computation-saving strategies turn out to be nonrepresentational, others (e.g., cultural artifacts) are external representations. Hence, C&T's hypothesis is consistent with antirepresentationalism.

Using achieved representations reduces computational labor while also enabling a computational device to solve what would otherwise be an intractable mapping problem. Such is the crux of Clark & Thornton's C&T's representation-dependent "trading spaces" hypothesis. There are two facets of their gloss on the hypothesis about which I shall comment, one pertaining to the representational status of the sorts of "ploys, stratagems, and tricks" (sect. 4, para. 6) they identify, the other pertaining to whether the hypothesis depends on internal representations.

This commentary is motivated by my research into the nature of representation and its relation to computation. Since C&T's hypothesis is predicated on the interrelationship between these notions, let me begin by exploring a bit of the dialectic between computationalists and their opponents.

Although C&T fail to state precisely what they take to be an internal representation – apart from "a recoding scheme" (sect. 3, para. 16) – it is clear that they feel internal representations play a crucial role in biological computational processing. They are not alone: *computationalism* and *representationalism* underlie virtually every naturalistic attempt to explain how the mind/brain works. The former is the view that a computational framework is essential to explain the workings of "intelligent" systems. The latter is the view that computational systems require a sophisticated system (or "medium") of internal representations, without which they cannot compute (Fodor 1975, p. 27). So, regardless of whether one's preferred computational framework is symbolic or nonsymbolic, the status of internal representations in explanations of computational processing seems to be secure. Or so we have been conditioned to believe.

The received view of computationalism is not without its gainsayers. Although attacks against it come in a variety of guises, most (though not all) varieties of anticomputationalism are explicitly antirepresentational. Some of the more significant critiques include: (1) attacks against a computational framework being a plausible framework within which to explain cognition (Port & Van Gelder 1995; Van Gelder 1995); (2) arguments against the notion that (biological) computation presupposes a medium of internal representations (Stufflebeam 1997, Chs. 3–4); (3) attacks against the biological plausibility of cognitivism and symbolic computation (Dreyfus 1992; Searle 1990); (4) attacks against the efficacy of computational simulations as a basis for explanations of how the mind/brain works (Fodor 1995; Searle 1990); (5) arguments in defense of situated action (Agre 1993; 1995; Brooks 1991); and (6) antirepresentational arguments regarding the role and status of "distributed representations" in explanations of PDP (Stufflebeam 1995). If anyone is familiar with these attacks, it's Andy Clark (Clark 1996; Clark & Toribio 1994).

Does any computation-saving ploy, trick, or stratagem qualify as a representation? Some do – for example, "public language" and certain cultural artifacts (sect. 4, para. 9). Because to use public language is to use achieved representations to reduce the computational complexity regarding (at least) the transmission of knowledge, since public language is clearly representational, it exemplifies C&T's hypothesis. But "real-world actions" are supposed to do that as well (sect. 4, para. 15; sect. 5., para. 4). Here's the rub: C&T are careful *not* to call such computation-saving ploys "representations," though they *do* feel real-world actions are consistent with their hypothesis. As such, it is odd that C&T are insensitive to the antirepresentationalist arguments coming from proponents of situated action.

More important, C&T's hypothesis does *not* seem to depend on trading internal representation for computation, as is their claim. Instead, it seems to depend rather on trading *something* for computation, even if that something is an *external* representation (as is the case with cultural artifacts and public language utterances). It may still be the case that "the computationally weak will inherit the earth" (sect. 5, para. 5). But, one could argue, it is the *external* representations that make the computationally weak "representationally rich enough to afford it" (sect. 5, para. 5). And since real-world actions *also* get traded for computation, their hypothesis is far less representation-dependent than Clark & Thornton seem to realize.

Prospects for automatic recoding of inputs in connectionist learning

Nicolas Szilas and Thomas R. Shultz

Department of Psychology, McGill University, Montreal, Quebec, Canada H3A 1B1. **nicolas@lima.psych.mcgill.ca; shultz@psych.mcgill.ca; www.psych.mcgill.ca/labs/lnsc/html/lab.-home.html; www-leibniz.imag.fr/reseaux/szilas/szilas.html**

Abstract: Clark & Thornton present the well-established principle that recoding inputs can make learning easier. A useful goal would be to make such recoding automatic. We discuss some ways in which incrementality and transfer in connectionist networks could attain this goal.

Clark & Thornton's (C&T) key point, that recoding inputs can make learning easier, is well established and widely known. The merit of their approach is to tackle the problem instead of avoiding it, by showing negative results obtained without recording and providing a general overview of how manual recoding might work.

The next challenge would be to make such recoding automatic. Unfortunately, a mechanism for automatic recoding is not proposed by C&T. Ideas that could serve as the basis for automatic encoding can be found in studies of techniques for making learning easier, only some of which are mentioned by C&T.

C&T wisely cite Elman's (1993) work on incremental learning. There are a few studies showing that learning is sensitive to the sequencing of tasks (Cloete & Ludik 1993; Szilas & Ronco 1995; Tetewsky et al. 1995). However, if the sequence of tasks is not very carefully chosen, learning can be impaired.

A growing subfield in connectionism concerns the study of knowledge transfer. Some of these studies show that a common hidden layer can be shared by several tasks, in either simultaneous (Caruana 1993) or sequential (Baxter 1995) learning.

Transfer from one task to another can be useful only if the tasks are related in some important way. Otherwise, the two tasks may merely interfere with each other. If a representation is to be retained and reused without interference, it should perhaps be frozen. This is what is achieved by an algorithm like cascade-correlation (Fahlman & Lebiere 1990), which builds a hierarchical layered structure in which input weights to hidden units are no longer adjusted once the unit has been installed.

Simulations confirm that cascade-correlation networks are less susceptible to retroactive interference and make better models of human learning on sequential learning tasks than more conventional back-propagation networks that do not freeze hidden units (Tetewsky et al. 1993). In these simulations, transfer is achieved by learning connections from an old structure to a new one, whereas C&T seem to discard this possibility unless the earlier subnet is copied. With freezing of input-side weights, subnets are not used up, but simply used.

The constraint that "the dimensionality of the inputs is identical for both the original task and any later ones" can likewise be overcome: once again, use connections! A three dimensional input can be connected to a four dimensional input by a set of weighted links. Furthermore, in referring to Karmiloff-Smith's (1992) Representation Redescription, C&T seem to identify recoding with abstract redescription. Even if abstract redescription does exist, the foregoing examples show that the reuse of knowledge can occur without abstraction.

C&T stress the importance of using old representations to facilitate new learning, in effect, trading representation for computation. However, it is worth noting that in open-ended sequential learning there may be as many representations as there are learning episodes. Consequently, using achieved representations implies searching among them, and such search does require computation. Psychological studies of analogical reasoning show that even when people have relevant knowledge, they may not be able to use it without extensive hints (Gick & Holyoak 1980; 1983). Because such search is not trivial and is often unsuccessful, C&T's space trading maneuver is not without potential problems. The more representations that are available, the more extensive the search computation is likely to be.

Relational problems are not fully solved by a temporal sequence of statistical learning episodes

A. Vinter & P. Perruchet
L.E.A.D., C.N.R.S., University of Bourgogne, 21000 Dijon, France.
vinter@satie.u-bourgogne.fr

Abstract: Clark & Thornton's conception finds an echo in implicit learning research, which shows that subjects may perform adaptively in complex structured situations through the use of simple statistical learning mechanisms. However, the authors fail to draw a distinction between, on the one hand, subjects' representations which emerge from type-1 learning mechanisms, and, on the other, their knowledge of the genuine abstract "recoding function" which defines a type-2 problem.

1. Power of statistical learning mechanisms. Much of the interest of Clark & Thornton's (C&T's) target article lies in the fact that it offers a straightforward demonstration of the power of statistical learning mechanisms for solving problems which seem, prima facie, to be beyond the scope of such mechanisms. Empirical support for this conclusion can be found in the recent literature on implicit learning (Dienes & Berry, in press). In an often-cited study (Lewicki et al. 1988) for example, participants were asked to track as fast as possible a long and continuous series of targets appearing apparently at random locations. Unknown to participants, the series was composed of a systematic alternation of two unpredictable and three predictable trials. The discovery of this structure implies that subjects recode the continuous succession of trials into adjacent blocks of five successive trials. The underlying structure of the series remained completely opaque to participants, even after practice, yet performances were better for the predictable trials than for the unpredictable ones. Perruchet et al. (1990) demonstrated that the surprising adaptive performance of subjects in this situation was a direct consequence of a sensitivity to the frequency of occurrence of certain small chunks of two or three trials generated by the rules structuring the series. One could say that subjects solved a type-2 problem after its reduction to a set of type-1 problems.

The analogy between C&T's position and some aspects of the literature on implicit learning may be taken a step further. Perruchet and Gallego (in press) have proposed a theoretical account of implicit learning which shares striking similarities with C&T's claims about the nature and the function of type-1 learning. In this account, implicit learning is devoted to the formation of the "subjective units" shaping the perception of events and objects. Statistical learning mechanisms result in the chunking of information into discrete units, the nature and size of which are a function of the salience of surface features, as well as of the subject's background knowledge and general processing constraints and abilities (active memory and attention mainly). These subjective units emerge from the association of the primitive features that are processed conjointly in an attentional focus, and determine how the environment is attentionally perceived and processed after experience. With training, these units become increasingly independent of the sensory input and hence form internal representations. In line with C&T's position, this account construes the notion of representation as the endproduct of statistical learning mechanisms, making it possible to deal efficiently with problems involving what are a priori powerful computational abilities.

2. Limits of statistical learning mechanisms. Placing C&T's conception of learning within the context of implicit learning research reveals a major limitation of this conception, however. First note that C&T do not distinguish between the formation of achieved internal representations of the world, which permits behavioral adaptation to a given situation, and subjects' knowledge about the structural features of this situation. Let us illustrate this distinction. Each of us can state the direction of the source from which a sound comes. This ability stems from the detection and analysis of subtle differences in intensity or phase between the auditory streams processed by each ear. Consequently, location detection belongs to the class of relational, type-2 problems. The distinction we refer to is between the formation of achieved representations of sound space and the knowledge of the principle which permits these representations, namely, that detection is possible thanks to the relation between the information provided to each ear (Vinter & Perruchet 1994). Now, as should be clear from this example, it makes no sense to endow laymen with knowledge of this principle. The idea of knowledge makes sense here only from the observer's point of view not from the subject's.

In location detection, the coding of relational information is the direct product of hard-wired mechanisms. Our proposal is that the very same logic holds for the recoding provided by type-1 mechanisms of learning. The sensitivity to frequency statistics, and the representation resulting from this sensitivity, must be carefully distinguished from the subject's knowledge of the relational properties embedded in the task. Let us return to the Lewicki et al. situation. We noted that the better performance of subjects on the predictable trials, which apparently indicated that subjects were sensitive to the underlying structure of the series, relied on the sensitivity to the frequency of certain chunks forming the series. The crucial point is that this sensitivity to the surface frequency features gave the subjects no access at all to the underlying structure, for the very reason that the relevant frequencies, although a byproduct of the rules, do not make it possible to infer the rules. Indeed, the rules were concerned with the trajectory defined by two successive locations, whereas the resulting frequency effects captured by the participants were mostly concerned with perceptually salient units such as back and forth movements involving three successive locations. In this situation, it is clear that there is no justification for inferring relational knowledge from improved performance.

3. The need to introduce higher-level processes. We suggest that the solution provided by statistical learning mechanisms to type-2 problems is only a first step in the full course of human learning. The genuine knowledge of the relation embedded in type-2 problems involves processes that C&T fail to consider. In order to gain knowledge about the mechanisms involved in the detection of sound location for instance, scientists need to proceed by reasoning, hypothesis testing, and logical inference. The fact that they are themselves able to detect, as can everyone else, the location of a sound is of no help. In other words, knowledge of the "recoding function" can only be achieved by using processes fundamentally different from those involved in statistical learning. These high-level processes are needed to infer any abstract relation and to integrate it into a coherent view of the world or even to transfer it to another domain. The formation of abstract knowledge implies the use of processes which rely on the specific power of conscious thought. Overall, C&T's suggestion that there is no other type of learning to be had than type-1 learning, needs revision.