Parallel distributed processing (PDP) is a theory of cognition that builds on developments in cognitive science over the past decade or so, but which emerged as a full-blown intellectual paradigm only in 1986 (McClelland, Rumelhart & the PDP Research Group 1986; Rumelhart, McClelland & the PDP Research Group 1986). PDP has already attracted intense interest and controversy in nearly every discipline concerned with human cognition, behavior and learning. The theory is intended to deal with topics as diverse as motor control, visual perception, language, and memory, and has a wildly optimistic goal: to provide a mathematical model that captures the essence of both neural processing, on the one hand, and thought on the other. PDP immediately raised a provocative challenge to established views of language and language learning by providing a stunning demonstration (by computer simulation) of how learners could acquire the English past tense through the operation of parallel distributed processing (Rumelhart & McClelland 1986).

PDP theory assumes that learning is based on the processing of input and in that respect is similar to most theories of SLA that warrant serious consideration. However, in nearly every other respect, PDP contradicts the dominant paradigms of the past three decades. In SLA, for example, under the influence of generative grammar, the dominant view has been that the essence of what learners acquire is a system of implicit (unconscious) and highly abstract rules. Language acquisition is thought of as the process of inducing such rules, a complex process that depends on an innate component that severely restricts the class of possible rules. The system of rules thus induced (competence) is only imperfectly revealed in performance, which is variable and noisy, although performance data, including acquisition facts, are sometimes used as arguments for the existence of such rules. A good example of such an argument from acquisition data has to do with the past tense in English, the general facts of which are well known: children begin by acquiring...
a few high frequency irregular forms (e.g. *went*), then produce not only regular past tense forms encountered in input but also incorrectly regularized forms such as *goed*, sometimes also producing doubly marked forms such as *wented* before finally marking certain verbs as exceptions and recovering the irregular forms. Stage 2 of this developmental sequence is crucial, since it is often considered evidence for the existence of an internalized, implicit rule which overrides evidence from input.

In contrast, the central claims of PDP theory are:

1. Information processing takes place through the interactions of a large number of simple units, operating in parallel, each sending excitatory and inhibitory signals to other units within networks of multiple, simultaneous and often mutual constraints. An explicit analogy between units and their interconnections on the one hand and neurons and synapses on the other suggests a direct link between neural hardware and cognitive processing.

2. Learning takes place through the strengthening and weakening of the interconnections in a particular network in response to examples encountered in input. Changes in the weights of interconnections spread throughout the entire system according to computationally complex principles. With sufficient input, connection strengths mirror the regular patterns found in the language.

3. The result of learning is a network of simple units that acts as though it knew the rules, but the "rules" themselves exist only in the form of association strengths distributed across the entire network. Higher level regularities are emergent properties of processes taking place on a micro-level. They are the result of the operation of the system itself, which is dynamic, interactive and self-organizing. However, they are not part of the system in the sense of being explicitly represented in it, and are not principles which in any sense guide or control the lower level processes.

As a demonstration, Rumelhart and McClelland constructed a mechanism that learned to produce the past tense forms of English verbs based on the phonological shape of their root forms. The simulation used a simple PDP system with just two types of units: input-units corresponding to the root forms of various verbs and output-units corresponding to various possible past tense forms, both of which were represented by clusters of phonological features. The model was trained through the presentation of
present-past tense pairs, fed into the mechanism according to their frequency in English. The goal of the model was to learn the right connection strengths between root and past-tense units so that whenever the root form of a new verb was presented the model would construct the appropriate past tense form. The model learned to do so. It handled both regular and irregular verbs, distinguished the regular allomorphs of the past tense morpheme, and produced past tense forms for verbs on which it was not trained. It went through the stages that children go through, including the stage of overregularization. It acquired subclasses of irregular verbs in a similar order to that documented for children. It did all of this without anything that could be called a rule ever being represented in the mechanism. When a new verb was presented, the mechanism produced a past tense form not by reference to a rule or general principle, but rather as the outcome of inhibitory and excitatory action from all root-past tense pairings currently represented in the system, a kind of massive on-line blending process. Rumelhart and McClelland conclude, "We have shown that a reasonable account of the acquisition of past tense can be provided without recourse to the notion of a ‘rule’ as anything more than a description of the language. We have shown that, for this case, there is no induction problem. The child need not figure out what the rules are, nor even that there are rules" (Rumelhart & McClelland 1986: 267).

Proponents of PDP intend the theory to be a major revolution in the way in which we think about human intelligence. It is not clear whether this intended revolution will succeed. Initial reactions have ranged from enthusiastic (Sampson 1987) to highly skeptical (Pinker & Prince 1987). Sampson suggests that no branch of linguistic inquiry will be unaffected by the PDP paradigm, and argues that PDP resolves with one stroke some of the most resistant problems of the generative paradigm: the relationship between competence and performance, variability, and the nature of the interactions between innate knowledge and primary linguistic data.

1 It can be argued that such networks are much less efficient than rule based systems in which generalizations are abstracted from the data. This might seem a convincing argument if one thinks of present-day computers as a metaphor for the mind, but one of the major arguments put forth by proponents of PDP (and many others) is that the human brain is not an efficient calculating device like the computer. PDP is an attempt to account not only for the inefficiency of the brain as a sequential calculator but also for the fact that, ultimately, humans are much smarter than computers.
Other initial reactions have been much more negative, for both legitimate and illegitimate reasons. I would classify as illegitimate a variety of responses that are essentially variants on the theme of "must I bother to find out about this stuff?" For example, the model may be unappealing to some on emotional grounds; PDP and other "connectionist" models may seem too reminiscent of a discredited behaviorism to be appealing, thus discouraging consideration of what real (as opposed to assumed) limitations PDP may have. The algebraic formalisms in papers about PDP are certain to drive many away; Sampson points out that most language teachers, arts undergraduates and linguists will immediately and correctly realize that they cannot deal with the mathematics of PDP analysis.

Much more serious objections have been raised by Pinker and Prince (1987), who claim that the PDP model does not do the job it is said to do in simulating the past tense forms of English verbs, not only because it produces some striking bloopers (e.g. membled as the past tense of mail) but because the model in principle cannot learn some rules that occur in natural languages but can learn rules found in no human language. Some of Pinker and Prince's criticisms are directed against the unorthodox theory of phonology (based on "wicklephones" and "wicklefeatures", after Wicklegren 1969) that is used to represent verb roots in the Rumelhart and McClelland computer simulation, a devastating criticism only if it can be shown that PDP models cannot operate with more defensible phonological theories. Other criticisms are directed at flaws in PDP architecture itself. Pinker and Prince especially aim to show that PDP fails because of the extreme reductionist nature of the model. Not only are "rules" and "exceptions" for past tense not represented in the model, none of the individual units or connections in the model corresponds to a word, a position within a word, or a morpheme. Except for wicklefeatures, the units are conceptually empty, with no symbolic representation at all, and it is this "eliminative connectionism" that Pinker and Prince show to be inadequate. For example, since the mechanism operates solely on the basis of the phonological shape of verb roots, it cannot in principle distinguish between such outputs as "the conductor braked ("broke") the train" and "he broke ("breaked") the dish."

On the other hand, Pinker and Prince praise two aspects of connectionist models (not limited to PDP): they are type- frequency sensitive
(the more the mechanism encounters a given pattern, the stronger are its graded representations of that pattern), and there is competition among various output candidates, the resolution of which depends on the strength of regularities in the input data. Pinker and Prince are generally sympathetic to what they call "implementational connectionism," in which PDP models occupy an intermediate level between symbol manipulation and neural hardware, with linguistic categories standing for the units in PDP networks (Stemberger 1985 and Waltz & Pollack 1985 exemplify such models). In this way, linguistic theories and connectionist models can be seen as complementary.

Whether implementational connectionism will resolve many potential points of conflict between PDP and linguistic accounts of particular phenomena remains to be seen. PDP theorists certainly appear willing to allow linguistic constructs of various sorts into their systems. McClelland and Kawamoto (1986) present a PDP mechanism for assigning case roles to NP's (output-units with conceptual content) on the basis of interactive word-order and semantic constraints (input-units that also have conceptual content), as a demonstration of some of the properties of the interface between syntax and semantics. On the linguists' side, there also seems to be room for compromise, at least on the question of rules. Chomsky, for example, discusses phrase structure and transformational rules as "anachronisms," now replaced by constraints, parameters and principles in linguistic theory, while for performance, "it also may be that the best theory will depart from the model of rule following altogether in any standard sense of the notion of rule, both for the receptive and the productive side of linguistic behavior" (Chomsky 1986: 243). On the other hand, linguistic theory of the type practiced by Chomsky has always taken an extreme rationalist position, assuming that the child (or a mechanism internal to the child) automatically draws all possible inferences from linguistic data, storing the most abstract generalizations possible. This is precisely what PDP claims does not happen, and it may be the notion of distributedness rather than the parallel processing aspect of PDP on which compromise will not be possible.

For SLA, PDP models suggest a new basis for understanding how learning can occur spontaneously, as a by-product of input processing. At the very least, if PDP prevails, it will force a revision in the most common metaphor we use when talking about acquisition, the learner as linguist. We
have grown used to accounts that talk about child or adult learners as grammarians (White 1987 is an excellent example). Learners are said to start with certain assumptions about language, to be on the lookout for certain crucial data, and to proceed to add, delete, and reorganize rules, or (alternatively) to confirm or reject hypotheses, all part of an attempt to construct a theory of the language they are learning. The problem with this kind of account has always been that the thought processes attributed to learners --making assumptions and using inductive and deductive reasoning to draw inferences and make abstract generalizations-- are typical of conscious reasoning of the highest order, while there is little or no evidence that learners (especially children) are consciously engaged in such a rational enterprise when learning language. The typical solution to this has been to assume that all of this reasoning is going on unconsciously, and that unconscious processes are similar to conscious thought processes except that for some reason, perhaps because they are carried out in a special code, they are inaccessible to introspection and awareness (Karmiloff-Smith 1986). Or perhaps it is not the learner himself or herself who is carrying on the rational enterprise, but the acquisition device or the grammar itself, as a kind of being within the being. PDP, on the other hand, asserts that there is a qualitative difference between conscious and unconscious thought processes. Conscious thought is serial, deliberate and slow, but displays great range and flexibility and is crucial for the manipulation of propositional representations, the formulation of abstract hypotheses, and so on. Unconscious thought, on the other hand, is fast and efficient, because it is massively parallel, but depends on the resolution of probabilistic constraints rather than engaging in problem solving by sequential reasoning or the abstraction of generalizations from data (Norman 1986).

PDP also suggests that SLA researchers were misguided --but nevertheless on to something-- in the 1970's, when there was great debate over the issue of when a rule could be said to have been acquired (first appearance?, 90% suppliance in obligatory contexts?). PDP suggests that what researchers were looking for, evidence that the learner had made an inductive leap from data to rule, simply does not occur. SLA research was on to something, however, in calling attention to the gradual, incremental modifications that are observed in acquisition over time, which in the PDP
view are the true reflection of underlying processes.

For SLA, perhaps even more significant than these conceptual shifts is the range of acquisitional phenomena that are relevant to the PDP model. At the present time, the intersection between linguistic theory and SLA research is extremely narrow, limited to the most productive rules of "core grammar," with most research concentrated on a few favorite rules (pro-drop, subjacency, etc.) that do not even approximate the variety of acquisitional tasks that learners must resolve. Linguistic theory deals poorly with aspects of acquisition that require the resolution of simultaneous, mutual constraints across linguistic levels (phonology, syntax, semantics, lexicon), whereas PDP analysis appears highly relevant for understanding the complex interplay of multidimensional constraints that are apparent in SLA data (Young, forthcoming). Current linguistic theory deals even less well with the acquisition of aspects of language that are only partially regular, falling somewhere in between complete idiosyncracy/idiomaticity and full productivity, a continuum that includes most of the grammar of most languages. The PDP paradigm is well suited for handling fuzzy categories and variability, since it exploits regularity in the input while allowing exceptions and regular patterns to coexist in the same network and does not ascribe any special status to the regular cases. Some acquisitional tasks for which PDP analysis looks promising are the English systems underlying word-stress, the articles, and the modals, all messy phenomena on which SLA theories based on linguistic theory have been essentially silent.

The PDP perspective may also have some implications for language teaching. Because it is input-based, it is in a sense compatible with all approaches to teaching that emphasize the importance of providing input, the more the better. However, the theory does not assume an anthropomorphized acquisition device that searches the input for whatever it needs. Instead, the model produces generalizations on the basis of whatever in the input is attended to or "noticed" by the learner. The McClelland and Rumelhart computer simulation of past tense learning assumes the following preliminary steps (taken care of by programmers, acting as "teachers" and providing input to the system): the problem to be solved is in the domain of verb tense; phonological shape is a likely basis for variation in inflectional morphology; and certain verb forms in input are to be paired with certain others as the basis for generalization. The first two of these prerequisites could
be satisfied by universal predispositions, but the third could not. The model is therefore compatible with a view of the role of consciousness in language learning (Schmidt, forthcoming) that requires that learners attend to form in input-processing and not only notice what is present but achieve a low level understanding of it, for example, that a verb form comprehended in input is the past tense equivalent of a present tense verb form that is already known. The model is similarly compatible with teaching approaches that emphasize not only the provision of comprehensible input but the calling of learners' attention to relevant aspects of the input.

What the theory suggests is that learners do not need to become aware of higher level generalizations implicit in the input, and do not have to discover rules or understand rules provided by a teacher. However, PDP is silent on the question of whether higher-level awareness and teaching of the traditional sort (including giving learners rules) might be facilitative, as opposed to necessary, in the acquisition of language. It seems to me plausible that conscious understanding at a more abstract level might be facilitative.

Consider the case of French gender. According to Tucker, Lambert and Rigault (1977), native speakers of French are unable to describe the basis for their correct, but seemingly arbitrary, gender classifications. When pressed for explanation, native speakers come up with implausible reasons for gender-choice, for example that a noun is masculine if it sounds "harsh", but feminine if it sounds "delicate." Native speakers apparently are not aware of the fact that gender is variably predictable, depending on the phonological ending of the noun, e.g. all words ending in -ation are feminine, whereas those ending in -oire are almost evenly distributed between masculine and feminine. This is a case where it is difficult to speak of "rules" in the usual sense, since what is found is a graded distribution from fully predictable cases to completely unpredictable ones, a good candidate for a PDP type explanation.2

Assuming that a PDP analysis could account for all the regularities in the input eventually being mirrored in output in this case, it would do so

---

2 Probabilistic regularities may lie behind dichotomous judgments. Tucker et al. found that when the ending reliably predicts the gender, native speakers consistently assign novel nouns to that gender. When phonological shape could only probabilistically determine gender, some subjects chose feminine and some masculine, though each subject made a dichotomous decision.
without ever requiring awareness of general principles. Learners would have to attend to the phonological shape of nouns in input, but not required to appreciate the significance of phonological shape for gender determination or to become aware of the strengths of variable correlations. However, it might be facilitative if learners knew that the endings of noun are what they need to pay attention to, thus improving the input-encoding on which PDP mechanisms can operate. It might also be facilitative if learners knew that some endings reliably predict gender whereas others do not, so that they would know in which cases they should pay particular attention to individual examples. Tucker *et al.* made the same suggestion (without benefit of the PDP paradigm), finding some support for it in a classroom acquisition study in which subjects were given explicit information regarding the regularities in gender assignment. Finally, although the model describes how generalizations can come about without any inductive leap, what about the fact that learners often do make such inductive leaps consciously? PDP theory is not very clear on the possibility of serial processing and conscious problem solving leading to the restructuring of systems operating in parallel, although Rumelhart, Smolensky, McClelland and Hinton (1986) consider the operation of mental models and mental simulations, suggesting that such representations can indeed be internalized, though in ways that cannot yet be specified.

PDP theory appears revolutionary in many respects. Neither enthusiastic conversion nor quick dismissal seem to be reasonable responses at the present time. I will hazard two predictions. (1) PDP is unlikely to survive counterattacks from the regime in power without considerable revision, and the greatest impact of PDP might ultimately be its role in calling attention to a wider variety of connectionist models that would not otherwise receive wide attention. (2) Current accounts of language and language acquisition will not remain unchanged, but will have to respond to the PDP challenge. We can expect some PDP-inspired accounts of SLA in the near future. (Should this not happen, it would be a rather disappointing sign of isolation from the intellectual mainstream.) Such accounts should be very healthy for SLA research. One of the difficulties with TG-inspired theories is

---

3 I am not arguing that learners need to consciously (i.e. deliberately) pay attention to noun endings in this case, since Slobin-type input processing universals suggest that we do that anyway, but only that awareness of the locus of the crucial data might facilitate improved input-encoding, affecting the rate of acquisition.
that the route from linguistic theory to predictions about acquisition often involves some rather tortuous logic, with alternative explanations readily available for whatever facts might turn up. This is not the case with PDP modeling. Even its harshest critics agree on one point: the model is explicit, and makes extremely precise predictions about what the acquisitional facts should be, some of which are not at all obvious until the appropriate simulations are run. Because PDP attempts to explain acquisition facts directly, instead of identifying abstract mental constructs that are said to be only imperfectly reflected in performance, it is responsible to the data of acquisition. Exploration of the possible limits of PDP should be an interesting enterprise for SLA researchers to be involved in.

Received January 20, 1988.

Author's address for correspondence:
Richard W. Schmidt
Department of English as a Second Language
University of Hawai'i at Manoa
1890 East-West Road
Honolulu, HI 96822
USA
REFERENCES


