Running head: Individual Strategies in AGL

Individual Strategies in Artificial Grammar Learning

Ingmar Visser and Maartje E. J. Raijmakers

Department of Psychology, University of Amsterdam

Emmanuel M. Pothos

Department of Psychology, Swansea University

Please address correspondence to:

Ingmar Visser

Department of Psychology

University of Amsterdam

Roetersstraat 15

1018 WB Amsterdam

The Netherlands

i.visser@uva.nl

fax: +31-20-6390279

Abstract

Artificial Grammar Learning (AGL) has been used extensively to study theories of learning. We argue that compelling conclusions cannot be forthcoming without an analysis of individual strategies. We describe a new statistical method for doing so, based on the increasingly popular framework of latent variable models, which is especially suited to capture heterogeneity in participants' responses. In the current study, we apply the method of latent class regression models, in which the intercept and regression coefficients can have different values in different latent groups of participants; each latent group represents different reliance on the (potentially) available sources of knowledge in AGL, such as grammaticality and fragment overlap. The results indicate that grammaticality and fragment overlap can be understood as distinct aspects of learning performance, as evidenced by different groups of participants adopting predominantly one or the other strategy in a series of comparable datasets from AGL studies.

Individual Strategies in Artificial Grammar Learning

In an Artificial Grammar Learning (AGL) experiment, participants are first asked to study a set of training stimuli, usually sequences of letters. Subsequently, they are presented with some test stimuli and they have to decide which of these are compatible with the training ones. This simple learning task seems to embody the main elements of human learning, that is, the extraction of some knowledge from a set of stimuli, which can subsequently guide generalization performance to other, novel ones.

AGL has been employed in learning research for more than five decades (Miller, 1958; Reber, 1967). One of the reasons for the popularity of the AGL paradigm is that it allows a precise instantiation and comparison of different theories of learning. For example, AGL stimuli are typically created on the basis of a finite state language, so that only certain symbols can follow other symbols (Figure 1 depicts the finite state language that was used in the experiments that are reanalyzed in the current paper). Stimuli consistent and inconsistent with the finite state language are called grammatical (G) and ungrammatical (NG) respectively. Various AGL investigators have suggested that participants develop a representation of the finite state language employed, in terms of either a tacit network of rules (Reber, 1967) or explicit tests for deciding whether a stimulus is G or NG (Dulany, Carlson, & Dewey, 1984; cf. Dienes, 1992). Vokey and Brooks (1992) advocated an exemplar similarity view, according to which test items for which there is a highly similar training item are more likely to be endorsed as grammatical. Perruchet and Pacteau (1990) and Knowlton and Squire (1996) observed that participants sometimes base their grammaticality judgments on whether a test item would contain many parts (pairs or triplets of symbols) that have been frequently observed in training ('fragment

overlap'; see also Boucher & Dienes, 2003, Tunney & Altmann, 2001). Such views of

AGL are, in principle, distinct and map onto more general debates in cognitive

science (Pothos, 2005, 2007).

It is important to note that all the above AGL accounts consider regularity that

can only be evident if the training items are considered relative to each other. In other

words, chunk strength takes into account co-occurrence statistics across all training

items and knowledge of grammaticality constraints likewise arises from a

consideration of the commonalties between training items. However, as Jamieson and

Mewhort (2005) pointed out, AGL knowledge may well reflect local constraints (that

is, regularity within strings; cf. Tunney & Altmann, 2001). These authors have

proposed alternative forms of grammars that unconfound possible sources of

regularity. In this research, we wanted to carry out an analysis with AGL stimulus sets

that were as standard as possible, but with future work we hope to apply our approach

to more carefully controlled stimulus sets, as proposed by Jamieson and Mewhort.

-----------------------------INSERT FIGURE 1 ABOUT HERE-------------------------

The key issue regarding the utility of the AGL paradigm is whether it is

possible to specify methodologies and analytical techniques that allow researchers to

examine the particular source(s) of knowledge (e.g., in principle, grammaticality,

fragment overlap etc.) that drives performance. The main premise of this work is that

this is not possible, unless a measure of individual strategy is incorporated in AGL

analyses. Note that this issue is distinct from the problem of whether AGL knowledge

is implicit or explicit. The implicit/explicit debate in AGL has led to intense

controversy (e.g., Dienes & Perner, 1999; Shanks & St. John, 1994; Tunney &

Shanks, 2003), which is unlikely to be resolved without either radical reformulations of the implicit/explicit distinction (Dulany, 1997; 2003; cf. Cleeremans, 2005; Pothos, 2007) or cognitive neuroscience data (e.g., Eldridge et al., 2002; Knowlton, 1999). The latter is assuming, of course, that such data can in principle be brought to bear on problems in cognition (for discussion see e.g. Henson, 2006; Kim, 1992; Poldrack, 2006). We believe that the problem of determining the type of knowledge guiding performance in test is more tractable and we proceed to describe our method of addressing it.

Knowledge Sources in AGL

Empirically disambiguating the possible influence of different sources of knowledge in AGL has historically taken the following form: the test stimuli in an AGL task are created in a way that selection of different test stimuli indicates an influence from different sources of knowledge. For example, suppose that grammaticality and exemplar similarity are counterbalanced, so that the average similarity between the G test items and the training ones is the same as the average similarity of the NG test items and the training ones; likewise, the test items which are 'highly similar' to the training ones are equally likely to be G and NG, and the same would apply to the training items 'not similar' to the training ones. Therefore, for example, if participants are influenced primarily by grammaticality they should be selecting as G the G items, irrespective of their similarity to the training items. Several studies have adopted this approach (e.g., Higham, 1997; Knowlton & Squire, 1996; Vokey and Brooks, 1992).

More recently, regression techniques have been employed to examine the influence of different sources of knowledge on each grammaticality selection. For example, suppose we are examining exemplar similarity, fragment overlap, and

grammaticality. Each test item would be associated with three numbers: one corresponding to its exemplar similarity, another to its fragment overlap, and a third to whether it is G or NG. Then, a regression analysis can be constructed to predict the probability with which the item is endorsed or not, as a function of grammaticality, exemplar similarity, and fragment overlap. Johnstone and Shanks (1999) first applied this technique in AGL, using the repeated measures regression method of Lorch and Myers (1990). Conducting regression analyses in this way allows an item-based analysis, but in a way that individual participant variance is correctly taken into account (cf. Raaijmakers, Schrijnemakers, & Gremmen, 1999). Kinder and Assman (2000) employed a similar approach to conclude that similarity based measures have the strongest influence on performance. Pothos and Bailey (2000) extended these analyses by considering the variance accounted for in grammaticality selections, from each possible predictor, independent of the other predictors. In this way, for example, in some conditions it was found that grammaticality status could explain a unique portion of variance in participants' performance, over and above exemplar similarity or fragment overlap.

It is worth noting that there is a technique alternative to regression analysis for investigating the type of knowledge acquired in AGL (and whether there is a single type of knowledge or multiple types). This is based on receiver-operating characteristic (ROC) curves, as used, for example, by Kinder and Assmann (2000). While this approach is very promising, there are two potentially limiting factors in its applicability. First, a ROC analysis depends on particular assumptions about the way putative rules and similarity influences manifest themselves in AGL. Specifically, rules are assumed to reflect an all-or-none influence, while similarity is assumed to reflect a continuous influence. However plausible these assumptions, ideally an

analysis would proceed in a less theory-laden way, that is, without requiring a

particular commitment to the form of rules/ similarity influences in AGL. Second,

ROC curves require 'confidence' responses on some appropriate scale, rather than the

typical binary responses. Accordingly, a ROC analysis would be less applicable to

results from the standard AGL paradigm (and indeed Kinder and Assmann, 2000, had

to modify the standard AGL procedure for their analytical approach to be applicable).

Individual Differences in AGL Performance

We believe that a measure of individual strategy is needed to understand whether

grammaticality and fragment overlap are really distinct aspects of AGL performance.

Averaged participant data can reflect (seemingly) distinct influences of (for example)

grammaticality and fragment overlap in either of two ways. First, it could be the case

that for each participant both grammaticality and fragment overlap influence his/ her

performance. Second, it could be the case that some participants employ primarily

grammaticality in their selections, while others use fragment overlap (in such a case

we would expect to find categorical individual differences). Crucially, only the

second possibility allows us to infer (reasonably) unambiguously that in AGL there

are distinct influences on performance from both fragment overlap and

grammaticality. By contrast, in the first case it could be that grammaticality and

fragment overlap are simply epiphenomenal to another knowledge influence, which

happens to partially correlate with both. Such an argument requires some additional

assumptions to work (e.g., a noise signal), however, its validity has been

demonstrated several times, both in AGL (in relation to the implicit/explicit

distinction; e.g., Berry, Shanks, & Henson, 2008; Kinder & Shanks, 2001; Shanks &

Perruchet, 2002), and more generally (Juola & Plunkett, 1998; Plunkett & Bandelow, 2006).

The issue of individual strategies in AGL has been particularly, and surprisingly, under-researched[1]. Trivially, for example, one could compare the grammaticality and fragment overlap scores of each participant. However, such comparisons are meaningless since the effect size is small and an individual's performance could vary for all sorts of random reasons. An analysis is needed which would allow us to establish whether there is an overall trend for some participants to generalize on the basis of grammaticality vs. fragment overlap. Clearly, generalization strategies might depend on task demands and stimulus characteristics, and the analytical procedure should be sensitive to such variability; particularly so because groups of participants following different strategies, as we propose, cannot be observed directly; instead, the division in groups is a latent variable.

McAndrews and Moscovitch (1985) are the first researchers we are aware of who attempted to study individual strategies in AGL in a quantitative way. They examined grammaticality and exemplar similarity. They divided their sample into two groups, by performing a median split on the basis of average grammaticality accuracy. In this way, they observed that some participants' performance reflected a very strong influence from grammaticality knowledge and little influence of exemplar similarity, while for other participants the converse pattern held. This result is interesting, but establishing its robustness is problematic: it would be possible to carry out such a median split of participants' performance under any circumstances. Therefore, ideally, we would have some statistical measure of whether it is *appropriate* to distinguish participants in high/low achievers. The objective of the present work is exactly to provide such a statistical measure.

In other related work, Shanks, Johnstone, and Staggs (1997) observed individual strategies with a task involving biconditional grammars (first used by Mathews et al., 1989). In biconditional grammars each stimulus is composed of two groups of four letters. A letter in the first group predicts another one in the second group. In this way, stimuli in a biconditional grammar can be fully specified on the basis of a simple set of rules. Shanks et al. (1997) used two training procedures, after Mathews et al. (1989). One emphasized the surface properties of the stimuli, the other forced participants to process the structural regularities in the stimuli. Some participants in the latter group displayed a performance accuracy of nearly 100%, but the rest of the participants failed to learn. Such results are highly suggestive of two learning processes. However, biconditional grammars are very different from the standard finite state languages typically employed in AGL. Therefore, the results of Shanks et al. (1997) and Mathews et al. (1989) cannot be assumed to readily generalize to AGL.

The regression analysis technique of Johnstone and Shanks (1999) could, in principle, be modified to examine individual strategies, in two ways. First, interaction terms with the 'participants' variable could be examined, but no such results were reported by either Johnstone and Shanks (1999) or Pothos and Bailey (2000). Interaction terms in regression analyses involving 1000+ values per variable are somewhat cumbersome to investigate. Second, Johnstone and Shanks (1999) carried out regression analyses for each participant individually. In this way, by examining the standardized beta coefficient corresponding to different knowledge influences, one could decide whether the performance of a particular participant is, for example, primarily determined by grammaticality vs. fragment overlap. However, in practice this approach fails: as Johnstone and Shanks reported, the coefficients for the

(individual) regression analyses were hardly ever significant (cf. Tunney & Altmann, 2001).

Establishing Individual Strategies in AGL

We propose and illustrate a novel technique for examining individual strategies in AGL. This technique, latent class regression analysis (Huang & Bandeen-Roche, 2004), combines key aspects of two techniques that have been applied earlier: regression analysis and individual regression analysis. Latent class regression belongs to the family of latent structure models (Lazarsfeld & Henry, 1968). The main aim of these types of models is to explain correlations between responses to different items by introducing a latent variable. In our case, the items are AGL test stimuli; but the same method could be applied, for example, in questionnaire analysis. For example, standard analysis of personality scales involves the factor model: the correlation among several personality scale items is explained by introducing a latent factor, e.g. extraversion. In such a case, the latent variable is continuous. By contrast, in latent class models, the latent variable is nominal, indicating the existence of a number of different *types* of people rather than a dimension (such as extraversion) on which people vary continuously, as is the case in the factor model of personality scales. Finally, in latent class *regression* models, the assumption is that there are a number of types of people, each having a unique set of *regression coefficients* (and intercepts).

As is the case in normal regression analysis, in latent class regression models, performance is modeled as a function of predictors, such as grammaticality and fragment overlap. The assumption in standard regression analysis is that the same regression coefficients apply to each participant. However, in latent class regression analysis, participants belong to a particular group, and the regression coefficients are

the same only for those in the same group. Importantly, group membership is not a

manifest variable, but only assigned as a result of the statistical procedure. Thus,

latent class regression analysis is in-between standard regression analysis and the

individual participant regression analyses carried out by Johnstone and Shanks

(1999). In sum, latent class regression models are suitable to model heterogeneity in

responses, without going as far as modeling each individual participant separately. As

a consequence, latent class regression does not suffer from the problem that Johnstone

and Shanks ran into, which was that individual regression coefficients were hardly

ever significant.

Finally, although we are presently interested in AGL, the above discussion

hopefully illustrates that latent regression analysis has a very wide scope of

applicability. Bouwmeester, Sijtsma and Vermunt (2004) have applied it in cognitive

development, to model how reasoning strategies depend on age, school grades and

task demands. Schmitz et al. (2007) employed the method to differentiate between

groups of psychotic patients. Yamaguchi (2000) identified groups of people having

qualitatively different gender-role attitudes. Finally, Wang and colleagues have

reported a series of applications in biology (Wang & Puterman, 1998; Wang,

Puterman, & Cockburn, 1996). The interested reader is referred to McLachlan & Peel

(2000; specifically chapter 5) for a general introduction into such models.

Data Sets

We analyzed the data of Pothos and Bailey (2000) and Pothos, Chater, and Ziori

(2006). In both these studies data were collected (and compared) for three types of

stimuli. A factor that is very likely to affect style of responding in AGL is stimulus

format. Therefore these data sets afford ample potential for observing possible

individual strategies. Also, both studies employed the same finite state grammar and stimulus set (the one of Knowlton and Squire, 1996, Experiment 1; Figure 1), so that the results of these studies are highly comparable.

In the letters condition, participants memorized letter strings generated from the finite state grammar. In the 'shapes' stimulus set of Pothos and Bailey (2000) the letters of a standard finite state language were mapped to simple geometric shapes (e.g., a square, a diamond, etc.), arranged so that the shapes corresponding to later letters in a stimulus enclosed all earlier ones. Thus, each stimulus in the shapes condition looked like an embedded arrangement of shapes and would so give an impression of a single object, rather than an arbitrary collection of distinct elements. In the 'lines' stimulus set stimuli consisted of lines arranged at different angles relative to each other; each symbol in the finite state language corresponded to a different angle. The lines stimuli were constructed to confuse information about individual symbols or pairs of symbols, since the particular form of a symbol would depend on its context. In the 'routes' stimulus set the letters of the finite state grammar were mapped onto names of major cities, suggesting a travel itinerary. Finally, in the 'sequences' stimulus set geometric shapes were arranged next to each other, on a straight line. This stimulus set was meant to be most equivalent to standard AGL studies, yet still have a perceptual gestalt, intended to allow similarity judgments (for more information see Pothos & Bailey, 2000). Examples of the letters, shapes and routes stimuli are given in Figure 2 (from Pothos et al., 2006).

-----------------------------INSERT FIGURE 2 ABOUT HERE-------------------------

Using the repeated measures regression analysis technique of Lorch and Myers (1990), Pothos and Bailey (2000) reported significant effects of grammaticality, fragment overlap, and exemplar similarity for the shapes and the lines stimuli. For the sequences stimuli significant effects were observed only for exemplar similarity and grammaticality. (Grammaticality performance was computed as the proportion of G items in test correctly identified as G and NG ones correctly rejected as NG; fragment overlap performance was computed as the proportion of high fragment overlap items endorsed as G and low fragment overlap items rejected as NG.)

Pothos et al. (2006, Experiment 3; in the other two experiments only grammaticality was examined) also ran a condition with the shapes stimulus set. Additionally, they employed standard letter strings ('letters') and, finally, stimuli as sequences of cities, so that each sequence was meant to correspond to a 'route' of an airline company. With the routes stimuli it was hypothesized that general knowledge expectations of which routes were more or less plausible would impair processing of the structural properties of the stimuli. Pothos et al. examined grammaticality and fragment overlap. For the letters stimuli there was a significant effect of both (as indeed was found by Knowlton and Squire, 1996). For shapes and cities there were significant effects of fragment overlap; however, grammaticality only approached significance.

Pooling together, the The results of Pothos et al. (2006) and Pothos and Bailey (2000) together, provides five distinct AGL conditions: the letters, shapes, lines, routes, and sequences conditions. These conditions are rather closely matched, in that the same finite state grammar was employed and the training/test stimuli had the same abstract structure (that of Knowlton and Squire, 1996, Experiment 1). (The data from

the two shapes experiments, one in Pothos & Bailey, 2000, and the other in Pothos et al., 2006, were combined, since these two experiments were methodologically identical and latent regression analysis works better with larger sample sizes.) They differ in terms of stimulus format and so allow us the opportunity to examine the effect of stimulus format on participant strategy. Other researchers have examined the potential effect of stimulus format on AGL (e.g., Altmann, Dienes, & Goode, 1995; Chan, 1992; Conway & Christiansen, 2005, 2006). It is not relevant for our purposes to fully review this important research. Our aim presently was to seek a set of matched AGL conditions, varying only in terms of stimulus format. Hence, the conditions of Pothos et al. (2006) and Pothos and Bailey (2000) were considered most appropriate for inclusion in our analyses.

There were data from 20 participants with the letters stimuli, 20 with the routes stimuli, 36 with the shapes stimuli (20 from Pothos et al., 2006; 16 from Pothos & Bailey, 2000), 16 with the lines stimuli, and 16 with the sequences stimuli. We examined individual strategies primarily on the basis of two possible influences on performance, grammaticality and fragment overlap (in an additional set of analyses we included some further predictors). The reason is that Knowlton and Squire (1996) created this stimulus set with a view to specifically balance grammaticality and fragment overlap. In other words, the average chunk strength of the G items in test was the same as the average chunk strength of the NG ones. Accordingly, a participant responding perfectly on the basis of grammaticality would demonstrate chance chunk strength performance, and vice versa. Therefore, in principle, there could be exactly equal possible influences of grammaticality and fragment overlap, making it more likely to observe interesting differences in individual strategy. Note that a few corrections had to be made to the fragment overlap (chunk strength) values

reported by Knowlton and Squire (1996); the corrected values are shown in the

Appendix.

Finally, Pothos et al. (2006) and Pothos and Bailey (2000) closely observed

the standard procedure in AGL experiments. Before the start of the training phase,

participants were instructed that they were about to see a set of stimuli and that they

would only need to observe these (in the routes experiment, some additional

instructions were provided to allow participants to make sense of the stimuli). Then,

participants were presented with the training stimuli. Subsequently, they were told

that all the training stimuli had been created on the basis of a complex set of rules,

that participants would see novel stimuli which either complied or violated these

rules, and that they would have to decide which stimuli were legal and which illegal.

Participants did not receive corrective feedback, nor were specifically told whether

the numbers of legal and illegal stimuli conformed to a certain proportion (although

most participants probably did make this assumption).

<center>Latent Class Regression Models</center>

Latent class regression models were fitted to the above datasets. The main goal of the

analysis was to examine the influence of grammaticality and fragment overlap on

participants' performance (i.e. whether they judged a particular test item to be

grammatical or non-grammatical), in a way that would take into account possible

heterogeneity in individual strategies. Such putative heterogeneity was assessed

statistically, against a null hypothesis of no (robust) heterogeneity in participant

strategies.

The fitted model is a mixture of logistic regressions; more formally, assume

there are N_c latent classes, that is, participants can be divided into N_c groups, such that

participants within each group can be said to have adopted the same strategy of

responding. For our purposes, a strategy is a particular way to take into account

grammaticality and fragment overlap in responding. For example, one latent class

might reflect a predominant influence of grammaticality, while another relatively

equal influences of grammaticality and fragment overlap. The model does not specify

in advance the *form* of a strategy; rather, this is determined in a way analogous to how

regression coefficients are determined. More formally, the specified model had the

form:

$$\text{logit}(p(y_i)) = \mu_c + \beta_{gc}*G + \beta_{fc}*F, \qquad c=1 \ldots N_c \qquad , i=1 \ldots n, \qquad (1)$$

where $p(y_i)$ is the probability of judging a test item to be grammatical, and logit is the

logistic function, $\log(p/(1-p))$; μ_c is the intercept (specific to the particular latent

class; the same applies to all the other coefficients, which have a 'c' index); β_{gc} is the

regression coefficient for the *grammaticality* (G) of an item; β_{fc} is the regression

coefficient for the *fragment overlap* value (F) of an item (as computed by Knowlton

& Squire, 1996, and subsequently slightly corrected); N_c is the number of latent

classes and n is the number of participants. Note that if the number of classes equals

1, an ordinary logistic regression results; on the other hand, if the number of classes is

set equal to n, the number of participants, the individual regression analysis of

Johnstone and Shanks (1999) results (in the latter case, the classes are not latent

anymore because each participant has his/ her own class).

Before fitting the model, the variables *grammaticality* and *fragment overlap*

were centered, so that grammatical items were coded as 1, and ungrammatical items

were coded as -1, and *fragment overlap* was normalized to have zero mean and unit

variance. Centering was carried out to ensure that the intercept μ_c could be interpreted

as the average rate of endorsing items as grammatical in the latent class, as is

frequently done in normal regression analyses. Note that it is hard to provide rules-of-thumb about the number of cases needed to carry out latent class regression analysis as this depends largely on the magnitudes of the differences between classes, i.e. in terms of the values of the coefficients (and of course on the number of classes). However, in the applications mentioned earlier, sample sizes were at least 100 (and would be as high as 800 when different age groups were included in the studies as well).

Results

Models were fitted to the entire dataset, consisting of the 108 participants in the experiments of Pothos and Bailey (2000) and Pothos et al. (2006). Note that the latent class regression analysis is best carried out in the aggregate dataset, because it is a lot more robust with greater sample size. Information about each of the data sets separately, and about each of the conditions can readily be obtained by examining posterior estimates, as is done below. Each participant responded to the 32 test items (see Appendix A for details of the items). Models with an increasing number of latent classes, up to 4, were fitted. As latent variable models tend to have more than one local maximum of the log-likelihood function, multiple sets of starting values are used to ensure stability of the results. Models were fitted using the `flexmix` package (Leisch, 2004) for the R statistical programming environment (R Development Core Team, 2007; see also Grün & Fleisch, 2007, for an example of fitting mixtures of logistic regressions in R). In Table 1, the goodness-of-fit measures of the models with 1 through 4 classes are provided.

------------------------------INSERT TABLE 1 ABOUT HERE------------------------

The Akaike and Bayesian Information Criteria (AIC and BIC respectively), are commonly used in comparing non-nested competing models (Akaike, 1973; Schwarz, 1978), in this case between models with an increasing number of latent classes (see Lin, 1997, for details on the specific uses of AIC and BIC in latent class models). In the case of non-nested models, traditional tests for comparing models such as the log likelihood ratio test,test are not applicable. Both AIC and BIC provide a trade-off between goodness-of-fit, in this case the log likelihood, and the number of parameters in the models; note that for each added latent class, 4 extra parameters need to be estimated, i.e., the intercept and regression coefficients of that class and the proportion of participants that it contains. Lower values of each of these criteria denote better models, in which goodness-of-fit and parsimony are balanced. As can be seen from Table 1, the 3-class model fits the data optimally, according to both AIC and BIC criteria. The regression coefficients of this model are provided in Table 2.

------------------------------INSERT TABLE 2 ABOUT HERE-------------------------

It can be readily seen that the three classes are associated with different regression coefficients, reflecting differences in response strategies. In Class 1, there is a significant coefficient for grammaticality, but the fragment overlap coefficient is not significant. Accordingly, participants in Class 1 would be the ones who relied primarily on grammaticality for their responses; we will refer to Class 1 as the 'grammaticality' class. In Class 2 we have significant coefficients for both the intercept and fragment overlap; we will refer to this class as the 'fragment overlap' class. The significant intercept in the 'fragment overlap' class indicates that

participants in that class had an overall higher rate of endorsing test items as grammatical. The significant coefficient for 'fragment overlap' indicates that these participants are strongly influenced by this factor in judging test items. Finally, participants in Class 3 do not appear to be influenced by either grammaticality or fragment overlap; accordingly, Class 3 will be referred to as 'neither'. Such a class is consistent with the research of Johnstone and Shanks (1999) and Pothos and Bailey (2000), which showed that factors other than fragment overlap or grammaticality could influence performance. Class 3 participants might have picked up on idiosyncratic strategies, such as the length of the stimuli (although note that the Knowlton and Squire, 1996, stimuli were specifically balanced only for grammaticality and chunk strength; therefore, it is with respect to the study of these influences that an analysis of performance is most appropriate).

Table 2 also contains the class sizes. As can be seen, the 'grammaticality' and 'fragment overlap' classes each contain about 44% of the participants, and the 'neither' class contains the remaining 12%. Accordingly, the latent class analysis replicates the commonly reported finding that the influences of grammaticality and fragment overlap are roughly equivalent (e.g., Higham, 1995; Knowlton & Squire, 1996). There is a key difference: we show that *different* participants rely on grammaticality and different participants rely on fragment overlap. This is a very distinct conclusion from one whereby each participant would rely to an equivalent degree on grammaticality and fragment overlap. The latter conclusion cannot be used to corroborate a view of grammaticality and fragment overlap as distinct influences on performance. The former conclusion, the one that was possible with latent class modeling, does.

Clearly, of interest to researchers is not only the relative influence of grammaticality and fragment overlap on performance in the *aggregate* data (that is, data from the entire set of 108 participants), but also how this relative influence might vary with stimuli of different types. This can be examined by considering assignments of participants to latent classes. Based on the responses that a participant has provided, it is possible to compute the probability that s/he belongs to each of the latent classes. These probabilities are so-called posterior probabilities. Participants can then be assigned to the class corresponding to their highest posterior probability. This was done for the 3-class model that was described earlier. Table 3 contains a cross-tabulation of the number of participants in each of the three classes and the stimulus format condition that they were in.

-----------------------------INSERT TABLE 3 ABOUT HERE-------------------------

Differences in stimulus format did not appear to lead to systematic, interpretable variations in individual strategy. Possibly, the open-ended nature of the AGL task leads to different participants simply adopting different strategies in an idiosyncratic manner. Despite this somewhat disappointing conclusion, some broad observations are worth making. First, the letters and sequences conditions were the ones that were most equivalent to the standard AGL paradigm. For these conditions, one would anticipate the influence of grammaticality and fragment overlap to be roughly equal, as has been the intention of Knowlton and Squire (1996). Our analyses showed that this was indeed the case. Second, the routes stimuli were effectively sequences of words. Accordingly, there would be less scope for a similarity process to operate, as opposed to a symbolic process (which can be assumed to broadly

correspond to grammaticality, regardless of the exact interpretation of grammaticality). Correspondingly, Table 3 shows that most participants with the routes stimuli were assigned in the grammaticality class. Finally, the shapes and the lines stimuli were created to provide particularly compelling perceptual Gestalts and so encourage a mode of responding based on similarity; our analyses show that this was clearly the case for one of the shapes conditions, however, the evidence was ambivalent in the other shapes condition and the lines condition (i.e., as said, idiosyncrasies in response strategies probably overrode perceptual biases). Apart from idiosyncratic strategies, participants also differ with respect to an overall bias in endorsing test items; this is evidenced by a significant intercept term in the 'fragment overlap' class (and possibly by the intercept term in the 'neither' class which tends to significance).

With respect to the 'neither' participants, Table 2 indicates that they do respond with above chance grammaticality accuracy. This is not surprising. There are several performance factors that somewhat covary with grammaticality performance (e.g., the length of the stimuli; Pothos & Bailey, 2000). Indeed, AGL researchers have tried very hard to eliminate such factors, so that performance could be driven by only the putative influences of interest (e.g., Redington & Chater, 1996; Tunney & Altmann, 2001).

A methodological point is worth making about the cross-tabulation in Table 3: In dataset 2, participants provided two responses to each test item, whereas in dataset 1 participants provided only one response to each of the test items. As a consequence, it is expected that the classification of participants from dataset 2 into the latent classes could be made more reliable than for participants from dataset 1. Assigning participants to latent classes is not error-free; each participant has a certain probability

of belonging to each of the latent classes. This can be tested by computing the mean

of the maximum posterior probabilities for the participants from each of the datasets;

for dataset 1 this turns out to be 0.749, whereas in dataset 2 this turns out to be 0.858.

This confirms that indeed the posterior assignment of participants in dataset 2 is more

reliable than the assignment of participants from dataset 1. The straightforward

implication of this demonstration is that multiple responses per participant are

desirable when the object of the analyses is to reliably detect individual differences in

response strategies.

Other influences in AGL

In the analyses carried out so far, we have only considered grammaticality and

fragment overlap, as possible sources of information that participants could base their

judgments upon. Grammaticality and fragment overlap were the performance factors

balanced by Knowlton and Squire (1996) in their stimulus set that we used and are

two of the most widely used measures of AGL performance. We carried out a second

set of analyses including a number of other predictors as well, in particular anchor

chunk strength, length of the AGL strings, and edit distance. It is important to note

that the results from this second set of analyses would be less reliable: Knowlton and

Squire ensured that grammaticality and fragment overlap did not correlate with each

other in their stimulus set, but this is not necessarily the case for these additional

performance factors. The modeling approach is identical to the one described earlier,

the only difference being that now there are five predictors included in the analyses:

grammaticality, fragment overlap, anchor chunk strength, length, and edit distance.

Before presenting the analyses, we briefly discuss each of the new predictors in turn.

Several investigators (for an early examination see Reber & Allen, 1978) have pointed out that the beginning and end parts of a stimulus (anchor points) are particularly salient to participants. Accordingly, knowledge of anchor bigrams and/or trigrams may well have an influence on participants' classifications. Anchor chunk strength is computed in the same way as chunk strength, but taking into account the frequency of only beginning and final bigrams and trigrams. Additionally, we have found length to be a useful predictor variable of participants' performance in earlier research of ours (Pothos & Bailey, 2000). The motivation for including length is that it is a particularly salient feature of stimuli (especially graphical ones). Note that where stimuli are presented as embedded shapes, the length variable effectively corresponds to size. Finally, edit distance between two strings is the measure of similarity adopted by Vokey and Brooks (1992). The edit distance between two strings is computed as the as the number of insertions and deletions that are required to map one string to another, in which substitutions are treated as an insertion and a deletion. Vokey and Brooks advocated a similarity view of AGL based on edit distance and, also, edit distance corresponds well to some algorithmic views of similarity (e.g., Chater & Hahn, 1997).

Clearly, additional predictors of AGL performance could have been included. A problem with AGL is that there is nois no limit to the kind of statistics whichstatistics that can be computed for the items. However, even if a particular statistic correlates highly with performance, this does not necessarily enhance our insight of psychological process. Accordingly, we restricted ourselves to including possible predictors which have (a) been used in the past in AGL (since the purpose of our study is primarily to illustrate the latent class regression technique), and (b) can have a clear psychological interpretation.

As in the above presented analyses, we fitted latent class regression models with an increasing number of latent classes on participants' responses, with the only difference that now five predictors were included. The results broadly corroborate the earlier findings. A four-class model turned out to be the best model, although the three- and four-class model had very similar goodness-of-fit statistics (log-likelihood and BIC). The four-class model had one class in which fragment overlap was the only significant predictor, one class in which grammaticality was the best predictor, although only marginally significant (p=0.06), and two classes which had no significant predictors. The proportion of participants in the grammaticality class was 0.29, which is slightly less than the 0.44 found in the first analysis. The proportion of participants in the fragment overlap class was 0.44, which is identical to what was found in the first analysis.

In order to better compare the results from the first and second analysis, we also inspected the three-class solution in the second analysis. Statistically, this is justified because the goodness-of-fit statistics of the three-class solution were very close to that of the four-class one. The three-class solution had the same structure as in the first analysis, with a grammaticality class, a fragment overlap class and a 'neither' class. Relative to the first analysis, the only difference was that in the fragment overlap class, string length also had a significant influence on participants' endorsement of items. In the grammaticality class, both string length and anchor chunk strength were found to have marginally significant influences.

In none of the solutions did edit distance have any influenceany influence, significant nor marginally significant, on participants' responses. Therefore, we reran the second analysis without edit distance included. This enabled us to check the effects of adding/removing a predictor from the analysis which highly correlated with

the other predictors. In particular, the correlation between fragment overlap and edit distance was -0.68. As expected, results were similar, with the three-class solution being the best fitting model. Again, this model had a grammaticality class, a fragment overlap class, and a neither class as in our first analysis. The only difference relative to the first analysis, is that length was an additional significant predictor in the grammaticality class.

Overall, despite including three additional predictors of performance, the results of the second analysis were highly similar to the results of the first analysis. Moreover, even though length and anchor chunk strength emerged as somewhat important predictors of performance in the second analysis, some caution is required in interpreting this result. Both length and anchor chunk strength correlate with fragment overlap, and, albeit to a lesser extent, with grammaticality. This obviously makes it harder to separate the importance of each of these influences on participants' performance. Grammaticality and fragment overlap, on the other hand, do not correlate, and hence their influences can be analyzed independently. Avoiding colinearity problems in regression analyses such as the above is only possible if putative influences on AGL performance are balanced in an a priori way.

Discussion

In this paper we analyzed data from several AGL experimental conditions, with a view to determine whether there is heterogeneity in the strategies participants employ in responding to test items. Response strategies primarily concerned grammaticality and fragment overlap. By using latent class regression models, we were able to achieve conclusions well beyond previous similar approaches, such as those of Pothos and Bailey (2000) and Johnstone and Shanks (1999). The key reason is that standard

regression analyses assume a single set of coefficients for all participants. By contrast, latent class regression models introduce a (nominal) latent class variable, and compute a different set of regression coefficients for the participants in each class.

Why is this distinction important? Consider a situation where a regression analysis reveals significant effects of chunk strength and grammaticality. It is possible that participants' performance is *really* driven by a third factor, call it X, which happens to partly correlate with both grammaticality and similarity; if this is true, then it is X we should be studying, rather than grammaticality and chunk strength (cf. Kinder & Shanks, 2001; Plunkett & Bandelow, 2006). By contrast, if we identify some participants to be responding predominantly on the basis of grammaticality and some on chunk strength, then we are a lot more certain that grammaticality and chunk strength are correctly understood as *separate* influences on participants' performance on AGL. This is the conclusion from the present analyses.

More specifically, the starting point in our analyses was that a single class model would be the best model, that is, there would be no variation in response strategies between individual participants. The analyses showed otherwise by identifying three classes, one showing a predominant influence of grammaticality on responses, one showing a predominant influence of fragment overlap, and one where responses could not be readily assigned to either grammaticality or fragment overlap. This result also provides further confirmation for the long AGL research tradition which assumes grammaticality to be separate from fragment overlap (Higham, 1997; Knowlton & Squire, 1996; Pothos & Bailey, 2000).

Furthermore, an examination of posterior probabilities enabled us to assign participants in the different experimental conditions to each of the three latent classes. Such an analysis could, in principle, reveal systematic influences of stimulus format

on response strategies. However, beyond some general conclusions, this approach did

not reveal any further structure in our data. This indicates that participants appear to

use individual strategies based on known or unknown stimulus characteristics, rather

than been uniformly biased by particular characteristics of the stimuli (such as

stimulus format).

Similar remarks could apply when different grammars are involved rather than

just different stimulus formats, and when different training sets of stimuli are used.

For example, Meulemans and Van der Linden (1997) showed that with few training

examples, participants based their responses predominantly on chunk strength (or

fragment overlap). On the other hand, when given more training examples,

participants based their responses relatively more on grammaticality. Latent class

regression analysis could confirm such differences in representational format between

the different conditions, while at the same time assess whether there are individual

differences with respect to the strategies that participants used; that is, it could be the

case that even in the many-training-examples-condition, some participants still

predominantly used chunk strength and vice versa in the other condition. With future

work we hope to analyze more datasets with the promising latent regression method.

Another issue concerns the development of classification strategies. In order to

study the development of classification strategies, one would need to have participants

carry out several judgments on the same test stimuli. However, a potential

complication with such a procedure is that successive judgments would not reflect

application of knowledge acquired from training, but rather memory of the previous

judgment. In an early investigation, Reber and Allen (1978) had test stimuli presented

twice in order to examine consistency of responses; they measured the frequency of

CC (correct—correct responses for two presentations of the same test item), EE

(erroneous—erroneous response), CE, and EC. One of their findings was that the combined frequency of CC, EE was much higher than that of CE, EC, showing that participants were consistently correct or wrong in the test part. Later investigations (including our own) replicated this finding. With respect to how the current statistical approach could deal with this issue, one could have participants make several responses for the same test stimuli and then analyze the response patterns for each set of responses separately. We could then find that the characterization of a participant relative to his/her first set of responses is different relative to his/her second set of responses. While on paper such an approach appears plausible, we believe one would require a much greater sample size before it is viable. Note that in categorization a corresponding analysis with latent regression models has already been carried out, by Raijmakers and Visser (submitted), who analyzed the data of Johansen and Palmeri (2002). Johansson and Palmeri created a task which involved successive classifications of the same stimuli, so as to study the evolution of category representation. Such data are suitable for the so-called latent Markov analysis (which is an extension of the latent class analysis in repeated measurements, cf. Visser, Schmittmann, & Raijmakers, 2007). Whether the approach of Raijmakers and Visser (submitted) might be a suitable way to address the problem of repeated measurements in AGL as well is an interesting topic for future research.

In sum, we hope to have illustrated the utility of our approach. A clear prescription for further research is greater sample sizes. The latent modeling approach to regression analysis, for all its additional explanatory power, is ideally applied to large sample sizes. We hope that there will be opportunities to apply our method to more extensive datasets in the future, and so corroborate the current conclusions.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood

principle. In Petrov, B. N. and Csaki, F., (Eds.), <u>Second international</u>

<u>symposium on information theory</u>, 267–281. Budapest: Academiai Kiado.

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality Independence of

Implicitly Learned Grammatical Knowledge. <u>Journal of Experimental</u>

<u>Psychology: Learning, Memory and Cognition</u>, <u>21</u>, 899-912.

Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008). A single-system account of

the relationship between priming, recognition, and fluency. <u>Journal of</u>

<u>Experimental Psychology: Learning, Memory, and Cognition</u>, <u>34</u>, 97-111.

Boucher, L. & Dienes, Z. (2003). Two ways of learning associations. <u>Cognitive</u>

<u>Science</u>, <u>27</u>, 807-842.

Bouwmeester, S., Sijtsma, K., & Vermunt, J.K. (2004). Latent Class Regression

Analysis to Describe Cognitive Developmental Phenomena: An Application to

Transitive Reasoning. <u>European Journal of Developmental Psychology, 1</u>, 67-

86.

Chan, C. (1992). <u>Implicit Cognitive Processes: Theoretical Issues and Applications in</u>

<u>Computer Systems Design</u>. Unpublished DPhil thesis, University of Oxford.

Chater, N. & Hahn, U. (1997). Representational distortion, similarity and the

Universal Law of generalization. In <u>Proceedings of the Similarity and</u>

<u>Categorization Workshop 97</u>, 31-36, University of Edinburgh.

Cleeremans, A. (2005). Computational correlates of consciousness. <u>Progress in Brain</u>

<u>Research</u>, <u>150</u>, 81-98.

Conway, C. & Christiansen, M. H. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. Journal of Experimental Psychology: Learning, Memory & Cognition, 31, 24-39.

Conway, C. M. & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus specific representations. Psychological Science, 17, 905-912.

Dienes, Z. (1992). Connectionist and Memory-Array Models of Artificial Grammar Learning. Cognitive Science, 16, 41-79.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. Behavioral and Brain Sciences, 22, 735-808.

Dulany, D. E. (1997). Consciousness in the Explicit (Deliberative) and Implicit (Evocative). In J.D. Cohen, & J. W. Schooler (Eds.), Scientific approaches to consciousness. (pp. 179-211). Mahwah, NJ: Lawrence Erlbaum Associates.

Dulany, D. E. (2003). Strategies for putting consciousness in its place. Journal of Consciousness Studies, 10, 33-43.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? Journal of Experimental Psychology: General, 113, 541-555.

Eldridge, L. L., Masterman, D., & Knowlton, B. J. (2002). Intact implicit habit learning in Alzheimer's disease. Behavioral Neuroscience, 116, 722-726.

Gebauer, G. F. and Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(1): 34–54.

Grün, B. and Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in R. Computational Statistics and Data Analysis, 51(11): 5247–5252.

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. TRENDS in Cognitive Sciences, 10(2):64–69.

Higham, P. A. (1997). Dissociations of grammaticality and specific similarity effects in artificial grammar learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 1029-1045.

Huang, G-H., & Bandeen-Roche, K. (2004). Building an Identifiable Latent Class Model with Covariate Effects on Underlying and Measured Variables. Psychometrika, 69(1), 5-32.

Jamieson, R. K. & Mewhort, D. J. K. (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: an analysis using information theory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 9-23.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? Cognitive psychology, 45, 482-553.

Johnstone, T. & Shanks, D. (1999). Two Mechanisms in Implicit Grammar Learning? Comment on Meulemans and Van der Linden (1997). Journal of Experimental Psychology: Learning, Memory, & Cognition, 25, 524-531.

Juola, P., & Plunkett, K. (1998). Why double dissociations don't mean much. In M.A. Gernsbacher & S.J. Derry (Eds.), Proceedings of the Twentieth Annual Conference of the Cognitive Science Society (pp. 561-566). Mahwah, NJ: Lawrence Erlbaum Associates.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. Philosophy and
    Phenomenological Research, 52(1):1–26.

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: No evidence for the
    acquisition of rules. Memory & Cognition, 28, 1321-1332.

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

Knowlton, B. J. (1999). What can neuropsychology tell us about category learning?
    Trends in Cognitive Sciences, 3, 123-124.

Knowlton, B. J., & Squire, L. R. (1996). Artificial Grammar Learning Depends on
    Implicit Acquisition of Both Abstract and Exemplar-Specific Information.
    Journal of Experimental Psychology: Learning, Memory and Cognition, 22,
    169-181.

Lazarsfeld, P. F., & Henry, N. W. (1968). Latent structure analysis. Boston: Houghton
    Mifflin.

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent
    class regression in R. Journal of Statistical Software, 11(8).
    http://www.jstatsoft.org/v11/i08/

Lorch Jr., R. F. & Myers, J. L. (1990). Regression Analyses of Repeated Measures
    Data in Cognitive Research. Journal of Experimental Psychology: Learning,
    Memory, and Cognition, 16, 149-157.

Lotz, A., & Kinder, A. (2006). Transfer in artificial grammar learning: The role of
    repetition information. Journal of Experimental Psychology: Learning,
    Memory & Cognition, 32, 707-715.

McLachlan, G.J., & Peel, D. (2000). Finite Mixture Models. Wiley.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., &
    Druhan, B. (1989). Role of Implicit and Explicit Processes in Learning From
    Examples: A Synergistic Effect. Journal of Experimental Psychology:
    Learning, Memory and Cognition, 15, 1083-1100.

Meulemans, T., Van der Linden, M. (1997). Associative Chunk Strength in Artificial
    Grammar Learning. Journal of Experimental Psychology: Learning, Memory,
    and Cognition, 23, 1007-1028.

Meulemans, T., van der Linden, M., and Perruchet, P. (1998). Implicit sequence
    learning in children. Journal of Experimental Child Psychology, 199–221.

Lin, T. H. and Dayton, C. M. (1997). Model selection information criteria for non-
    nested latent class models. Journal of Educational and Behavioral Statistics,
    22(3):249–264.

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

Miller, G. A. (1958). Free Recall of Redundant Strings of Letters. Journal of
    Experimental Psychology, 56, 485-491.

Perruchet, P., & Pacteau, C. (1990). Synthetic Grammar Learning: Implicit Rule
    Abstraction or Explicit Fragmentary Knowledge? Journal of Experimental
    Psychology: General, 119, 264-275.

Plunkett, K. & Bandelow, S. (2006). Stochastic approaches to understanding
    dissociations in inflectional morphology. Brain and Language, 98, 194-209.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?
    TRENDS in Cognitive Sciences, 10(2):59–63.

Pothos, E. M. (2005). The rules versus similarity distinction. Behavioral & Brain Sciences, 28, 1-49.

Pothos, E. M. (2007). Theories of Artificial Grammar Learning. Psychological Bulletin, 133, 227-244.

Pothos, E. M., & Bailey, T. M. (2000). The Importance of Similarity in Artificial Grammar Learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 847-862.

Pothos, E. M., Chater, N., & Ziori, E. (2006). Does stimulus appearance affect learning? The American Journal of Psychology, 119, 277-301.

R Development Core Team (2007). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with the "the language-as-fixed-effect fallacy": common misconceptions and alternative solutions. Journal of Memory and Language, 41, 416-426.

Raijmakers, M.E.J. & Visser, I. (submitted). Statistical models of intra- and inter-individual differences and errors in category learning. Manuscript submitted for publication.

Redington, F. M., & Chater, N. (1996). Transfer in Artificial Grammar Learning: Methodological Issues and Theoretical Implications. Journal of Experimental Psychology: General, 125, 123-138.

Reber, A. S. (1967). Implicit Learning of Artificial Grammars. Journal of Verbal Learning and Verbal Behavior, 6, 855-863.

Reber, A. R., Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functional interpretation, Cognition, 6, 189-221.

Schmitz, N., Malla, A., Norman, R., Archie, S., & Zipursky, R. (2007). Inconsistency

in the relationship between duration of untreated psychosis (DUP) and

negative symptoms: Sorting out the problem of heterogeneity. Schizophrenia

Research, 93, 152-159.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6,

461–464.

Shanks, D. R., St. John, M. F. (1994). Characteristics of dissociable human learning

systems. Behavioral and Brain Sciences, 17, 367-447.

Shanks, D. R., & Perruchet, P. (2002). Dissociation between priming and recognition

in the expression of sequential knowledge. Psychonomic Bulletin & Review,

9, 362-367.

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial

grammar learning. Quarterly Journal of Experimental Psychology, 50A, 216-

252.

Tunney, R. J. & Altmann, G. T. M. (2001). Two models of transfer in artificial

grammar learning. Journal of Experimental Psychology: Learning, Memory,

and Cognition, 27, 614-639.

Tunney, R. J. & Shanks, D. R. (2003). Subjective measures of awareness and implicit

cognition. Memory & Cognition, 31, 1060-1071.

Visser, I., Schmittmann, V. D., and Raijmakers, M. E. J. (2007). Markov process

models for discrimination learning. In van Montfort, K., Oud, H., and Satorra,

A., editors, Longitudinal models in the behavioral and related sciences,

chapter 14, pages 337–365. Lawrence Erlbaum Associates, Mahwah (NJ).

Vokey, J. R., & Brooks, L. R. (1992). Salience of Item Knowledge in Learning

    Artificial Grammar. Journal of Experimental Psychology: Learning, Memory

    & Cognition, 20, 328-344.

Wang P., Puterman M. L. (1998). Mixed Logistic Regression Models. Journal of

    Agricultural, Biological, and Environmental Statistics, 3(2), 175- 200.

Wang P., Puterman M. L., Cockburn IM, Le ND (1996). Mixed Poisson regression

    models with covariate dependent rates. Biometrics, 52, 381-400.

Yamaguchi, K. (2000). Multinomial Logit Latent-Class Regression Models: An

    Analysis of the Predictors of Gender-Role Attitudes among Japanese Women.

    American Journal of Sociology, 105(6), 1702-1740.

Appendix

The stimulus set of Knowlton and Squire (1996, Experiment 1), used by Pothos and

Bailey (2000) and Pothos et al. (2006, Experiment 3). The letters correspond to the

ones employed by Knowlton and Squire. vtvjjSec indicates the second instance of

vtvjj (Knowlton and Squire accidentally repeated in their test set one item). Finally,

higher fragment overlap values indicate higher similarity to the training items.

item	Grammaticality	Fragment overlap
vjtvt	G	6
vjtxvx	G	6.777778
vtv	G	5
vtvj	G	5.6
vtvjj	G	5.142857
vtvjjSec	G	5.142857
vx	G	12
vxj	G	9.333333
xvjtvt	G	6.666667
xvjtvx	G	7.444444
xvtv	G	6.8
xvtvj	G	6.714286
xvtvjj	G	6.111111
xxvtv	G	7.857143
xxvtvj	G	7.555556
xxvxj	G	10.28571
jxvt	NG	5
tvj	NG	6.666667
vjjxvt	NG	4.888889
vjtv	NG	7
vxjjx	NG	5.857143
vxjtj	NG	4.857143

vxvj	NG	8.2
xjj	NG	7
xvxt	NG	7
xvxv	NG	10
xvxvj	NG	9.142857
xxjj	NG	6.8
xxtx	NG	2.8
xxv	NG	12
xxvjjj	NG	7.666667
xxvvjj	NG	6.111111

Author note

Footnote

[1]Note that there is a related but different discussion in the implicit learning literature on individual differences in implicit learning abilities. In this case, the question of interest is whether participants reliably differ in their ability to judge novel items as grammatical or not, and most importantly whether this ability covaries with other cognitive abilities, e.g. IQ (cf. Gebauer & Mackintosh, 2007). In such studies individual differences are considered to be quantitative, whereas in the current paper it is examined whether there are qualitative (i.e., categorical) individual differences.

Table 1. Goodness of fit for latent class regression models with 1 through 4 classes.

classes	loglik*	AIC	BIC
1	-3041.3	6088.6	6096.6
2	-3003.9	6021.8	6040.6
3	-2993.7	6009.5	6038.9
4	-2991.1	6012.1	6052.4

*loglik refers to the log likelihood of the models.

Table 2. Regression coefficients for the best, 3-class, model.

class	size	intercept	gram β_g (se)	frag β_f (se)	correct
1	0.44	-0.02 (0.11)	0.25 (0.11)	0.10 (0.12)	0.572*
		p=0.84	p<0.05*	p=0.39	
2	0.44	0.32 (0.12)	0.13 (0.12)	0.51 (0.13)	0.524*
		p<0.01*	p=0.28	p<0.001*	
3	0.12	1.35 (0.79)	0.39 (0.79)	-0.09 (0.75)	0.560*
		p=0.09	p=0.61	p=0.91	

Note: gram denotes the 'grammaticality' coefficient; frag denotes the 'fragment overlap' coefficients; *'s indicate the significant coefficients. Standard errors (se) of parameters are shown in parentheses. The final column provides the proportion of correctly judged items for participants in each of the classes (all three proportions are significantly above chance level, as assessed with single-sample t-tests).

Table 3. Distribution of participants over the latent classes and stimulus format
conditions.

Dataset	Condition	Neither	Fragment	Grammaticality	Total
1	Letter	2	8	10	20
	Routes	3	5	12	20
	Shapes	9	9	2	20
2	Shapes	1	7	8	16
	Lines	0	10	6	16
	Sequences	1	6	9	16
	Total	16	45	47	108

Note: Dataset 1 refers to the conditions of Pothos et al. (2006), dataset 2 to the ones of

Pothos and Bailey (2000).

Figure Captions

Figure 1. An example of a finite state language (from Knowlton & Squire, 1996), where the symbols corresponding to the different transitions are letters and therefore the resulting sequences letter strings. The circles are the states of the language. Every time a legal transition is made between states, the letter corresponding to this transition is added, until a transition is made to one of the OUT states. For example, while string XXVT is G, string XT is not.

Figure 2. Examples of the types of stimuli used in the letters, shapes, and routes conditions respectively (from Pothos et al., 2006).
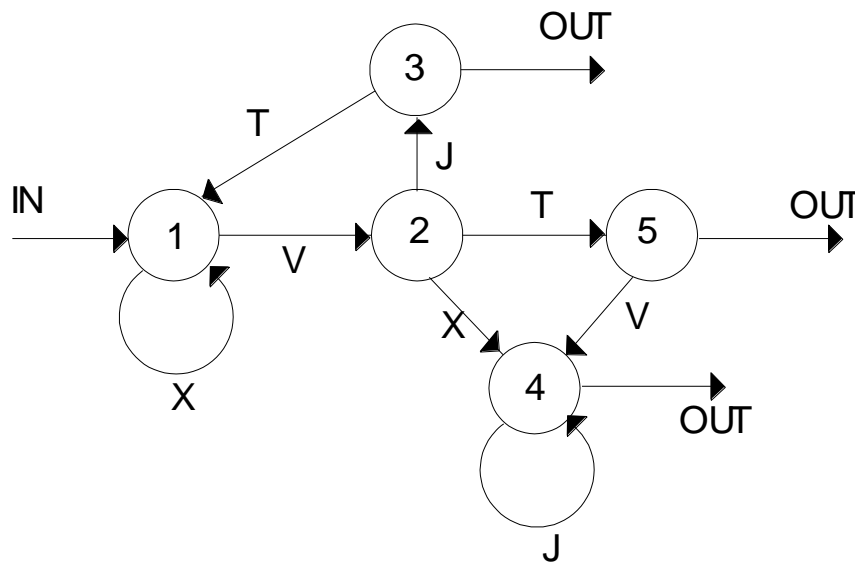
Figure 1

Figure 2

MSSSV

Athens --> London --> London --> London --> Berlin

MSSSV

Athens --> London --> London --> London --> Berlin