Language and Cognitive Processes

ISSN: 0169-0965 (Print) 1464-0732 (Online) Journal homepage: http://www.tandfonline.com/loi/plcp20

The transfer of implicit knowledge across domains

Zoltán Dienes , Gerry Altmann , Shi-Ji Gao & Alastair Goode

LANGUAGE AND COGNITIVE PROCESSES, 1995, *10* (3/4), 363–367

The Transfer of Implicit Knowledge Across Domains

Zoltán Dienes, Gerry Altmann, Shi-Ji Gao and Alastair Goode

Laboratory of Experimental Psychology, University of Sussex, Brighton, UK

INTRODUCTION

In a number of studies, Reber (e.g. 1967) and others have shown that asking subjects to memorise strings of letters generated by a finite stage grammar enables them later to classify new strings as obeying the rules or not. Reber (1989) regarded this knowledge as having two key properties. First, he argued that the knowledge was *implicit* because subjects found it very difficult to report what the rules of the grammar were. This claim has aroused a lot of controversy, but all seem to agree that the knowledge is at least difficult to articulate freely, and so is implicit in this minimal descriptive sense (for discussions, see Berry & Dienes, 1993; Dienes & Perner, in press). Second, Reber argued that the knowledge was abstract. This paper will focus on this second claim, and in particular on the claim that the knowledge (or a substantial part of it) is not tied to any specific perceptual features. First, experimental evidence relevant to the claim that the knowledge is abstract will be overviewed. Our extension of Elman's (1990) Simple Recurrent Network model to allow more abstract knowledge will be described and fitted to experimental data.

Reber (1969) provided the first evidence that incidentally acquired knowledge of artificial grammars is not strongly tied to specific perceptual features. Reber asked subjects to memorise strings of letters generated by a finite state grammar. The more strings subjects had previously studied, the easier they found it to memorise new strings generated by the grammar. This benefit remained intact even when the new strings were constructed from a different letter set, but the same grammar. Altmann, Dienes and Goode (in press) extended these findings by showing that subjects trained in one

Requests for reprints should be addressed to Zoltán Dienes, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

modality (e.g. sequences of tones) can classify in another modality (e.g. sequences of letters), even when the mapping between the domains was arbitrary and no accuracy feedback was given. These experiments are discussed next.

EXPERIMENTS WITH PEOPLE

Experiment 1

A finite state grammar was used to generate simple melodies or sequences of letters. The letters M, T, V, R and X were mapped onto the tones c, d, g, e and a.

The subjects were asked to memorise 20 grammatical stimuli; half the subjects memorised strings of letters and the other half memorised melodies. All subjects then classified new test letter strings and new test melodies without feedback. Finally, control groups were run who were trained on a different grammar, or who only received the test phase. Relative to the controls, classification was improved for the subjects tested in the same modality as learning (a 9% advantage). Relative to the controls, classification was also improved when a switch in modality had occurred (a 6% advantage), even though the subjects were not informed of the mapping.

Experiment 2

Experiment 2 used a similar design as Experiment 1, but with a different grammar and different stimuli. The subjects listened to sequences of nonsense syllables and were tested on sequences of unusual graphics characters. Again, the transfer group outperformed the control group during the test phase (58% *vs* 47%).

In this experiment, there were some strings that had no repeated elements—the subjects could classify these strings just as well as other strings, so the subjects must have induced a mapping. However, the subjects were at chance in specifying what the mapping was. In fact, in the test phase, the subjects often believed that they were just guessing.

In a subsequent experiment, in which the subjects were first exposed to sequences of graphics and were tested on written sequences of nonsense syllables, the transfer group outperformed the control group (64.5% *vs* 49.4%), but performed significantly worse than a no-transfer group who classified graphics sequences in the test phase (64.5% *vs* 70.7%).

SIMULATIONS WITH NETWORKS

Dienes, Altmann and Gao (1995) developed a computational model of the ability to transfer knowledge of an artificial grammar across domains for which the mapping is not *a priori* known. The process of forming a mapping between two disparate domains is a fundamental aspect of human cognition; indeed, Hofstadter (1985) regarded the problem of how analogies are formed as *the* fundamental problem in cognitive science. We attempted to model one part of that problem: How can an analogy be formed when the domains have already been carved up into the corresponding components and the task is to determine the mapping? This is the problem faced by the subjects in Experiments 1 and 2 above. Our aim is to provide an account of how that mapping could be established in the context of a plausible model of artificial grammar learning.

Training

We used a version of Elman's (1990) SRN which can transfer knowledge across different domains. The input layer is divided into two parts: D1 for coding information in the first domain, and D2 for coding information in the second domain. The first layer of hidden units then recodes both domains, and this recoding is used as an input to a standard SRN, with separate D1 and D2 output layers.

Let us say the first domain the model is trained on is syllable sequences. The first syllable in a sequence is coded by the D1 input units. This is successively recoded by the hidden layers in order to predict the second syllables of the sequence. Weights are adjusted by back-propagation. Then the second syllable is applied to the D1 input units, and so on. In a subsequent test phase, the network more successfully predicts successive syllables of grammatical rather than non-grammatical sequences when they are applied to D1, and this fact can be used to produce equivalent same-domain classification performance as people for equivalent training.

Testing in a New Domain

Now let us say the network is tested on a new domain, for example sequences of graphics. The "core weights" (the recurrent weights and the weights between the hidden layers) are frozen, and only the D2 input and output "mapping" weights are changed (i.e. the weights from the D2 input layer to the first layer, and the weights from the second hidden layer to the D2 output layer). The D2 mapping weights start at arbitrary random values. The first graphic is applied to the D2 input units and the network attempts to predict the second graphic. Back-propagation changes the mapping weights, and the network then attempts to predict the third graphic given the second, and so on. By the time the network has reached the end of the string, the mapping weights have changed, so the network can iterate around the string a number of times before moving on to the next test string. As in the same-domain case, the network will classify a string as grammatical if on the last iteration around the string it can predict successive graphics in the string well.

Because the core weights implicitly encode the structure of the new domain, the network just needs to learn the mapping to and from the abstract encoding formed by the SRN.

Results

The network does indeed learn the mappings, despite the noise introduced by adjusting the mapping weights when presented with non-grammatical strings. The advantage of training the core weights on the same grammar (the "transfer" data), as opposed to a different grammar (the "control" data) to that of the test stimuli, reaches up to 10% and is relatively stable over a range of parameter values. When the network is trained to produce equivalent same-domain performance as people, it can produce equivalent cross-domain transfer. If we take the same-domain performance to define the maximum amount of cross-domain transfer that could in principle be shown, then the model, like people, can perform at about 70% of the maximum possible. The correct mapping is gradually and partially induced over the first dozen test items. This partial induction of the mapping allows the network, like people, to classify correctly strings that are composed entirely of different graphics (i.e. with no graphic appearing more than once in the sequence) at that same level as other strings (which do contain repeated elements).

CONCLUSION

The purpose of this paper has been to explore the way people and connectionist models could apply knowledge of the structure of one domain to another. People are surprisingly successful in implicitly transferring their knowledge. The success of the SRN in modelling these data illustrates how abstract knowledge of an artificial grammar can be understood simply in terms of sensitivity to statistical structure. It also illustrates how knowledge of artificial grammars can be seen as implicit (see Dienes & Perner, in press, for a discussion of the relation between connectionist networks and implicit knowledge).

REFERENCES

Altmann, G.T.M., Dienes, Z., & Goode, A. (in press). On the modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Berry, D.C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues.* Hove: Lawrence Erlbaum Associates Ltd.

Dienes, Z., & Perner, J. (in press). Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit cognition.* Oxford: Oxford University Press.

Dienes, Z., Altmann, G., & Gao, S.-J. (1995). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. In L.S. Smith & P.J.B. Hancock (Eds), *Neural computation and psychology: Proceedings of the 3rd Neural Computation and Psychology Workshop (NCPW3)*, pp. 19–33. London: Springer Verlag.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Hofstadter, D.R. (1985). *Metamagical themas: Questing for the essence of mind and pattern.* Harmondsworth: Penguin.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.

Reber, A.S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115–119.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.