THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2003, *56A* (1), 117–123

Measuring learning using an untrained control group: Comment on R. Reber and Perruchet

Zoltán Dienes
University of Sussex, Brighton, UK
Gerry Altmann
University of York, York, UK

R. Reber and Perruchet (this issue) argue that use of control groups without training is unsound for establishing that learning has occurred. We show that inferring learning from a difference between a trained group and an untrained control in no way relies on their implausible additivity assumption, and that untrained control groups can be an invaluable aid to the researcher.

What is the nature of human learning? How do people learn? Are there different types of learning? These questions exercise many researchers in different disciplines. One increasingly active literature that has been motivated by such questions is the implicit learning literature, started in 1967 when A.S. Reber asked if people could learn artificial grammars in an implicit way. The artificial grammar learning task has remained a dominant paradigm for investigating implicit learning ever since. R. Reber and Perruchet (this issue) seek to show that basic ideas in the artificial grammar learning literature (and the learning literature generally) regarding how to answer even such a simple question as "Has learning occurred?" have been deeply misguided.

First let us note what we find useful about the paper. R. Reber and Perruchet provide implicit learning researchers with valuable information on particular biases that subjects may possess in classifying artificial grammar learning strings in the absence of learning. They further show that the particular set of biases that they found in their untrained subjects were likely to be the same as those in the untrained subjects in a number of previous studies conducted in a number of different countries. This is what we take to be the positive contribution of their paper. Learning (it seems to us) involves the replacement of one set of biases with another, and so it is potentially useful in investigating learning to know what biases subjects start with.

Requests for reprints should be sent to Zoltán Dienes, Experimental Psychology, Sussex University, Brighton, East Sussex, BN1 9QG, UK. Email: dienes@biols.susx.ac.uk

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980244000305

However, R. Reber and Perruchet believe that they have established something further. They state that the standard method of inferring learning from a difference in performance between a trained group and an untrained group relies on what they call the "additivity assumption". The additivity assumption is the assumption that learning consists of the strict addition of a new set of biases on the *enduring* initial set of biases. They point out that it is highly unlikely that any researcher actually explicitly makes this assumption. Presumably, when asked, most researchers (and that includes us) would assume that learning in general involves a *replacement* of one set of biases by another. Replacement rather than addition of biases is consistent with any of the empirical models of artificial grammar learning that exist in the literature (e.g., Berry & Dienes, 1993; Dienes, 1992; Dienes, Altmann, & Gao, 1999; Kinder & Shanks, 2001; Servan-Schreiber & Anderson, 1990) and with models of implicit learning generally (e.g. Cleeremans, 1993; Dienes & Fahey, 1995, 1998; Sun, 2002). That is, according to the models, the initial set of biases are at least attenuated, and even completely eliminated, when learning is complete. R. Reber and Perruchet's claim is that although researchers (which include themselves) a priori regard the additivity assumption as unlikely, researchers implicitly assume additivity whenever they conclude that learning occurred because a trained group classified better than an untrained group. R. Reber and Perruchet then show empirically that when trained on an artificial grammar, subjects do not persist with their old biases and just add some relevant additional ones; subjects instead display a completely different, if not opposed, set of biases from those of an untrained group. Consistent with theoretical expectation, the additivity assumption is wrong. R. Reber and Perruchet conclude that learning should not be established by comparing a trained group with an untrained group.

We agree with R. Reber and Perruchet that the additivity assumption is false in most circumstances. In fact, Dienes, Kurz, Bernhaupt, and Perner (1997) had earlier shown that untrained subjects do not respond randomly to strings but in very consistent ways and that these biases change with training. What we disagree with is that it follows that learning cannot be established by comparing a trained group with an untrained control. In fact, we are quite perplexed by this conclusion.

We note that the issue for R. Reber and Perruchet is not whether subjects have learnt the rules of the grammar (whatever "the" grammar is) or something else, like exemplars, or fragments. The issue is not about what is learnt, only whether learning has occurred at all. They make the implicit proviso that they are interested in whether subjects have learnt any "genuine features of the grammar", but this is "regardless of whether these features are construed in terms of rules, exemplar memory, or fragmentary information". What feature would count as a "genuine feature" of the grammar is left unspecified (just as the question of what counts as the grammar is). Maybe the point of the paper turns on this; but one cannot give a substantial answer to the question of when a feature is a "genuine" one or not in the absence of a theory of what type of learning one is interested in. On the other hand, the authors only wish to take us to the "limits inherent in any methodological recommendations", leaving details of "theoretically motivated tasks" to individual experimenters. If one's particular learning theory is not relevant, the content of the learning is not relevant to their recommendations. Any content will do. For current purposes, we take learning to be *acquiring some content positively correlated with the experimenter's grammar*, or *acquiring content more positively correlated with the experimenter's grammar than when the subject started learning*. By "positively correlated" we mean that the content when used to classify some set of items will produce classification responses correlated with the responses required by the experimenter's grammar (Dulany, Carlson, & Dewey, 1984). (Even this is a restricted definition; the subject may have thoroughly learnt a grammar that just happened to produce classification responses uncorrelated with those produced by the experimenter's grammar for the test set in question even though perfectly correlated for the training set.)

Do difference scores assume the additivity assumption?

If we assume that the additivity assumption is true, then a subject starts with a set of initial biases together with some variance in classification performance that is literally random (if it was not random it would be some bias, we presume). Learning consists of replacing some of the random variance with new biases, leaving the old ones intact, just as strong as they were. If the learning mechanism is reacting appropriately to the experimenter's grammar, classification performance will improve. That is, given that the additivity assumption is true, a trained group of subjects will have a higher classification performance than an untrained group when learning has occurred.

Now let us assume that the additivity assumption is false. A subjects starts with an initial set of biases. Learning consists of the replacement of these biases with a different set more correlated with the experimenter's grammar. The prediction is that a trained group will have a higher classification performance than an untrained group when learning has occurred.

Whether we assume the additivity assumption or not, we predict in general that the trained group will have a higher classification performance than the untrained group when learning has taken place. Taking a difference score in no way assumes the additivity assumption, contrary to the repeated (though never justified) claims of R. Reber and Perruchet.

It may be, as R. Reber and Perruchet suggest, that the initial biases are correlated with the grammar. Take the case where the correlation is negative. For example, grammatical strings may have less symmetry than non-grammatical strings in a way that untrained subjects are responsive to. Untrained subjects would then classify below chance. Trained subjects may learn that grammatical strings are asymmetric in this way. This knowledge may bring performance up to chance. Has learning occurred? Redington and Chater (1996) sensibly suggest not treating chance performance as evidence of learning. If we did in this last case, the chance performance of the trained subjects would lead us to say that they have not learned. In the sense of having acquired a set of biases that are positively correlated with the grammar, this is true: We should not say the subjects have learnt. On the other hand, the subjects have acquired a set of biases that are *more* positively correlated with the grammar than the pre-existing biases, and this could be taken to be a form of learning. (We would need to establish that subjects were not responding randomly but in consistent ways to different stimuli to establish that the trained subjects did indeed have a set of biases. Note that we do not need any additional control groups to establish this.)

If we allow learning of the relative symmetry of the strings to continue so that performance is above chance, then learning has occurred by either criterion. R. Reber and Perruchet might object that symmetry is not a "genuine feature" of a grammar (R. Reber and Perruchet describe it as a "non specific feature"). That just depends on the grammar and therefore one's theory of what should be learnt. Many context-free grammars produce inherent symmetries or anti-symmetries (see, e.g., Dienes & Perner, in press, for an example with artificial grammar learning); becoming sensitive to these symmetries is part of learning the grammar.

Let us assume that the subjects' biases are initially correlated with the grammar. Untrained subjects would then perform above chance. If this set is replaced with another set that as a whole is correlated with the grammar even better, then classification performance would improve. But maybe subjects, for some strange reason, stop using some biases that are correlated with the grammar and replace these with other biases correlated to just the same degree, a possible scenario raised by Reber and Perruchet. There is apparently no overall learning—no difference in classification performance between trained and untrained subjects—but change has occurred. It is not net learning, but it is learning of a sort when one considers not the set of biases as a whole but particular subsets of biases. Determining what has been learnt—and unlearnt—is the important question. This can be tracked without using any other control groups; indeed, R. Reber and Perruchet show just how this can be done. It is precisely the comparison of the trained group with the untrained control that allows learning (in terms of the whole set of biases or just a sub-set) to be monitored, as we see next.

The content of what is learnt

The important question in any investigation of learning is not simply whether learning has occurred, but what it is that subjects have learnt. A model or theory may predict a difference score between trained and untrained groups; finding such a difference therefore corroborates the model. (This assumes the model can model the existing biases of subjects. We suggest this is a useful characteristic to build into models, as was explicitly done, for example, by Dienes & Fahey, 1995, Sun, 2002.) The model or learning theory makes a falsifiable prediction (the trained group should perform better than the untrained group), and empirically validating the prediction corroborates the model. There is no need to assume the additivity assumption in drawing this conclusion.

Although the model has been corroborated, it has only been weakly corroborated. The subjects have learnt something, but have they really learnt what the model or theory predicts? It is important to look at how subjects classify particular items. Contrasting theories of what could have been learnt need to be set up, and item characteristics need to be manipulated to distinguish the theories. The artificial grammar learning literature has been replete with examples of just this sort of practice since McAndrews and Moscovitch (1985; conceptually replicating the delayed publication of Vokey & Brooks, 1992). Indeed, a prominent example of this sort is Perruchet and Pacteau (1990), who manipulated test items that had non-permissible bigrams or permissible bigrams in non-permissible locations. More recently, Tunney and Altmann (1999, 2001) manipulated training and test item characteristics in a series of artificial grammar learning studies to determine both the nature of the knowledge that can be transferred across domains (cf., Altmann, Dienes, & Goode, 1995) and the kinds of learning process that could underpin such transfer.

A useful way of investigating what item characteristics subjects are sensitive to was used by Johnstone and Shanks (1999). They performed a multiple regression over items to see which item characteristics predicted subject responses. This is a technique that we ourselves are using in current research, and R. Reber and Perruchet also illustrate how the technique can help indicate the content of subjects' knowledge. As an aside, we would initially like to comment on a difference in use between Johnstone and Shanks, on the one hand, and R. Reber and Perruchet on the other. Johnstone and Shanks performed a regression for each subject separately and then analysed the regression coefficients over subjects, allowing statistical generalization to other subjects. Significance testing cannot be performed for each subject separately, because of non-normality of the residuals, but an unbiased estimate is provided for the least squares coefficients for each subject (e.g., Gujarati, 1995), and significance testing can be performed over subjects. R. Reber and Perruchet performed the regression only over items, allowing in principle statistical generalization to other items but not to subjects. (What other items? Almost all grammatical items with the constraints used in item selection would have been used up in the study.) R. Reber and Perruchet comment that they think their method is preferable because of the assumption that the "incremental effect" of the independent variable "remains constant throughout", and this is problematic for a dichotomous variable or probability. It is of course equally problematic for a percentage (or a linear transform thereof), and so if they regard the constant incremental effect as a problem, it is a problem for them equally. If in doubt, one can use logistic regression for each subject, and then analyse over subjects. When the effects of each independent variable are relatively small, there is no need to use logistic regression.

R. Reber and Perruchet use multiple regression to show how training changes the biases used in classification. There is a statistically significant change in the regression coefficients between untrained subjects and trained subjects, and R. Reber and Perruchet use this to indicate that one set of biases have been replaced with another set by the process of training. Because this satisfies the definition of learning that R. Reber and Perruchet say most researchers hold, they have demonstrated how the use of an untrained control group is sufficient for demonstrating learning, and not only that learning has occurred but also, through the use of multiple regression using item characteristics as independent variables, what learning has occurred. Of course, the multiple regression does not tell one definitively what subjects have learned; there is no one methodological tool that can do that. Multiple regression can help distinguish competing theories that the experimenter has mustered, given sufficient statistical power, but no more. The true content might merely be correlated with some of the variables that the experimenter has isolated; the usefulness of the technique depends on how imaginative the researcher is in proposing potentially relevant variables.

Multiple regression is only one technique for investigating the content of what is learned. Converging evidence could be obtained by training subjects on items constructed in different ways. For example, Perruchet and Pacteau (1990) helped bolster their case that subjects predominantly learned bigrams by training subjects on bigrams and showing that such subjects were still good at classifying normal test items. Similarly, R. Reber and Perruchet showed that for their materials, subjects may not have learned content isomorphic with the experimenter's grammar, because when subjects were trained on materials constructed by a simpler constraints, the experimenter's grammar was still predictive in multiple regression. In order to take these findings further, one would need to come up with a theory of what could plausibly be learned from both training regimes, given the particular items used. There can be no theory-independent method of determining the content of learning, and therefore what training regimes are best contrasted.

Curiously, R. Reber and Perruchet suggest that as long as control subjects respond in non-random ways, the comparison of the group with another trained group is problematic. It follows that the only valid control group is not one that is motivated by some theory, but one in which subjects respond randomly. If this is the case, why not just compare trained subjects' responding to a chance baseline? To use control groups only in this way, it seems to us, would miss the point. The aim of a theory of learning is to specify how biases change, and there is no reason why the change has to be from a random state to a systematic state.

Conclusion

R. Reber and Perruchet claim that their additivity assumption is necessary for using untrained control groups to establish learning, while admitting that probably no researcher subscribes to this assumption. They go on to use untrained control groups to show the assumption is false and that learning proceeds by the replacement of one set of biases by another. Because this satisfies the definition of learning that R. Reber and Perruchet say most researchers hold, they have themselves demonstrated how the use of an untrained group is sufficient for demonstrating learning (even if they would not themselves subscribe to this view of their own work). Their experiments also illustrate how any further methodological requirement for establishing what learning has occurred is completely theory dependent. No general pre-theoretic rules can be laid down as to what control groups should or should not be used; it all depends on what the researcher wants to show. In addition, given insightful theories, analysing differences in the item characteristics that trained and untrained subjects are responsive to can be an invaluable aid to the researcher.

REFERENCES

Altmann, G., Dienes, Z., & Goode, A. (1995). On the modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 899–912.

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

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.

Dienes, Z. (1992). Connectionist and memory array models of artificial grammar learning. *Cognitive Science*, *16*, 41–79.

Dienes, Z., Altmann, G., & Gao, S.-J. (1999). Mapping across domains without feedback: A neural network model of implicit learning. *Cognitive Science, 23,* 53–82.

Dienes, Z., & Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 848–862.

Dienes, Z., & Fahey, R. (1998) The role of implicit memory in controlling a dynamic system. *Quarterly Journal of Experimental Psychology, 51A*, 593–614.

Dienes, Z., Kurz, A., Bernhaupt, R., & Perner, J. (1997). Application of implicit knowledge: Deterministic or probabilistic? *Psychologica Belgica*, *37*, 89–112.

Dienes, Z., & Perner, J. (in press). Unifying consciousness with explicit knowledge. In A. Cleeremans (Ed.), *The unity of consciousness: binding, integration, and dissociation.* Oxford: Oxford University Press.

Dulany, D.E., Carlson, R., & Dewey, G. (1984). A case of syntactical learning and judgement: How concrete and how abstract? *Journal of Experimental Psychology: General*, *113*, 541–555.

Gujarati, D.N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw–Hill.

Johnstone, T., & Shanks, D.R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 524–531.

Kinder, A., & Shanks, D.R. (2001). Amnesia and the declarative/procedural distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience*, *13*, 648–669.

McAndrews, M.P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. *Memory and Cognition*, *13*, 469–475.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119*, 264–275.

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

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A re-evaluation. *Journal of Experimental Psychology: General, 125*, 123–138.

Servan-Schreiber E., & Anderson, J.R. (1990) Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592–608.

Sun, R. (2002). *Duality of the mind: A bottom up approach to cognition*. London: Lawrence Erlbaum Associates Inc.

Tunney, R.J., & Altmann, G.T.M. (1999) The transfer effect in artificial grammar learning: Re-appraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1322–1333.

Tunney, R.J., & Altmann, G.T.M. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 614–639.

Vokey, J.R., & Brooks, L.R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 328–344.

Original manuscript received 4 February 2002