight and psychological ones such as its perceived heaviness. He found that he could make the same generalization across a variety of dimensions: "we perceive not the difference between the things, but the ratio of this difference to the magnitude of the thing compared" (p. 172). Later this was expressed as Weber's law ($\Delta I / I = k$), where I is the intensity of a stimulus, ΔI is the *just noticeable difference* (the minimum increment over I that is detectable), and the value of k was constant except at extreme values for a given domain (e.g., approximately 0.15 for loudness). Gustav Fechner (1860) added an assumption of cumulativity to Weber's law to obtain a logarithmic function: $\Psi = c \log (I / I_0)$. That is, the intensity of a sensation is proportional to the logarithm of the intensity of the stimulus (relative to threshold intensity). The constant c depends on k and the logarithmic base.

[1] According to Estes (2002), the publication of the three-volume *Handbook of Mathematical Psychology* (Luce, Bush, & Galanter, 1963-1965) galvanized development of professional organizations for mathematical psychologists. The *Journal of Mathematical Psychology* began publishing in 1964. The Society for Mathematical Psychology began holding meetings in 1968, although official establishment of the society and legal incorporation only occurred in 1977.

This would seem definitive, but even in this most physically grounded area of psychology conflicts sometimes emerge. Notably, Stevens (1957) proposed a power law that made stronger, and in some cases more accurate, predictions than Fechner's law. Although often regarded as theoretical due to its elegance and breadth of application, Stevens' law (like its predecessors and like the Boyle-Charles' law in physics) is essentially an empirical law. Particular percepts can be explained by appeal to the law, but the law itself has not been explained—there has been no appeal to more fundamental laws and no plausible proposals regarding a mechanism. (Weber, p. 175, in applying his finding to perception of line length, noted that it ruled out a mechanism by which "the mind . . . counts the nerve endings touched in the retina." However, he had no suggestions as to what kind of mechanism would make ΔI relative to I rather than constant.)

The next arena in which psychologists expressed empirical laws in equations was learning theory. Notably, a logarithmic retention curve was found to accurately relate the number of items retained from a list to the time elapsed since the list was studied (Ebbinghaus, 1885). As the field developed, a number of other empirical phenomena were identified, and ambitions to account for them culminated in the mathematico-deductive theory of Clark Hull (1943). Explicitly inspired by logical positivism, Hull crafted a formal theory in which empirical relationships between such observables as number of reinforcements and response latency were taken to be derived from theoretical laws (axioms) that included operationally-defined *intervening variables*. For example, reinforcements affected habit strength, which was multiplied by drive to get excitatory potential, which affected response latency. Eventually Hull's theory came to be regarded as bloated and insufficiently explanatory. It was replaced by a *mathematical psychology* in which equations were expected to achieve explanatory power through parsimony:

> In principle, such a theory entails an economical representation of a particular set of data in mathematical terms, where 'economical' means that the number of free parameters of the theory is substantially smaller that the number of degrees of freedom (e.g., independent variables) in the data (Falmagne, 2002, p. 9405).

The Markov models of William Estes (1950) and R.R. Bush and F. Mosteller (1951) satisfied this criterion and energized the field by elegantly accounting for a variety of empirical data and relationships. Their new mathematical psychology surpassed Hull in successfully arriving at equations that were more akin to the explanatory theoretical laws of physics than to its descriptive empirical laws.

On their own, equations are not mechanistic models. One equation might characterize a psychological phenomenon, and another might recharacterize it so as to provide a highly satisfactory theoretical explanation. Equations do not offer the right kind of format, however, for constructing a mechanistic explanation—they specify neither the component parts and operations of a mechanism nor how these are organized so as to produce its the behavior. This suited the mathematical psychologists of the 1950s who, like other scientifically-oriented psychologists, avoided any proposals that hinted at mentalism. When the computer metaphor made notions of internal information processing respectable, though, they began to ask what sorts of cognitive mechanisms might be responsible for phenomena that could be characterized, but not fully explained, using equations alone.

Mathematical psychology combined with symbolic mechanistic models

The development of *information processing models* in the 1960s and 1970s signaled that certain experimental and mathematical psychologists had become committed to mechanistic explanation. They did this by positing various representations that were processed (e.g., compared, moved, transformed) by operations similar to those in computer programs. Often the models were further influenced by symbolic disciplines such as linguistics and logic, making "rules and representations" a familiar phrase. (Models for visual imagery, though, usually specified analog representations and operations rather than discrete ones.) Another option was the introduction of equations with variables specifying quantitative properties of a part (e.g., the familiarity or concreteness rating of a word being encoded in memory) or of an operation (e.g., the rate of encoding).Thus, information processing models integrating computational, symbolic, and quantitative perspectives were became widespread (and still are).

In one well-known exemplar, Saul Sternberg (1966) devised a task in which a set of items, typically a list of digits, must be held in short-term memory and retrieved for comparison to a probe item. Sternberg explicitly brought together the symbolic and computational perspectives in his very first sentence: "How is symbolic information retrieved from recent memory?" That is, each item on the list was regarded as an external symbol that needed to be represented as a mental symbol, and then subjected to discrete computational operations like retrieval. He found that $RT = 392.7 + 37.9\ s$, where s is set size and RT is the mean reaction time (in msec.) to respond whether or not a single probe item was in the set of items on the just-presented list. He interpreted this as sufficient basis for rejecting a common "implicit assumption that a short time after several items have been memorized, they can be immediately and simultaneously available for expression in recall or in other responses, rather than having to be retrieved first" (p. 652). Instead, it appeared that each item was retrieved and compared to the probe in succession (a process that has been called *serial search*, *serial retrieval*, *serial comparison*, or simply *scanning*.) Moreover, because the reaction time functions were almost identical for trials in which there was and was not a match, Sternberg contended that this process was not only serial but also exhaustive. If, to the contrary, the process terminated once a match was found, positive trials should have had a shallower slope and averaged just half the total reaction time of negative trials for a given set size.

Sternberg's deceptively simple paper illustrates several key points.

- The linear equation initially played the same straightforward role as Fechner's logarithmic equation: it precisely characterized a phenomenon.
- Sternberg aspired to explain the phenomenon that he characterized, and departed from the mathematical psychology of the 1950s by proposing a mechanism, rather than a more fundamental equation, that could produce the phenomenon of a linear relation between set size and RT.
- In the proposed mechanism—one of the earliest information processing models—the most important **parts** were a short-term memory store and mental symbols representing the probe and list items, the most important **operation** was retrieval, and the system was **organized** such that retrieval was both serial and exhaustive.
- The mechanism combined computational and symbolic approaches, in that its operations were performed on discrete mental symbols.

- In addition to the computational and symbolic approaches, the mechanistic explanation incorporated the quantitative approach of mathematical psychology as well. That is, Sternberg used his linear equation not only to characterize the phenomenon, but also as a source of detail regarding an operation. Specifically, he interpreted the slope (37.9 msec.) as the time consumed by each iteration of the retrieval operation.
- The underdetermination of explanation is hard to avoid. For example, although Sternberg regarded his data as pointing to a mechanism with serial processing, competing mechanisms based on parallel processing have been proposed that can account for the same data. The slope of the linear function is interpreted quite differently in these accounts.

Information processing models of cognitive mechanisms like Sternberg's took shape in the 1960s, became more complex and dominant in the 1970s, and still play a major role today. They also were a major antecedent to connectionism and have competed with this offspring since the 1980s. Mathematical psychologists generally ally themselves with either information processing or connectionist models. As mentioned above, however, a different mathematical modeling approach from outside psychology—dynamical systems theory—caught the attention of certain psychologists focusing on perception or motor control in the 1980s and then of cognitive scientists in the 1990s.We consider it next.

Dynamical systems theory and its contributions to cognitive science

Dynamical systems theory (DST)

Researchers like linear equations; they are simple, well-behaved, and amenable to statistical analysis. Numerous phenomena of interest to cognitive scientists can be characterized using such equations, even when time is one of the variables. For example, mean IQ scores increased linearly across the years of the 20th century (a little-known and surprising fact called the *Flynn effect*, Flynn, 1987; Neisser, 1997). Another example, in which time is a dependent rather than independent variable, is Sternberg's finding that reaction time in his memory scanning task increased linearly with set size. When relationships are not linear, sometimes they can be made linear by transforming the data. For example, population growth is exponential. For purposes of analysis (unfortunately not in reality) the relationship of population size to time can be made linear by performing a logarithmic transformation. Exponential functions are common in mathematical psychology, both in characterizing phenomena and in theoretical laws proposed to explain phenomena, and in these uses present no undue difficulties.

Nature, though, is much less attached to simple linear and nonlinear functions than are the researchers trying to make sense of nature. It presents us with numerous phenomena in which the changes of state are nonlinear in complex ways. Working outside the scientific mainstream in the 20th century, the pioneers of *dynamical systems theory (DST)* developed mathematical, graphical, and conceptual tools for characterizing such phenomena and learning from them. Though DST is best-known for such exotic concepts as chaotic trajectories, Lorenz attractors, and fractals, some of its essentials can be conveyed in a relatively straightforward example (adapted from Abraham & Shaw, 1992, pp. 82-5). In the 1920s, Lotka and Volterra considered how the number of prey (x) and predators (y) in an idealized two-species ecosystem would change over time. They saw

that cyclic population swings, in which neither predators nor prey can retain a permanent advantage, are obtained from a system of just two nonlinear differential equations (if appropriate values are provided for its four parameters A-D):

$dx / dt = (A - By)\, x$

$dy / dt = (Cx - D)\, y$

The notation dx / dt refers to the rate of change in the number of prey x over time t and dy / dt to the rate of change in the number of predators y over time t. Figure 3(a) shows just three of the cyclic trajectories that can be obtained from these equations by specifying different initial values of x and y. They differ in the size of the population swings, but share the fate that whichever trajectory is embarked upon will be repeated *ad infinitum.* If the initial values are at the central equilibrium point, also shown, there are no population swings—it is as though births continuously replace deaths in each group. A more interesting situation arises if the equations are modified by adding "ecological friction" (similar to the physical friction that brings a pendulum asymptotically to rest by damping its oscillations). Now the swings decrease over time as the system spirals towards, but never quite reaches, its point of equilibrium. In this modified system the point of equilibrium is an *attractor state* (specifically, a *point attractor* in *state space*, or equivalently, a *limit point* in *phase space*)—one of the simplest concepts in DST and very influential in some parts of cognitive science. Figure 3(b) shows a point attractor and one of the trajectories spiraling towards it. The type of graphic display used in Figure 3, called a *phase portrait*, is one of DST's useful innovations. Time is indicated by the continuous sequence of states in each sample trajectory. This leaves all dimensions available for showing the state space, at a cost of not displaying the rate (or changes in rate) at which the trajectory is traversed. If graphic display of the rate is desired, a more traditional plot with time on the abscissa can be used.

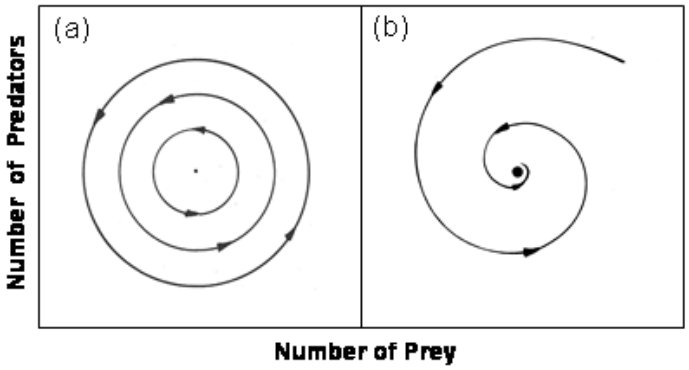

Figure 3. Two kinds of trajectories through state space for a predator and prey population: (a) cyclic trajectory (equilibrium point is not an attractor) and (b) spiraling trajectory (equilibrium point is a point attractor).

In the 1980s DST captured the attention of a few psychologists who recognized its advantages for characterizing phenomena of perception, motor behavior and development. By the 1990s DST was having a broader impact in cognitive science. As in the case of mathematical psychology before it, the use of DST progressed through phases:

- first, primarily exploring how dynamic phenomena, especially of motor behavior and development, could be characterized using the mathematical, graphical, and conceptual tools of DST;

- increasingly with experience, also extracting or applying theoretical principles, for example, metastability (see below) or the well-known "sensitivity to initial conditions" with regards to those dynamic systems that behave chaotically;
- finally, combining the theoretical principles and the tools for characterizing phenomena with other commitments so as to achieve a dynamical framework specifically tailored to cognitive science.

In recognition of the additional commitments involved in tailoring dynamical accounts to cognitive science, reference is most often made to the *dynamical approach* to cognition rather than *DST* as such. And given that individuals disagree about some of those commitments, it is best understood as a family of approaches. One of the major disagreements involves mechanistic explanation. On one view, the dynamical approach is valuable for its way of characterizing phenomena and its theoretical explanations, and that style of explanation is preferable to mechanistic explanation. Thus, the dynamical approach should be adopted in preference to symbolic, connectionist, or any other mechanistic accounts (see, for example, papers in Port & van Gelder, 1995). Opposing this are dynamical approaches that combine aspects of DST with mechanistic modeling of cognition (e.g., Elman's use of dynamical concepts such as attractors to better understand interactive connectionist models, or van Leeuwen's use of dynamical tools in constructing small-grained coupled map lattice models). Another axis of difference is the timescale involved. In addition to models targeting the fleeting dynamics of motor, perceptual, or cognitive acts, there are highly influential models of change across developmental time. In each of the next three sections we offer a glimpse of one of these dynamical approaches to cognition.

A dynamical approach with no mechanisms

Dynamicists in cognitive science who refrain from mechanistic explanation tend to advocate holism. Making the assumption that "all aspects of a system are changing simultaneously, [they] focus on how a system changes from one total state to another. . . . The distinctive character of some cognitive process as it unfolds over time is a matter of how the total states the system passes through are spatially located with respect to one another and the dynamical landscape of the system" (van Gelder and Port, 1995, pp.14-15). Timothy van Gelder contrasts this holism to several characteristics he attributes to the symbolic approach—homuncularity, representation, computation, and sequential and cyclic operation—and maintains that "a device with any one of them will standardly possess others" (van Gelder, 1995, p. 351). *Homuncularity* refers to Dennett's (1978) conception of cognitive explanation as a decomposition of cognitive processes into little agents, each responsible for one operation in the overall activity of the cognitive system. This approach is often referred to as *homuncular functionalism* (Lycan, 1979) and involves a kind of decomposition that is characteristic of mechanistic explanations. In arguing instead for a dynamical alternative, van Gelder cites James Watt's governor for the steam engine (Figure 4) and maintains that it is best understood in terms of the dynamical equations Maxwell developed for describing its activity. But Bechtel (1997) argued that the smooth overall functioning of Watt's governor offers insufficient grounds for the extreme holism espoused by van Gelder and some other dynamicists. As with any mechanism, component parts and operations can be identified: the flywheel rotates, the spindle arms become extended, the valve closes. The parts are spatially organized in ways that help organize the operations. It makes sense, for example, to ask why Watt attached the spindle arms to the spindle coming out of the

flywheel and to answer that question in terms of the resulting relationship between two of the operations: as the flywheel rotates, the angle arms become extended. These relationships can be quantified by defining variables corresponding to properties of the operations and measuring their values under a variety of circumstances; Maxwell's equations economically express the outcome of such a procedure. They permit, for example, prediction of the angle of the spindle arms given any particular speed at which the flywheel rotates. This is very useful, but does not imply holism.

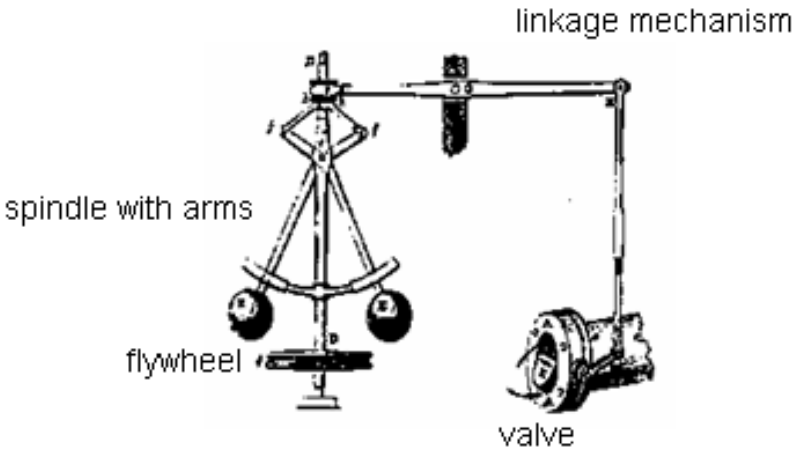

Figure 4. Watt's centrifugal governor for a steam engine. Drawing from J. Farley, *A Treatise on the Steam Engine: Historical, Practical, and Descriptive* (London: Longman, Rees, Orme, Brown, and Green, 1827).

In addition to their holism, these dynamicists also have a preference for parsimony that is similar to that of classic mathematical psychologists (those who had not yet adopted a mechanistic information processing framework within which to apply their mathematical techniques). A simple dynamical account typically targets phenomena involving a small number of quantitative variables and parameters. One of the earliest examples relevant to cognitive science is Scott Kelso's account of the complex dynamics evoked in a deceptively simple task used to study motor control. Participants were asked to repeatedly move their index fingers either in phase (both move up together, then down together) or antiphase (one moves up while the other moves down) in time with a metronome (Kelso, 1995). As the metronome speed increases, people can no longer maintain the antiphase movement and involuntarily switch to in-phase movement. In DST terms, the two moving fingers are *coupled oscillators*. Slow speeds (low frequencies) permit a stable coupling either in-phase or antiphase: the state space has two attractors. At high speeds only in-phase coupling is possible: the state space has only one attractor. Kelso offered a relatively simple equation that provides for characterization of the attractor landscape (V) as parameter values change:

$$V = -\varphi\delta\omega - a\cos\varphi - b\cos 2\varphi$$

Here φ is the degree of asynchrony between the the two fingers (*relative phase*), $\delta\omega$ reflects the difference between the metronome's frequency and the spontaneous oscillation frequency of the fingers, and a and b indirectly reflect the actual oscillation frequency of the fingers. When the ratio b / a is small, oscillation is fast and only the in-phase attractor exists. When it is large, there are two attractors: people can produce in-phase or antiphase movement as instructed or voluntarily traverse a trajectory between them.

Things get interesting when the ratio *a* / *b* is intermediate between values that clearly provide either one or two attractors. The attractors disappear but each "leaves behind a remnant or a phantom of itself" (p. 109). The system's trajectory now exhibits *intermittency*, approaching but then swinging away from the phantom attractors. Putting it more concretely, the two index fingers fluctuate chaotically between in-phase and antiphase movement. Although such intermittency may not seem particularly important, if we shift to a different domain we can recognize it as a significant characteristic of cognition. Most people have had the experience, when looking long enough at an ambiguous figure such as the Necker cube, of an irregular series of shifts between the two interpretations. The temporal pattern of these spontaneous shifts is *chaotic*—a technical concept in DST that refers to trajectories in state space in which no point is revisited (such trajectories appear random but in fact are deterministic). Kelso remarked on the similarity between the finger-movement and Necker cube tasks, not only in the switching-time data he displayed but also in the quantitative and graphical analyses of the systems. Given the wide range of applicability of these analyses, their elegance, and their depth, they clearly go beyond characterizing phenomena to explaining them via recharacterization. Nonetheless, they provide only one of the two types of explanation we have discussed. In the next section we introduce an alternative dynamical approach that offers intriguing new ideas about mechanistic explanation.

A dynamical approach with subsymbolic mechanisms

Kelso chose to focus on switching time phenomena and center his explanation on a single dynamical equation with variables capturing aspects of the system as a whole. Here we look at a mechanistic explanation for the same phenomenon. Cees van Leeuwen and his collaborators (van Leeuwen, Steyvers, & Nooter, 1997) developed a model that is both dynamical and mechanistic by using a dynamical equation to govern (not only to describe) the operations of the fine-grained component parts of a mechanism. It is similar to a connectionist model insofar as its component parts are numerous homogeneous units, some with pairwise connections, that become activated in response to an input. However, in this case the units are more sparsely connected and are designed as oscillators that can synchronize or desynchronize their oscillations. Particular patterns of synchronization across the units are treated as constituting different interpretations of the input.

More specifically, van Leeuwen et al. employed a *coupled map lattice (CML)* of the sort first explored by Kaneko (1990). A *lattice* is a sparsely connected network in which only neighboring units are connected (*coupled*); the basic idea is illustrated by a large construction from tinkertoys or a fishnet. A *map* is a type of function in which values are iteratively determined in discrete time. Kaneko employed the logistic equation

$$x_{t+1} = \mathrm{A}\, x_t (1 - x_t)$$

to govern the activation (x) of units at a future time t+1 on the basis of their current activation. Depending on the value of the parameter (A), such units will oscillate between a small number of values or behave chaotically. For nearby units to influence each other, there must be connections between them. Although van Leeuwen's proposed mechanistic model used 50 x 50 arrays of units, with each unit coupled to four neighbors, his analysis of a CML with just two units

suffices for illustration. In his account, the net input to unit *x* is determined from the activation of *x* and the other unit *y*:

$$netinput_x = C\ a_y + (\ 1 - C\ )\ a_x$$

The logistic function is then applied to the resulting net input to determine the activation of unit *x*:

$$a_{x,t+1} = A\ netinput_{x,t}\ (1 - netinput_{x,t})$$

The behavior of the resulting network is determined by the two parameters, A and C. For a range of values of A relative to C, the two units will come to change their activation values in synchrony. Outside this range, however, the system exhibits the same kind of intermittency as did Kelso's higher-level system. For some values, the two units approach synchrony, only to spontaneously depart from synchrony and wander through state space until they again approach synchrony. With larger networks, one constellation of nearby units may approach synchrony only to break free; then another constellation of units may approach synchrony. These temporary approaches to synchrony were interpreted by van Leeuwen et al. as realizing a particular interpretation of an ambiguous figure. Thus, there are echoes of Kelso not only in the achievement of intermittency but also in its interpretation. Nonetheless, van Leeuwen et al's fullscale CML is clearly mechanistic: it explains a higher-level phenomenon as emerging from the dynamics of numerous, finer-grained parts and operations at a lower level. This contrasts with the more typical dynamical approach, exemplified by Kelso, of describing a phenomenon at its own level using just a few variables in a global equation.

Dynamical approaches to development

The earliest dynamical approaches to development were more in the vein of Kelso than of van Leeuwen. Notably, Esther Thelen (1985) initially drew attention to dynamics by showing how global equations could account elegantly for phenomena of infant motor development. She saw bipedal locomotion, for example, as residing "in the dynamics of two coupled pendulums" (limbs). Collaboration with colleagues at Indiana University, especially Linda Smith and Michael Gasser, yielded a more ambitious, potent package of theoretical elaborations and extended the empirical work to cognitive and linguistic development. Their fresh ideas are best conveyed in quotations (taken from Smith & Gasser, 2005): "Babies begin with a body richly endowed with multiple sensory and action systems." Those systems are coupled to a heterogeneous physical world and also, via "smart social partners," to social and linguistic worlds. The developing intelligence of babies is embodied. That is, it resides "not just inside themselves but . . . distributed across their interactions and experiences in the physical world"—interactions that are exploratory, multimodal, incremental, simultaneous, and continuous. The physical world is "full of rich regularities that organize perception [and] action" and offer opportunities for offloading aspects of cognition that might otherwise require internal inferences or predicate bindings. Experience with this world "serves to bootstrap higher mental functions," but once in the system, language in particular provides computational power that "profoundly changes the learner."

These ideas did not take shape in an armchair, but rather emerged as the Indiana team extended their range to a variety of developmental domains and moved towards a more mechanistic style of dynamical modeling. Though some of the mechanisms they have proposed are connectionist networks, most innovative is their dynamic field model of the A-not-B error in Piaget's classic

object concept task (Thelen, Schöner, Scheier, & Smith, 2001). They explored many variations of this task, but essentially a researcher places two cups in front of a baby (7 to 12 months old) and repeatedly hides an enticing object under the cup on the left. Each time it is hidden, the baby retrieves it. Then, making a great show of it, the researcher hides the object under the cup on the right. The baby watches what is happening on the **right** like a hawk, but then picks up the cup on the **left** and hence fails to get the desired object. What went wrong? Piaget's interpretation involves an intermediate stage in development of the concept of object. In the dynamic field model, equations specify activation functions that are continuous across the left-right spatial field, coupled, and dynamically changing at multiple timescales. Essentially, the influence of the memory field (which has a peak on the left) can override that of the perception field (which has a peak on the right during the critical trial) as they dynamically mesh in the motor planning field. If an above-threshold peak on the left dominates that field, the baby will reach to the left. Nothing in these exquisite equations could be interpreted as a representation of an object. Although this account does not rule out positing that such a representation also exists, this strikes Thelen and her colleagues as superfluous and nonparsimoneous.

Dynamical approaches to development have also been adopted by cognitive scientists whose initial domains of interest were language and cognition rather than motor development and whose modeling preferences have ranged from the competition model to connectionist networks. For a lucid review bristling with examples and recent innovations see MacWhinney . For a longer exploration that is thought-provoking but accessible even to novices, see Elman et al.'s (1996) *Rethinking Innateness*, The product of an unusual collaboration involving six leading developmental theorists, this wide-ranging book includes a number of illustrations of different roles played by equations.

We have already seen, in the differing accounts of the U-shaped function relating age and past-tense formation for irregular verbs, that dissimilar mechanistic proposals can compete as explanations for a single nonlinear developmental phenomenon. Sometimes these disputes are rooted in choices researchers make at the outset about the equations used to characterize the phenomenon. *Rethinking Innateness* offers a very readable guide to the kinds of equations used and their implications. It also covers both kinds of explanation for nonlinear phenomena: mechanistic models (in the form of connectionist networks) and global theoretical equations. One example discussed by Elman et al. involves the shape of early vocabulary acquisition: slow growth to about 50 words and then a "vocabulary burst" initiating a period of rapid growth that can add several hundred words in a few months. The most obvious way to characterize this phenomenon mathematically is to posit two separate linear equations with different slopes, one for the first 50 words and the other beyond 50 words. Among the explanations invited by this characterization are a sudden insight that things have names (Baldwin, 1989) and a change in categorization (Gopnik & Meltzoff, 1987). These explanations seem far less apt when the same vocabulary data are characterized using a single nonlinear equation. Elman et al. offered the following exponential function for data from 10-24 months of age (p. 182):

$$y = y_0 e^{b(t-t_0)}$$

where the number of words known at t_0 (10 mo.) is y_0 and at any other t is y. When displayed graphically, y is seen to increase slowly, then bend upwards, then increase rapidly. Thus, it can fit the acquisition data about as well as the two straight lines—so well that a choice cannot be made based on fit. What is appealing about the single nonlinear function is that there is an

underlying simplicity. The amount *y* will increase in the next unit of time is proportional to the current value of *y*:

$$dy / dt = by$$

where *dy / dt* is a notation indicating that we are concerned with the number of additional words to be added per unit of time, *b* is the percentage increase across that time (which does not change), and *y* is the current number of words (which increases with *t*). This equation arguably is explanatory, in that it recharacterizes the first equation so as to explicate that a constant *percentage* increase is responsible for its exponential *absolute* increase.

Elman et al. emphasize that a single mechanism can produce slow initial vocabulary growth as well as a later vocabulary burst. This does not mean nothing else can be involved; they themselves go on to consider a more complicated version of their initial account. The point is that the burst, as such, does not require any special explanation such as a naming insight; "the behavior of the system can have different characteristics at different times although the same mechanisms are always operating" (p. 185). In another part of their book (pp. 124-9), Elman et al. describe an autoassociative connectionist network that, though limited to simplified representations, suggests a mechanistic explanation. It replicates not only the quantitative phenomenon of a vocabulary burst (exponential growth), but also such phenomena as a tendency to underextension prior to the burst and overextension thereafter.

Abrupt transitions are not uncommon in development, and connectionist networks excel at producing this kind of nonlinear dynamic. Another example discussed by Elman et al. targets the transitions observed by Robert Siegler (1981) when children try to solve balance scale problems. They appear to progress through stages in which different rules are used (roughly: weight, then distance, then whichever of these is more appropriate). McClelland and Jenkins (1991) created a connectionist network that offered a simplified simulation of the abrupt transitions between these three stages. While intrigued by this capability, Elman et al. commented that "simulating one black box with another does us little good. What we really want is to understand exactly what the underlying mechanism is in the network which gives rise to such behavior" (p. 230).

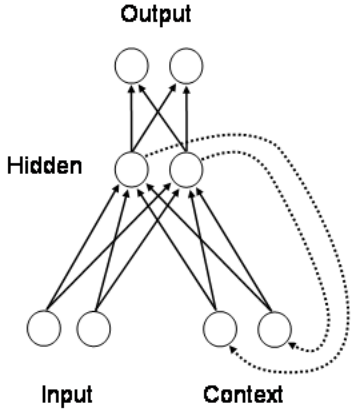

Figure 5. Simple recurrent network as proposed by Jeffrey Elman (1990). For clarity, fewer units are shown than would be typical. The solid arrows represent feedforward connections used to connect each unit in one layer with each unit in the next layer. The dotted arrows indicate the recurrent (backwards) connections that connect the hidden units to specialized input units called *context units*.

To address this question they explored a type of connectionist network that exhibits time dynamics in its behavior. Unlike feedforward networks, such as those used in the past-tense, vocabulary, and balance-scale simulations, *recurrent networks* include recurrent (backwards) connections between certain sets of units (Figure 5). Rather than a single forward pass of activation changes, each unit in a recurrent network repeatedly has its activation recalculated as the network (typically) settles gradually into a stable response to the input—an attractor state. Elman et al. began by training a recurrent network to perform a simple demonstration task: determining parity for a long string of 1s and 0s (that is, deciding whether the number of 1s is odd or even). They succeeded in simulating the phenomenon of abrupt transitions; in fact, after 17,999 training cycles the network could not do the task and after just one more training trial it could. To understand why, they simplified further by examining a single unit with a nonlinear activation function commonly employed in connectionist modeling. This unit starts with an initial activation and on successive time-steps receives input from a constant input (called a *bias*) and a recurrent connection from itself. With a weight on the recurrent connection of 1.0 and a bias of -0.5, the unit after several iterations settles to an activation of 0.5 and remains there. With a weight of 10.0 and a bias of -5.0, however, the unit will settle to an activation of 0.0 when its initial activation is below 0.5, to an activation of 1.0 when its initial activation is above 0.5, and to 0.5 when its initial activation is 0.5. These two kinds of behavior reflect a network unable to retain information versus one able to retain information. Further exploration of values of weight and bias enabled the researchers to map the response of such a unit. They found a region in which a very small change in the weight created a large change in behavior, and noted that DST could provide exactly the right tools for pursuing and understanding this kind of nonlinearity.

Conclusions

Our strategy through this paper has been to show that the range of phenomena for which mechanistic models are sought is extremely varied and to illustrate briefly some of the kinds of models of mechanisms that have been advanced to account for different phenomena. The focus in the philosophical literature on systematicity and other general properties of cognitive architectures presents a distorted view of the actual debates in the psychological literature over the types of mechanisms required to account for cognitive behavior. Even in the domain of language, where systematicity is best exemplified, many additional phenomena claim the attention of cognitive scientists. We discussed two that are quantitative in nature: the U-shaped acquisition of irregular past tense forms and the exponential acquisition of early vocabulary. Beyond language we have alluded to work targeting the rich domains of perceptual and motor behavior, memory, and problem solving. Phenomena in all of these domains are part of the explanandum of mechanistic explanation in cognitive science. Such explanatory attempts, which like the phenomena themselves often are quantitative, go back as far as Weber's psychophysics and currently are moving forward in dynamical approaches to perception, cognition and development.

We have also emphasized that cognitive science, despite its many disputes, has progressed by continually combining and recombining a variety of influences. The use of equations both in characterizing and explaining phenomena are among these. When combined with other influences and commitments, the outcomes discussed here have ranged from information processing models with quantified operations to connectionist networks to both global and

mechanistic dynamical accounts. Each of these approaches has provided a different answer to the question of whether the mind processes language-like representations according to formal rules, and we have argued that the overall answer need not be limited to just one of these. Cognitive science takes multiple shapes at a given time, and is protean across time.

"I am large, I contain multitudes." (Walt Whitman, 1855, Leaves of Grass. Brooklyn: Rome Brothers)

References

Abraham, R. H., & Shaw, C. D. (1992). *Dynamics: The geometry of behavior*. Redwood City, CA: Addison-Wesley.

Baldwin, D. A. (1989). Establishing word-object relations: A first step. *Child Development, 60*, 381-398.

Bechtel, W. (1997). Dynamics and decomposition: Are they compatible? *Proceedings of the Australasian Cognitive Science Society*.

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks* (Second ed.). Oxford: Blackwell.

Bechtel, W., & Abrahamsen, A. (in press). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Bechtel, W., Abrahamsen, A., & Graham, G. (1998). The life of cognitive science. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 1-104). Oxford: Basil Blackwell.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Princeton, NJ: Princeton University Press.

Brown, R. (1973). *A first language: The early stages*. Cambridge: Harvard University Press.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review, 58*, 313-323.

Cummins, R. (2000). "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117-144). Cambridge, MA: MIT Press.

Dennett, D. C. (1978). *Brainstorms*. Cambridge, MA: MIT Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Ervin, S. (1964). Imitation and structural change in children's language. In E. Lenneberg (Ed.), *New directions in the study of language*. Cambridge, MA: MIT Press.

Estes, W. K. (1950). Towards a statistical theory of learning. *Psychological Review, 57*, 94-107.

Estes, W. K. (2002). Mathematical psychology, History of, *International encyclopedia of the social and behavioral sciences* (pp. 9412-9416). New York: Elsevier.

Falmagne, J.-C. (2002). Mathematical psychology, *International encyclopedia of the social and behavioral sciences* (pp. 9405-9412). New York: Elsevier.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.

Garcia, J., McGowan, B. K., Ervin, F. R., & Koelling, R. A. (1968). Cues: Their relative effectiveness as a function of the reinforcer. *Science, 160*, 794-795.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 50-71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342-S353.

Gopnik, A., & Meltzoff, A. N. (1987). The development of categorization in the second year and its relation to other cognitive linguistic developments. *Child Development, 58*, 1523-1531.

Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331-496). New York: Macmillan.

Kaneko, K. (1990). Clustering, coding, switching, hierarchical ordering, and control in a network of chaotic elements. *Physica D, 41*, 137-142.

Kelso, J. A. S. (1995). *Dynamic patterns: The self organization of brain and behavior*. Cambridge, MA: MIT Press.

Kitcher, P. (1999). Unification as a regulative ideal. *Perspectives on Science, 7*, 337-348.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16*, 589-600.

Luce, R. D., Bush, R. R. B., & Galanter, E. (Eds.). (1963-1965). *Handbook of mathematical psychology*. New York: Wiley.

Lycan, W. G. (1979). Form, function, and feel. *Journal Philosophy, 78*, 24-49.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

McClelland, J. L., & Jenkins, E. (1991). Nature, nurture and connectionism: Implications for connectionist models of development. In K. van Lehn (Ed.), *Architectures for intelligence: The twenty-second (1988) Carnegie Symposium on cognition* (pp. 41-73). Hillsdale: Erlbaum.

McClelland, J. L., & Patterson, K. (2002a). 'Words or Rules' cannot exploit the regularity in exceptions. *Trends in Cognitive Sciences, 6*, 464-465.

McClelland, J. L., & Patterson, K. (2002b). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6*, 465-472.

Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace.

Neisser, U. (1997). Rising scores on intelligence tests. *American Scientist, 85*, 440-447.

Newport, E. L., & Aslin, R. N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 1-21). Somerville, MA: Cascadilla Press.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73-193.

Pinker, S., & Ullman, M. T. (2002a). The past and future of the past tense. *Trends in Cognitive Sciences, 6*, 456-463.

Pinker, S., & Ullman, M. T. (2002b). Combination and structure, not gradedness, is the issue. *Trends in Cognitive Sciences, 6*, 472-474.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron. *Cognition, 38*, 43-102.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21-69.

Port, R., & van Gelder, T. (1995). *It's about time*. Cambridge, MA: MIT Press.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Rutgers Center for Cognitive Science, TR-2. Rutgers Optimality Archive 537): Rutgers University.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science, 275*, 1604-1610.

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986a). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2. Psychological and biological models*. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986b). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations*. Cambridge, MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Siegler, R. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46*(2).

Smith, L. B., & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life, 11*, 13-30.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153-181.

Suppe, F. (1974). The search for philosophical understanding of scientific theories. In F. Suppe (Ed.), *The Structure of Scientific Theories* (pp. 3-241). Urbana: University of Illinois Press.

Tesar, B., Grimshaw, J., & Prince, A. (1999). Linguistic and cognitive explanation in optimality theory. In E. Lepore & Z. W. Pylyshyn (Eds.), *What is cognitive science* (pp. 295-326). Oxford: Blackwell.

Thelen, E. (1985). Developmental origins of motor coordination: Leg movements in human infants. *Developmental Psychobiology, 18*, 1-22.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences, 24*, 1-34.

van Gelder, T. (1995). What might cognition be, if not computation. *The Journal of Philosophy, 92*, 345-381.

van Leeuwen, C., Steyvers, M., & Nooter, M. (1997). Stability and intermittency in large-scale coupled oscillator models for perceptual segmentation. *Journal of Mathematical Psychology*.

Weber, E. H. (1834). *De pulsu, resorpitione, auditu et tactu: Annotationes anatomicae et physiologicae*. Leipzig: Koehlor.