Psychological Research (2017) 81:204–218
DOI 10.1007/s00426-015-0715-4

ORIGINAL ARTICLE

# Implicit learning is order dependent

**Randall K. Jamieson[1] · John R. Vokey[2] · D. J. K. Mewhort[3]**

Received: 21 April 2015 / Accepted: 5 October 2015 / Published online: 20 October 2015


**Abstract** We report two experiments using the artificial-grammar task that demonstrate order dependence in implicit learning. Studying grammatical training strings in different orders did not affect participants' discrimination of grammatical from ungrammatical test strings, but it did affect their judgments about specific test strings. Current accounts of learning in the artificial-grammar task focus on category-level discrimination and largely ignore item-level discrimination. Hence, the results highlight the importance of moving theory from a category- to an item-level of analysis and point to a new way to evaluate and to refine accounts of implicit learning.

## Implicit learning is order dependent

People become sensitive to the regularities in a stimulus domain, often without deliberate effort or explicit awareness. In response, several theorists have argued for a specialized learning system, one that abstracts contingencies automatically and that guides behaviour adaptively.

Reber (1967) was the first to discuss the idea under the name implicit learning. In his now classic experiment, participants studied strings of letters constructed according to the rules of an artificial grammar (see Fig. 1).

Next, participants were informed that the studied strings had been constructed using rules and then were asked to discriminate novel grammatical (i.e., rule based) from ungrammatical test strings. Although participants could judge the grammatical status of the test strings, they could not articulate the rules of the grammar. Reber (1967) took the ability to discriminate the two kinds of test strings as evidence that participants had abstracted the grammar. Because they could not describe the rules, he surmised that the participants' knowledge of the grammar must be implicit.

Other theorists have argued that, instead of abstracting the grammar, the participants store the training exemplars and, at test, endorse strings that remind them of the studied exemplars. Because no implicit knowledge of the grammar is assumed, the discrepancy between peoples' performance and their inability to describe the rules is inconsequential. From this perspective, implicit-learning tasks define an interesting class of phenomena but do not point to a unique class of processing mechanisms (e.g., Brooks, 1978; Dulany, Carlson, & Dewey, 1984; Jamieson & Mewhort, 2009, 2011; Pothos & Bailey, 2000; Vokey & Brooks, 1992).

Since Reber's (1967) initial demonstration, an empirical database has developed that delineates the constraints, limits, and regularities of performance in the artificial-grammar test. For example, people learn adjacent letter dependencies more easily than nonadjacent ones (Kinder, 2010; Poletiek & Lai, 2012). People infer grammaticality based on fragment similarity (Johnstone & Shanks, 2001; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). People infer grammaticality based on whole-string similarity (Jamieson & Mewhort, 2009, 2011; Vokey & Brooks, 1992). People infer grammaticality based on serial redundancy (Brooks & Vokey, 1991; Jamieson et al., 2015; Lotz & Kinder, 2006; Vokey & Higham, 2005). Discrimination improves with the redundancy in the generative

✉ Randall K. Jamieson
randy.jamieson@umanitoba.ca

[1] Department of Psychology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

[2] University of Lethbridge, Lethbridge, Canada

[3] Queen's University at Kingston, Kingston, Canada

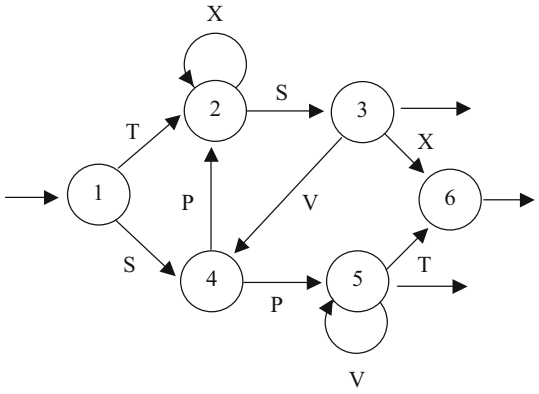


**Fig. 1** The artificial grammar that was used to construct materials in Experiment 1. To generate a grammatical stimulus, one enters the grammar at the leftmost node marked *1* and follows successive paths, picking up associated letters until reaching an exit (i.e., the *arrows* that depart from nodes *3*, *5*, and *6*). For example, the route from nodes *1* to *4*, *4* to *5*, *5* to *5*, and *5* to *6* produces the string SPVT

grammar (Jamieson & Mewhort, 2009). Performance is better if the training list includes better statistical coverage of the grammar (Jamieson et al., 2015; Poletiek & van Schijndel, 2009). Discrimination improves from the start to the end of the test phase (Reber & Perruchet, 2003). Explicit learning disrupts implicit learning (Reber, Kassin, Lewis, & Cantor, 1980; cf. Van den Bos & Poletiek, 2008). People discriminate grammatical from ungrammatical test strings, even if the strings are composed using different letters (Brooks & Vokey, 1991; Jamieson et al., 2015; Manza & Reber, 1997; Reber, 1969; Vokey & Higham, 2005). Both amnesia and forgetting impair recognition of studied exemplars more than judgement of grammaticality (Higham & Vokey, 1994; Jamieson, Holmes, & Mewhort, 2010; Knowlton, Ramus, & Squire, 1992; Nosofsky & Zaki, 1998; Vokey & Higham, 1999).

In this paper, we extend the list of phenomena by showing that judgment of grammaticality is affected by the order in which training strings are studied. To test for an effect of study order, we administered a standard artificial-grammar experiment that tested judgements of grammaticality after participants had studied training strings presented in alphabetical order, in a fixed but random order, in different random orders, or not at all. Our measure of the effect of order is based on agreement about the grammaticality of individual test strings.

The underlying principle is that performance on test trials reflects a comparison of the probe against memory of the studied training strings. Memory of the studied strings, in turn, reflects the way in which they are encoded where training strings encoded early influence encoding of subsequent strings. For example, participants who study SPS first may encode SPSVPS differently than those who study SPSVPS first.

Assuming that the order in which items are studied affects memory for those items, participants who study training strings in different orders should disagree with one another about which test exemplars are grammatical. Conversely, participants who study training strings in the same order should largely agree about which test strings are most and least grammatical. Hence, agreement about which test strings are most and least grammatical provides a measure of the effect of order of presentation. To foreshadow our results, the experiments show that study order does affect judgements of grammaticality, but in a selective manner.

Differences in encoding based on study order touch two important issues: First, current accounts of performance in implicit-learning tasks neglect the effects of study order; hence, any effect of study order in an artificial-grammar experiment challenges existing accounts to come up with a mechanism with which to match performance. Second, effects of study order in implicit-learning tasks should force theorists to acknowledge cue-interaction at study.

## Experiment 1

We conducted a standard artificial-grammar task to evaluate the hypothesis that study order influences judgment of grammaticality. In the experiment, participants studied grammatical training strings and, then, judged the grammaticality of unstudied test strings. One quarter of the participants studied the training strings presented in alphabetical order (fixed-alphabetical), one quarter studied the strings presented in a fixed but random order (fixed-random), one quarter studied the training strings presented in different random orders (random), and one quarter did not study the training strings at all (no-study). The fixed-alphabetical and fixed-random conditions were conducted to evaluate judgment of grammaticality under two different fixed study orders. The random condition was conducted to evaluate judgment of grammaticality with the conventional randomized training procedure and to provide a point for comparison against the two fixed study conditions. Finally, the no-study control condition was conducted to assess judgment of grammaticality independent of study altogether.

We expected that discrimination of grammatical from ungrammatical strings following study would be better than discrimination without study. More critically, if learning depends on order, participants in the fixed-alphabetical study condition should agree with one another but disagree with participants in the fixed-random study

condition about which test strings are most and least grammatical, and vice versa; there may also be a difference in the overall rate of discrimination.

## Method

### Participants

Two-hundred undergraduate students from the University of Manitoba participated in the experiment. The participants were assigned randomly, but in even numbers, to the four study conditions: fixed-alphabetic, fixed-random, random, and no-study control.

### Apparatus

The experiment was administered on eight desktop computers. Each computer was equipped with a standard keyboard, mouse, and a 22-in LCD monitor.

### Materials

The training and test items were generated using the finite-state grammar shown in Fig. 1. The grammar has been widely used (see Schiff & Katan, 2014).

Grammatical strings were sampled from the subset of grammatical items that are at least three and no more than six letters in length. Ungrammatical strings were constructed by randomly sampling between three and six letters from the set *P*, *S*, *T*, *V*, and *X* (i.e., the same letters used to construct the grammatical training strings).

We sampled 40 unique grammatical and 20 unique ungrammatical strings. Half of the grammatical strings were randomly assigned to the training list. The test list included the 20 remaining grammatical strings and the 20 ungrammatical strings. The full set of materials is presented in Table 1. The training strings are presented twice to show the order in which the strings were presented in the fixed-alphabetic and fixed-random conditions.

### Procedure

Participants were tested in groups of six to eight. Each participant was seated at a different computer.

Participants assigned to the fixed-alphabetic, fixed-random, and random study conditions were told that they would be shown 20 nonsense letter strings and that they should do their best to memorize each one. The participants assigned to the no-study condition were told that the 20 strings would be presented subliminally and that they should look at the computer screen until instructions appeared for the test.

**Table 1** Training and test strings presented in Experiment 1

| Item number | Training strings | | Test strings | |
|---|---|---|---|---|
| | Fixed-alphabetic | Fixed-random | Grammatical | Ungrammatical |
| 1 | SPSVP | TXXXXS | SPS | PSVST |
| 2 | SPSVPS | SPSX | SPSVPV | SPP |
| 3 | SPSX | TSVPSX | SPV | SPVST |
| 4 | SPT | TXSVPT | SPVT | SPVVX |
| 5 | SPVV | SPXSVP | SPVVV | SPVX |
| 6 | SPVVT | TXS | SPVVVT | SSPVT |
| 7 | SPVVVV | TXSVP | SPXS | SVVVT |
| 8 | SPXSVP | SPXSX | SPXXS | TPSVPT |
| 9 | SPXSX | SPSVP | SPXXXS | TPSX |
| 10 | SPXXSX | TSVP | TSVPT | TTSVP |
| 11 | TSVP | SPVVT | TSVPV | TXSS |
| 12 | TSVPS | SPVVVV | TSVPVT | TXSVXV |
| 13 | TSVPSX | SPSVPS | TSX | TXT |
| 14 | TSVPVV | SPT | TXSVPS | TXVSVP |
| 15 | TSVPXS | TSVPVV | TXSVPV | TXXSXV |
| 16 | TXS | SPXXSX | TXSX | VSX |
| 17 | TXSVP | TSVPXS | TXXS | VVVPS |
| 18 | TXSVPT | SPVV | TXXSVP | XPVVT |
| 19 | TXXSX | TXXSX | TXXXS | XSXXPS |
| 20 | TXXXXS | TSVPS | TXXXSX | XXSX |

When the participant clicked on the word "Start", the screen was cleared. In the fixed-alphabetic, fixed-random, and random study conditions, the first training string was presented after a 750 ms pause and, then, remained for 4 s. Following the presentation, the screen was cleared. After another 750 ms pause, the next training string was presented. The cycle continued until all of the training strings had been presented once. The procedure was identical in the no-study condition, except that no training strings were presented.

After all of the training strings had been presented (or an equal amount of time had passed in the no-study control condition), instructions for the test phase appeared on the computer screen. The instructions informed participants that the training strings had been constructed according to rules of an artificial grammar and that they would next judge the grammatical status of novel test strings. The participant clicked a button labeled "Start" to begin the test phase.

On each test trial, a string was presented along with two response alternatives labeled "grammatical" and "ungrammatical". A button labeled "OK" was presented beneath the response alternatives. The participant responded by selecting one of the two response alternatives and, then, clicking "OK". When they did so, the screen was cleared. One second later, the next string was displayed. The cycle continued until all of the test strings had been presented, and the participant had provided a response to each one.

Following the test phase, a text editor appeared along with a message that invited the participant to type the rules that they thought had been used to construct the training strings.

## Results and discussion

Table 2 presents the mean percentage of grammatical and ungrammatical strings that participants endorsed as grammatical. Endorsement rates for grammatical strings are hit rates. Endorsement rates for ungrammatical strings are false alarm rates. Standard deviations are shown in parentheses.

**Table 2** Experiment 1: mean percent endorsement rates and standard deviations for grammatical and ungrammatical test strings as a function of study condition

| Study condition | Probe type | |
|---|---|---|
| | Grammatical | Ungrammatical |
| Fixed-alphabetic | 57.9 (21.1) | 41.4 (15.9) |
| Fixed-random | 59.1 (20.6) | 43.6 (17.1) |
| Random | 61.8 (21.4) | 44.9 (15.6) |
| No-study | 30.4 (22.7) | 27.2 (23.0) |

The data were analyzed using a 4 × 2 mixed-factors ANOVA. Study condition was a between-subject factor with four levels (i.e., fixed-alphabetic, fixed-random, random, and no-study), and probe type was a within-subject factor (i.e., grammatical versus ungrammatical).

As shown in Table 2, there was a 16.3 % discrimination advantage for grammatical over ungrammatical test strings in the three trained conditions compared to 3.2 % in the no-study control condition; a test of the two-way interaction confirmed the difference, $F(1, 196) = 23.27$, $p < .0001$. Additional orthogonal tests to partition the two-way interaction confirmed (a) that performance in the two fixed-study conditions was statistically equivalent to performance in the random study condition, $F(1, 196) = .10$, $p > .80$, and (b) that performance in the fixed-alphabetic condition was statistically equivalent to performance in the fixed-random study conditions, $F(1, 196) = .09$, $p > .80$. In summary, studying grammatical strings benefited discrimination of the grammatical status of the test strings, but the order in which the training strings were studied had no effect.

To assess performance at the level of the individual items, we aggregated the data across subjects in each group and then plotted the percentage of endorsements for individual test items in each of the six pairs of study conditions. Figure 2 shows the resulting relations.

Each panel of Fig. 2 shows the item-level relationship between the percentage endorsements for individual strings in two of the four study conditions. Open circles show the relationships for ungrammatical test strings; closed circles show the relationships for grammatical test strings. If the participants in any pair of conditions agreed perfectly about which test strings were least to most grammatical, all points would lie on a straight line with a slope greater than zero.

As shown in top row of Fig. 2, participants in the two fixed conditions and the random condition agreed on the grammatical status of the ungrammatical test strings (open circles), but had much less agreement about the status of the grammatical test items (closed circles). As shown in the bottom row of Fig. 2, participants in the no-study control group did not agree with participants in any of the other three study conditions. The results in Fig. 2 suggest that study order influenced judgements about the grammatical but not ungrammatical strings. For completeness, Table 9 in Appendix presents the mean endorsement per item in each group.

To provide statistical support for the conclusions suggested by Fig. 2, we conducted a Monte-Carlo bootstrap analysis of agreement between all six pairs of study groups. We adopted the bootstrap analysis because it allowed us to index the intragroup as well as the intergroup agreement

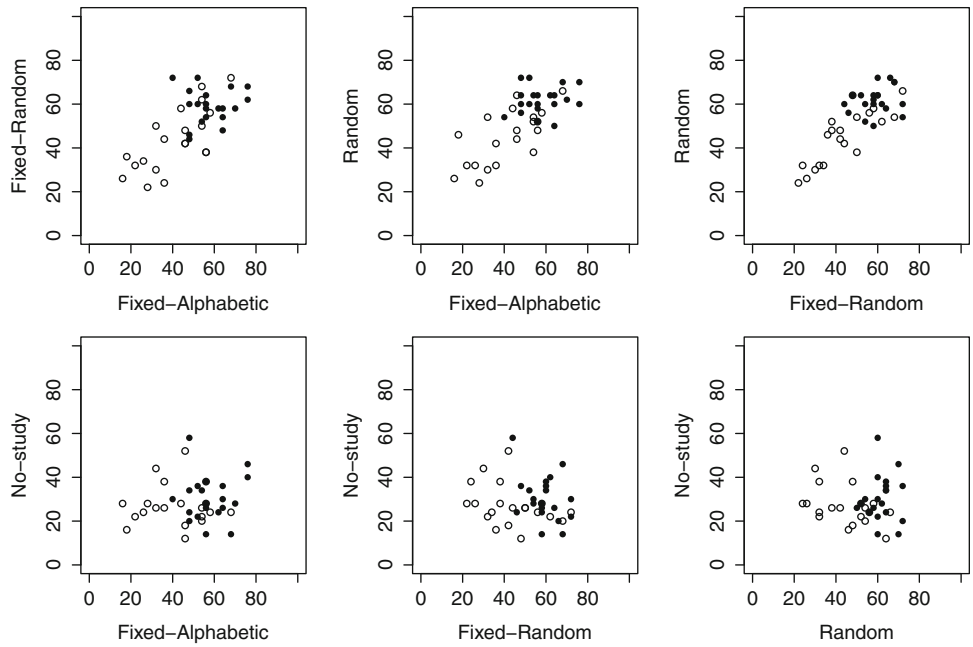


**Fig. 2** Percent endorsement for individual test strings as a function of study condition. *Closed circles* show endorsement for grammatical items and *open circles* show endorsement for ungrammatical items

**Table 3** Experiment 1: mean intergroup correlations and 99 % confidence intervals (end points in parentheses) of item endorsements in the four study conditions

| | Fixed-alphabetic | Fixed-random | Random |
|---|---|---|---|
| *Grammatical strings* | | | |
| Fixed-random | .04 (−.34, .42) | | |
| Random | .08 (−.30, .47) | .11 (−.35, .58) | |
| No-study | .04 (−.37, .44) | −.20 (−.63, .27) | .03 (−.41, .46) |
| *Ungrammatical strings* | | | |
| Fixed-random | .56 (.27, .79)* | | |
| Random | .51 (.18, .78)* | .62 (.25, .86)* | |
| No-study | .01 (−.29, .33) | −.19 (−.53, .12) | −.22 (−.61, .16) |

* $p < .01$

about which test strings were most and least grammatical—something that could not be accomplished using a standard correlation analysis. On each trial of the bootstrap, we sampled (with replacement) 50 participants from each of the four study conditions (i.e., the fixed-alphabetic, fixed-random, random, and no-study). Next we computed the item-level endorsements in each of the four samples and used the endorsement profiles to compute a range-corrected Pearson Correlation (Thorndike Type II) for grammatical and ungrammatical strings in each pair of study conditions. We repeated the procedure 10,000 times and calculated the mean correlation computed for the grammatical and the ungrammatical strings in each pair of conditions. We also calculated the corresponding 99 % confidence interval for each mean. The correlations index the intergroup (i.e., between group) agreement and correct for complications associated with range restriction.

Table 3 summarizes the bootstrap analysis. Note, first, that item-level endorsements in the no-study condition were uncorrelated with the corresponding endorsements by participants in the fixed-alphabetic, fixed-random, and random conditions for both grammatical and ungrammatical strings. Clearly, studying the training strings affected participants' judgements of grammaticality. Next, although there was good agreement on the grammatical status of the ungrammatical items in the fixed-alphabetic, fixed-random, and random groups (i.e., $r \geq .51$), there was little agreement about the grammatical status of the grammatical items (i.e., $r \leq .11$). Finally, note that the 99 % confidence intervals confirm both points: The mean correlation for grammatical strings in the fixed-alphabetic and fixed-random groups (i.e., $r = .04$) falls outside the range of the 99 % confidence interval for the mean of the ungrammatical strings (i.e., 99 % CI .27 −.79), and vice versa. We

conclude that study order affected participants' judgements about the grammatical status of the grammatical but not ungrammatical test items.

The bootstrap analysis summarized in Table 3 considered the intergroup agreement among raters in different study groups. The next analysis considers the intragroup agreement among raters within a group using a split-half correlation. As before, we used a Monte-Carlo bootstrap. The analysis was designed to index how well the item-level endorsement rates from one half of the participants in a study condition predict the item-level endorsements from the other half of participants in the same condition. If study order influences judgements of grammatical but not ungrammatical test strings, judgements by participants in the random study condition should be more consistent for the ungrammatical than for the grammatical items. In contrast, participants in both of the fixed-study conditions should agree about both the grammatical and the ungrammatical test strings.

To implement the bootstrap analysis, we (a) randomly divided the participants into two equal sized groups of 25 (b) computed the item-level endorsements in the two samples (c) computed separate range-corrected correlations of the item-level endorsements for the grammatical and ungrammatical strings, and (d) computed the mean correlation and corresponding 99 % confidence intervals over 10,000 replications.

A summary of the intragroup bootstrap correlations is presented in Table 4. Note that participants in the random condition agreed more strongly with one another about the status of the ungrammatical ($r = .50$) than grammatical strings ($r = .00$), $p < .01$. The result confirms the key prediction. However, whereas when study order was held constant participants agreed about both the grammatical ($r = .38$) and ungrammatical ($r = .64$) strings in the fixed-alphabetic condition, participants in the fixed-random condition agreed about the ungrammatical ($r = .62$) but not the grammatical ($r = .20$) strings. Thus, although the intergroup correlation analysis described earlier supports the idea that study order selectively affects judgements of grammatical test strings, support for the selective effect of study order is mixed in the intragroup analysis.

**Table 4** Experiment 1: mean intragroup correlation and 99 % confidence interval (end points in parentheses) within each study condition

| Study condition | Probe type | |
|---|---|---|
| | Grammatical | Ungrammatical |
| Fixed-alphabetic | .38 (.03, .72)* | .64 (.33, .87)* |
| Fixed-random | .20 (−.14, .59) | .62 (.33, .85)* |
| Random | .00 (−.36, .43) | .50 (.17, .80)* |
| No-study | .50 (.18, .79)* | .45 (.13, .76)* |

* $p < .01$

In Experiment 1, study order did not affect participants' discrimination of grammatical from ungrammatical test strings (see Table 2). Likewise, it did not affect their item-level judgements about the ungrammatical test strings (see Table 3). However, it did affect participants' item-level agreement about the grammatical test strings.

We did not anticipate evidence for order dependence at the item- but not the category-level. Moreover, we were surprised by the selective influence of study order on the participants' judgments about grammatical but not ungrammatical test items. Hence, we decided a replication was in order.

## Experiment 2

Experiment 2 was identical to Experiment 1, with the exception that we used different materials derived from a different grammar. Assuming that Experiment 1 is replicable, we should find the same selective influence of study order on participants' item-level judgments about grammatical but not ungrammatical test strings.

## Method

### Participants

Two-hundred-and-four undergraduate students from the University of Manitoba participated in the experiment. Equal numbers of participants were assigned randomly to the four study conditions: fixed-alphabetic, fixed-random, random, and no-study.

### Apparatus

See "Experiment 1".

### Materials

The strings were constructed using the grammar in Fig. 3 and were at least two and no more than six letters in length. The grammar has been widely used (see Schiff & Katan, 2014).

A set of 46 grammatical and 23 ungrammatical strings was constructed. Half of the grammatical strings were sampled to a training list; the test list included the remaining 23 grammatical strings and the 23 ungrammatical strings. The full set of materials is presented in Table 5. The training strings are presented twice to show the order in which they were presented in the fixed-alphabetic and fixed-random study conditions.

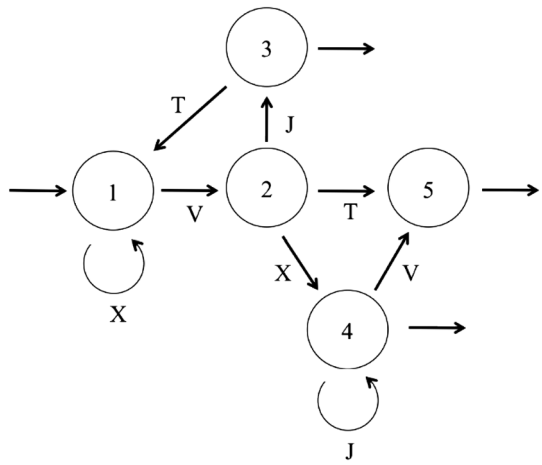


**Fig. 3** The grammar used materials in Experiment 2

## Procedure

The procedure was identical to Experiment 1; only the materials differed.

## Results and discussion

Table 6 shows the mean percent endorsements for both grammatical and ungrammatical test strings as a function of study condition.

As in Experiment 1, the results were analyzed using a 4 × 2 mixed-factors ANOVA. Study condition was a between-subject factor with four levels (i.e., fixed-alphabetic, fixed-random, random, and no-study), and probe type was a within-subject factor (i.e., grammatical versus ungrammatical).

As shown in Table 6, there was a 19.6 % discrimination advantage for grammatical over ungrammatical test strings in the three trained conditions compared to −0.3 % in the no-study control condition. A test of the two-way interaction confirmed the difference, $F(1, 200) = 41.58$, $p < .001$. Two additional orthogonal tests of the two-way interaction failed to confirm a difference in discrimination by the two fixed versus the random study group, $F(1, 200) = .01$, $p > .90$, or by participants in the two fixed study conditions, $F(1, 200) = .25$, $p > .60$. In summary, as

**Table 5** The training and test strings used in Experiment 2

| Item number | Training strings | | Test strings | |
|---|---|---|---|---|
| | Fixed-alphabetic | Fixed-random | Grammatical | Ungrammatical |
| 1 | VJTVT | XXVTVJ | VJ | JJXV |
| 2 | VJTVXJ | VJTVXJ | VJTVJ | JTVJV |
| 3 | VJTXVJ | VX | VJTVTV | JTVTX |
| 4 | VJTXVT | VXJJJ | VJTVX | JVVJJJ |
| 5 | VJTXVX | VTVJ | VTVJJ | JX |
| 6 | VT | VJTXVT | VTVJJJ | JXJ |
| 7 | VTV | XVXJJJ | VXJ | JXJVJT |
| 8 | VTVJ | XXXXVT | VXJJ | JXVXXJ |
| 9 | VX | XVTVJ | VXJJJJ | TTJTVJ |
| 10 | VXJJJ | XXXXVX | XVJ | TTJXX |
| 11 | XVT | XXVT | XVJTVJ | TVTVVX |
| 12 | XVTVJ | XVXJ | XVJTVT | TVVJJV |
| 13 | XVXJ | VJTXVX | XVJTVX | TXJ |
| 14 | XVXJJJ | VT | XVTV | TXTVJ |
| 15 | XXVT | XXXVXJ | XVTVJJ | VJT |
| 16 | XXVTVJ | XXXVJ | XVX | VJXXX |
| 17 | XXVX | XVT | XVXJJ | VVJT |
| 18 | XXXVJ | XXXXVJ | XXVJ | VVVXTX |
| 19 | XXXVT | VTV | XXVTV | VXTXXV |
| 20 | XXXVXJ | XXXVT | XXVXJ | XJVVVV |
| 21 | XXXXVJ | VJTVT | XXVXJJ | XTVTXT |
| 22 | XXXXVT | VJTXVJ | XXXVTV | XVJVTJ |
| 23 | XXXXVX | XXVX | XXXVX | XXVTJ |

**Table 6** Experiment 2: mean percent endorsements and corresponding standard deviations in parentheses for grammatical and ungrammatical test strings as a function of study condition

| Study condition | Probe type | |
|---|---|---|
| | Grammatical | Ungrammatical |
| Fixed-alphabetic | 62.2 (23.3) | 41.6 (18.9) |
| Fixed-random | 59.8 (20.6) | 41.1 (18.6) |
| Random | 55.3 (22.0) | 35.9 (18.5) |
| No-study | 28.6 (24.7) | 28.9 (25.1) |

in Experiment 1, studying grammatical strings benefited discrimination of the test stimuli; however, the order in which the training strings were studied had no effect.

To assess performance at the level of the individual items, we averaged the data across participants in each group and plotted the percentage of endorsements for individual test items in each pair of study conditions. Figure 4 shows the resulting relations. Table 10 in Appendix shows the item-level endorsements in each condition.

Results in Fig. 4 confirm the results from Experiment 1 (see Fig. 2). As shown in top row of Fig. 4, participants in the fixed-alphabetic, fixed-random, and random study conditions agreed on the status of the ungrammatical test strings but showed far less agreement for the grammatical test items. The relations shown in the bottom row of Fig. 4 show no agreement on grammatical status of test strings between participants who studied the training strings and those who did not. The data suggest that study order affected participants judgements of grammatical but not ungrammatical test strings.

To provide statistical support for the conclusions suggested by Fig. 4, we conducted the same intergroup Monte-Carlo bootstrap analysis that we used in Experiment 1. Results of the analysis are summarized in Table 7.

As shown in Table 7, the intergroup correlations were consistent with those from Experiment 1. First, item-level endorsements by participants in the no-study condition differed from those in the fixed-alphabetic, fixed-random, and random conditions. Similarly, as in Experiment 1, although endorsements for ungrammatical strings were consistent amongst participants in the fixed-alphabetic, fixed-random, and random groups (i.e., $r \geq .60$) endorsements for grammatical strings were not (i.e., $r \leq .28$), $p < .01$. The analysis confirms that study order affected participants' judgements about grammatical but not ungrammatical test strings.

We also conducted an intragroup analysis of item-level judgements. We did so using the same method from the analysis in Experiment 1. The results are presented in Table 8.

Unlike the intragroup analysis in Experiment 1, the intragroup analysis of data from Experiment 2 provide unambiguous evidence that study order affected judgement of grammatical but not ungrammatical strings: Participants in the fixed-alphabetic and fixed-random conditions agreed with one another about both the grammatical strings and ungrammatical strings. Finally, as in Experiment 1, participants in the random condition agreed with one another about the ungrammatical strings ($r = .52$) but not the grammatical strings ($r = .17$).

In summary, Experiment 2 corroborates the results identified in Experiment 1. Study order did not affect

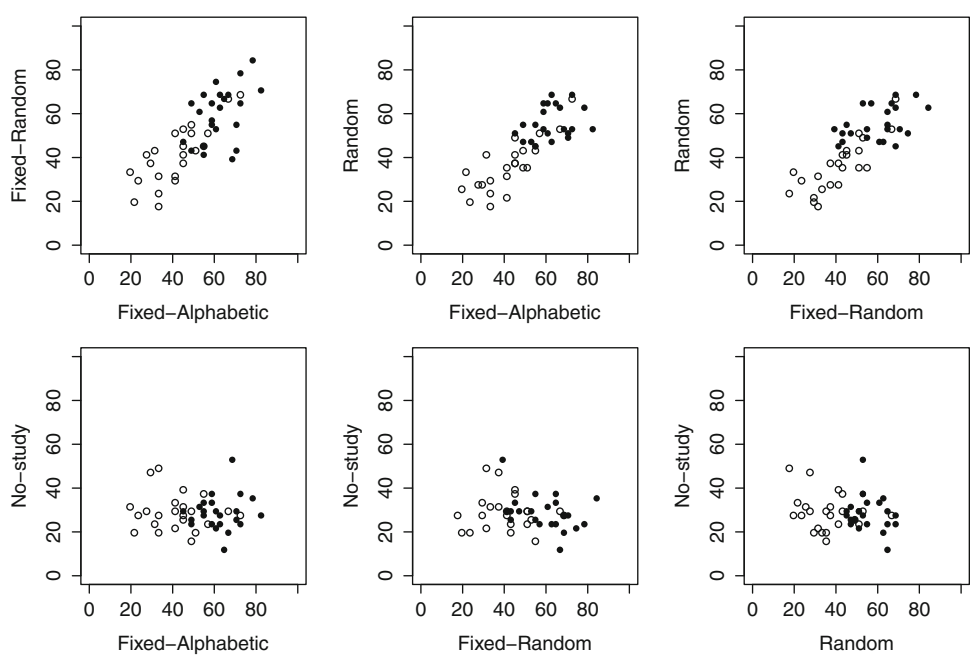


**Fig. 4** The relation between percent endorsement for individual test strings as a function of study condition. *Closed circles* show mean number of endorsements for grammatical items and open *circles* show the corresponding number for ungrammatical items

**Table 7** Experiment 2: mean intergroup correlations and 99 % confidence intervals to measure item-level agreement by participants in the different study conditions

| | Fixed-alphabetic | Fixed-random | Random |
|---|---|---|---|
| *Grammatical strings* | | | |
| Fixed-random | .28 (−.03, .59) | | |
| Random | .17 (−.17, .50) | .23 (−.12, .58) | |
| No-study | .07 (−.25, .40) | −.20 (−.52, .09) | −.11 (−.54, .36) |
| *Ungrammatical strings* | | | |
| Fixed-random | .62 (.33, .83)* | | |
| Random | .60 (.29, .82)* | .60 (.27, .84)* | |
| No-study | −.05 (−.29, .21) | −.02 (−.30, .29) | −.13 (−.42, .14) |

* $p < .01$

**Table 8** Experiment 2: mean intragroup correlations and 99 % confidence intervals to measure item-level agreement by participants in the same study condition

| Study condition | Probe type | |
|---|---|---|
| | Grammatical | Ungrammatical |
| Fixed-alphabetic | .39 (.10, .70)* | .61 (.32, .83)* |
| Fixed-random | .56 (.27, .80)* | .59 (.31, .83)* |
| Random | .17 (−.17, .56) | .52 (.22, .79)* |
| No-study | .33 (.03, .66)* | .37 (.07, .69)* |

* $p < .01$

overall discrimination performance but did affect item-level judgments. Specifically, study order affected participants' judgments about the grammatical but not the ungrammatical test strings.

## General discussion

We conducted two experiments to measure judgment of grammaticality after participants studied training strings in different orders. Study order did not affect category-level discrimination of grammatical from ungrammatical test strings, nor did it affect item-level judgments about ungrammatical test strings. However, the manipulation did affect participants' judgments about the grammatical test strings.

Current theories of implicit learning are ill equipped to explain the selective influence of study order on judgments of grammaticality. Jamieson and Mewhort's (2009, 2011) exemplar model, for example, postulates that participants judge a test string's grammatical status in terms of its global similarity to studied exemplars. Although the model explains much of the artificial-grammar database, it encodes training strings independently of other strings in the training list. Consequently, it cannot handle the effects of study order.

Pothos and Bailey's (2000) adaptation of Nosofsky's (1986) Generalized Context Model also can explain a goodly number of effects in the database, but it cannot address study-order effects: To model memory for the training list, participants provide pairwise similarity ratings for all training and test strings presented in random order. Consequently, memory for the training list is order independent.

Neural-network models also fail, but for different reasons. First, to train such models, the training list must be studied several times. Second, the training list must be presented in different orders (i.e., interleaved learning)—a technique that helps to establish a stable pattern of weights. Third, the learning algorithm requires error-corrective feedback during training—a technical demand that conflicts with the practice of untutored training in implicit-learning procedures (Ratcliff, 1990; Reber, 2002).

Finally, statistical models such as PARSER (Perruchet & Vinter, 1998) and the Competitive Chunking Model (Servan-Schreiber & Anderson, 1990) recognize an influence of order dependence, but both fail on the same pragmatic constraints as the neural-network models: They require the training list to be presented multiple times to derive a stable representation.

Clearly, to handle the order effects documented here, we need to re-think the grounds on which people make their judgements in the artificial-grammar task. There is some evidence that the judgements are not strictly judgements of grammaticality. Sometimes participants make their judgements on the basis of pattern goodness rather than grammaticality (see Jamieson et al., 2015). Further, the information people use to make categorical judgements needs closer scrutiny. Conventional theory assumes that participants actively endorse grammatical strings and passively reject strings that lack grammaticality: That is, the

choice "not grammatical" is a default response applied whenever test strings appear to be insufficiently grammatical. Instead, participants may actively reject exemplars that in some way contradict the information remembered from inspection of the training strings. One might remember, for example, that none of the training strings included a repeated letter and use that fact to actively reject a test string. If so, one might anticipate a stable pattern of judgments about ungrammatical test strings (based on contradiction) while also anticipating an unstable pattern of judgements about grammatical test strings (based on similarity).

In our view, judgement of grammaticality is accomplished by the same mechanisms and processes that underlie explicit learning (e.g., Brooks, 1978; Dulany, et al., 1984; Miller, 1958; Redington & Chater, 1996; Vokey & Higham, 1999, 2004). That is, implicit learning is a phenomenon, not a process. It does not reflect specialized unconscious learning processes. Rather we acknowledge that people can be affected by structure that they cannot articulate and that their behavior can be misconstrued because the structure is correlated with structure that the experimenter prefers (Jamieson et al., 2015). Of course, we are hardly the first to argue this position (e.g., Brooks, 1978; Vokey & Brooks, 1992; Whittlesea & Wright, 1997). Several theorists have attacked the logical models used to divide memory (e.g., Dunn, 2008; Dunn & Kirsner, 1988, 2003; Shanks & St. John, 1994; Van Orden & Kloos, 2003; Van Orden, Pennington, & Stone, 2001) and others have used computational theory to demonstrate that examples of implicit and explicit learning can be accommodated in a single model using the same mechanisms and representations (e.g., Jamieson & Mewhort, 2009; Kinder & Shanks, 2001, 2003; Vokey & Higham, 2004; Zaki & Nosofsky, 2001).

The idea that participants actively reject test stimuli because they contradict information stored during training has precedence in work on implicit learning (e.g., Dulany, Carlson, & Dewey, 1984; Redington & Chater, 1996; Tunney & Altmann, 1999; Vokey & Higham, 1999, 2004) as well as recognition memory. For example, Rotello, Macmillan, and van Tassel (2000), have argued that when faced with a difficult discrimination, participants use a recollection process to remember the studied item closest to the probe—a recall-to-reject strategy. Mewhort and Johns (2000, 2005) have argued for a single process that retrieves a continuum of information starting with superficial familiarity information and ending with item-specific retrieval (see also Johns & Mewhort, 2002, 2003, 2009, 2011; see Pothos, 2005, for a related idea in peoples' judgments of grammaticality).

The present data support the idea that people actively reject ungrammatical test stimuli. For example, some of our participants reported that they deliberately rejected test strings that began with an illegal letter (see Redington & Chater, 1996, for precedence on this point), a strategy borne out by the data. In Experiment 1, participants endorsed 48.0 % of the ungrammatical strings that began with a legal letter but only 32.2 % of the ungrammatical strings that began with an illegal letter. In Experiment 2, participants did the same, with endorsement rates of 48.7 and 33.7 %, respectively. However, it is equally clear that the reported first-letter rule provides neither a complete account of participants' decisions (i.e., some participants endorsed strings that began with illegal letters) nor a complete explanation of order dependence in the data. Nevertheless, the point stands. If participants used contradiction to judge grammatical status, judgments about ungrammatical strings (i.e., strings that include contradictions) should be more stable than judgments about grammatical strings (i.e., strings that do not).

People sometimes may also use memory of already studied strings to facilitate encoding of strings that followed. If a participant studied the string MTXX prior to MTXXVP, for example, they might encode MTXXVP as "MTXX plus VP"; if they studied the string XXVP prior to MTXXVP, they might encode MTXXVP as "MT plus XXVP". Similarly, Vokey & Brooks (1992) showed that a history of learning where strings are encoded using mnemonics (e.g., asking participants to encode VXM as "Virgins eXpect Miracles") can influence how later strings are encoded and test strings are judged. On such occasions, study order would affect memory of the study list and, consequently, judgements of grammaticality. The use of memory in this way is consistent with the spirit of but not the computations in the Competitive Chunking Model (Servan-Schreiber & Anderson, 1990) and Holographic Exemplar Model (Chubala & Jamieson, 2013; Jamieson & Mewhort, 2011) both of which propose that participants notice subunits in strings (e.g., bigrams, trigrams) and that once those subunits are noticed, they go on to influence how subsequent strings are encoded. At test, judgement of grammaticality reflects knowledge of the subunits noticed in training and the organization of those subunits in the training strings.

Dienes (1992) argued that a competent model should predict item- as well as category-level performance: an argument that we call Dienes's dictum and that is valuable on three counts. First, it forces theorists to develop more complete and precise accounts of learning. Second, it forces theorists to develop theories that predict peoples' decisions about stimuli rather than abstract stimulus properties (e.g., grammaticality and similarity). Third, it supports a better model comparison process: whereas every model of implicit learning can predict the discrimination of grammatical from ungrammatical test strings, not every model can predict peoples' judgements about particular test

strings. Our data lend force to Dienes's position: The fact that study order affects item-level judgments without affecting category-level discrimination warns against evaluating theories against category-level data alone.

Fortunately, the problem can be easily addressed by testing models against item- rather than category-level data. Admittedly, the strategy of testing at an item level is not without a cost for both empiricists and theorists: experimentalists must run large-sample experiments to resolve measurement error at the level of individual items, and theorists must reconfigure their models to handle performance at the level of individual items (not item properties). However, the long-term gain is clear. Whereas theories of category-level discrimination will still need to be validated against item-level data, theories of item-level judgments will pre-emptively solve the corresponding category-level problem. Given category-level performance is solved in an account of item-level performance (but not necessarily vice versa), it makes sense to move now towards an item-level approach to modeling.

Dienes's dictum also opens questions about the empirical analysis of learning in the judgement of grammaticality task. Dienes (1992) reported one of the only systematic item-level analyses of peoples' performance in the artificial grammar task (see also Jamieson & Hauri, 2012; Jamieson & Mewhort, 2010). To conduct the analysis, he constructed an empirical database of item-level responses from a number of different experiments using different study procedures and study orders. Once the data were in place, he applied several models to the materials and assessed fit by measuring the rank order correlation between the empirical and simulated item-level endorsement rates. A consistent outcome of the analysis was that the models did a better job of fitting participants' judgments about the ungrammatical than the grammatical test strings. At the time, Dienes argued that the systematic misfit signaled a principled shortcoming in the models he tested. However, he was unable to identify the particular problem. Our data and analysis provides a plausible even if retrospective explanation.

If Dienes's (1992) item-level data were contaminated by an influence of study order (which they were), and the models he tested neglected those influences (which they definitely did), one might argue in hindsight that the neglect of study order caused the models to mispredict judgments of grammatical but not ungrammatical items. The argument points not only of the benefit of analyzing data at a more precise level of analysis but also the insights that can be gained from an increasingly precise examination of behaviour and model fit.

Although the present data show that learning in the judgement of grammaticality task depends on the order in which training string are presented, the demonstration does not compromise the existing database. In most experiments, training items are presented in random order (as in our random condition). In other experiments, training items are presented multiple times in different orders. In yet other experiments, the training items are presented in groups or even all at once. In all cases, order dependence should wash out. Indeed, the conventional and sensible practice of presenting strings in a different random order for each participant presents an excellent method to control for order dependence on measures of category-level discrimination. However, it would be misguided to argue that order dependence does not matter. The order in which items are presented affects learning. More importantly, it has a replicable empirical signature: different study orders affect judgment of grammatical more than ungrammatical test items. Thus, any account that predicts the appropriate influence of order would be an improvement over current theory. By the same token, a model that does not predict order dependence must be incomplete.

Finally, what do our findings say to larger questions of theory? The traditional analysis of artificial-grammar learning focuses on abstraction and a goal to either support the abstractionist position or to explain learning in relation to other factors. The present paper, however, finesses the traditional debate in favor of a core methodological issue: Is performance consistent at the category and item levels? By showing a difference at the two levels, we are drawn into two debates. The first addresses the decision mechanism and, in particular, whether or not people seek positive evidence to endorse exemplars (with failure to endorse as a default) versus taking negative evidence to actively reject them. As we noted earlier, we take the latter view, as does Vokey and Higham (2004) PCA of pixel-maps approach: if a test item contradicts information obtained during study, participants use that information to reject the test item.

The second issue addresses potential cross-fertilization across two traditions in the study of learning. As we have noted, study order has been neglected in the artificial grammar literature, perhaps because people did not expect order effects in an implicit learning task to parallel those in explicit learning. By contrast, study order plays a central role in the study of explicit learning, particularly associative learning. Indeed, study order defines important phenomena in associative learning. Consider, for example, the blocking phenomenon discovered by Kamin (1969, see also Shanks, 1985). If participants learn A+ and then AB+, they report that A predicts the outcome and that B does not. If they learn A+ at the same time as AB+, they report that both A and B predict the outcome. If they learn AB+ and then A+, they report that A predicts the outcome whereas B inhibits the outcome. Indeed, the problem of path dependence in associative learning presents a clear parallel to the problem of order dependence in implicit learning (see Miller, 2006).

Although it is unlikely that study order affects learning for precisely the same reasons in the two situations (see Shanks, 2010), our focus on core methodological issues points to a potential gain for both fields that should help to motivate interest in integrating data and theory in implicit and associative learning (Herbranson & Shimp, 2008; Jamieson, Crump, & Hannah, 2012; Vokey & Jamieson, 2014).

**Acknowledgments** This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to all three authors. We thank Steve Lindsay for advice on the statistical analysis.

## Appendix

See Tables 9 and 10.

**Table 9** Experiment 1: item-level percent endorsement rates (items 1–20 are grammatical; items 21–40 are ungrammatical)

| Number | Item | Fixed-alphabetic | Fixed-random | Random | No-study |
|---|---|---|---|---|---|
| 1 | SPS | 48.0 | 44.0 | 60.0 | 58.0 |
| 2 | SPSVPV | 52.0 | 72.0 | 60.0 | 22.0 |
| 3 | SPV | 76.0 | 68.0 | 70.0 | 46.0 |
| 4 | SPVT | 70.0 | 58.0 | 62.0 | 28.0 |
| 5 | SPVVV | 64.0 | 48.0 | 64.0 | 36.0 |
| 6 | SPVVVT | 64.0 | 54.0 | 60.0 | 30.0 |
| 7 | SPXS | 52.0 | 60.0 | 72.0 | 36.0 |
| 8 | SPXXS | 62.0 | 58.0 | 64.0 | 24.0 |
| 9 | SPXXXS | 56.0 | 54.0 | 52.0 | 28.0 |
| 10 | TSVPT | 48.0 | 60.0 | 64.0 | 34.0 |
| 11 | TSVPV | 56.0 | 64.0 | 58.0 | 26.0 |
| 12 | TSVPVT | 48.0 | 66.0 | 72.0 | 20.0 |
| 13 | TSX | 76.0 | 62.0 | 60.0 | 40.0 |
| 14 | TXSVPS | 68.0 | 68.0 | 70.0 | 14.0 |
| 15 | TXSVPV | 64.0 | 58.0 | 50.0 | 26.0 |
| 16 | TXSX | 48.0 | 46.0 | 56.0 | 24.0 |
| 17 | TXXS | 54.0 | 52.0 | 64.0 | 34.0 |
| 18 | TXXSVP | 56.0 | 58.0 | 60.0 | 14.0 |
| 19 | TXXXS | 56.0 | 60.0 | 64.0 | 38.0 |
| 20 | TXXXSX | 40.0 | 72.0 | 54.0 | 30.0 |
| 21 | PSVST | 54.0 | 62.0 | 52.0 | 22.0 |
| 22 | SPP | 32.0 | 30.0 | 30.0 | 44.0 |
| 23 | SPVST | 68.0 | 72.0 | 66.0 | 24.0 |
| 24 | SPVVX | 58.0 | 56.0 | 56.0 | 24.0 |
| 25 | SPVX | 54.0 | 68.0 | 54.0 | 20.0 |
| 26 | SSPVT | 54.0 | 50.0 | 38.0 | 26.0 |
| 27 | SVVVT | 32.0 | 50.0 | 54.0 | 26.0 |
| 28 | TPSVPT | 44.0 | 58.0 | 58.0 | 28.0 |
| 29 | TPSX | 56.0 | 38.0 | 52.0 | 28.0 |
| 30 | TTSVP | 36.0 | 44.0 | 42.0 | 26.0 |
| 31 | TXSS | 56.0 | 38.0 | 48.0 | 38.0 |
| 32 | TXSVXV | 18.0 | 36.0 | 46.0 | 16.0 |
| 33 | TXT | 46.0 | 42.0 | 44.0 | 52.0 |
| 34 | TXVSVP | 46.0 | 48.0 | 64.0 | 12.0 |
| 35 | TXXSXV | 46.0 | 42.0 | 48.0 | 18.0 |
| 36 | VSX | 36.0 | 24.0 | 32.0 | 38.0 |
| 37 | VVVPS | 16.0 | 26.0 | 26.0 | 28.0 |
| 38 | XPVVT | 22.0 | 32.0 | 32.0 | 22.0 |
| 39 | XSXXPS | 26.0 | 34.0 | 32.0 | 24.0 |
| 40 | XXSX | 28.0 | 22.0 | 24.0 | 28.0 |

**Table 10** Experiment 2: item-level percent endorsement rates (items 1–23 are grammatical; items 24–46 are ungrammatical)

| Number | Item | Fixed-alphabetic | Fixed-random | Random | No-study |
|---|---|---|---|---|---|
| 1 | VJ | 68.6 | 39.2 | 52.9 | 52.9 |
| 2 | VJTVJ | 52.9 | 60.8 | 47.1 | 31.4 |
| 3 | VJTVTV | 54.9 | 41.2 | 45.1 | 29.4 |
| 4 | VJTVX | 62.7 | 62.7 | 47.1 | 23.5 |
| 5 | VTVJJ | 70.6 | 54.9 | 49.0 | 25.5 |
| 6 | VTVJJJ | 54.9 | 45.1 | 54.9 | 33.3 |
| 7 | VXJ | 49.0 | 43.1 | 47.1 | 25.5 |
| 8 | VXJJ | 70.6 | 43.1 | 51.0 | 29.4 |
| 9 | VXJJJJ | 45.1 | 47.1 | 51.0 | 29.4 |
| 10 | XVJ | 58.8 | 54.9 | 52.9 | 37.3 |
| 11 | XVJTVJ | 54.9 | 68.6 | 45.1 | 27.5 |
| 12 | XVJTVT | 60.8 | 74.5 | 51.0 | 21.6 |
| 13 | XVJTVX | 49.0 | 64.7 | 54.9 | 23.5 |
| 14 | XVTV | 58.8 | 64.7 | 60.8 | 33.3 |
| 15 | XVTVJJ | 64.7 | 66.7 | 64.7 | 11.8 |
| 16 | XVX | 72.5 | 64.7 | 52.9 | 37.3 |
| 17 | XVXJJ | 60.8 | 52.9 | 64.7 | 29.4 |
| 18 | XXVJ | 62.7 | 68.6 | 68.6 | 27.5 |
| 19 | XXVTV | 82.4 | 70.6 | 52.9 | 27.5 |
| 20 | XXVXJ | 66.7 | 68.6 | 62.7 | 19.6 |
| 21 | XXVXJJ | 58.8 | 56.9 | 64.7 | 23.5 |
| 22 | XXXVTV | 72.5 | 78.4 | 68.6 | 23.5 |
| 23 | XXXVX | 78.4 | 84.3 | 62.7 | 35.3 |
| 24 | JJXV | 41.2 | 31.4 | 31.4 | 21.6 |
| 25 | JTVJV | 54.9 | 45.1 | 43.1 | 37.3 |
| 26 | JTVTX | 49.0 | 54.9 | 35.3 | 15.7 |
| 27 | JVVJJJ | 33.3 | 17.6 | 23.5 | 27.5 |
| 28 | JX | 29.4 | 37.3 | 27.5 | 47.1 |
| 29 | JXJ | 33.3 | 31.4 | 17.6 | 49.0 |
| 30 | JXJVJT | 27.5 | 41.2 | 27.5 | 29.4 |
| 31 | JXVXXJ | 45.1 | 37.3 | 37.3 | 31.4 |
| 32 | TTJTVJ | 23.5 | 29.4 | 19.6 | 27.5 |
| 33 | TTJXX | 19.6 | 33.3 | 25.5 | 31.4 |
| 34 | TVTVVX | 31.4 | 43.1 | 41.2 | 23.5 |
| 35 | TVVJJV | 21.6 | 19.6 | 33.3 | 19.6 |
| 36 | TXJ | 41.2 | 29.4 | 21.6 | 33.3 |
| 37 | TXTVJ | 41.2 | 51.0 | 35.3 | 29.4 |
| 38 | VJT | 45.1 | 45.1 | 41.2 | 39.2 |
| 39 | VJXXX | 66.7 | 66.7 | 52.9 | 29.4 |
| 40 | VVJT | 51.0 | 43.1 | 35.3 | 19.6 |
| 41 | VVVXTX | 45.1 | 41.2 | 37.3 | 27.5 |
| 42 | VXTXXV | 56.9 | 51.0 | 51.0 | 23.5 |
| 43 | XJVVVV | 33.3 | 23.5 | 29.4 | 19.6 |
| 44 | XTVTXT | 49.0 | 51.0 | 43.1 | 29.4 |
| 45 | XVJVTJ | 45.1 | 52.9 | 49.0 | 25.5 |
| 46 | XXVTJ | 72.5 | 68.6 | 66.7 | 27.5 |

## References


Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 170–211). Hillsdale: Erlbaum.

Brooks, L. R., & Vokey, J. (1991). Abstract analogies and abstracted grammars: comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology General, 120*, 316–323.

Chubala, C. M., & Jamieson, R. K. (2013). Recoding and representation in artificial grammar learning. *Behavior Research Methods, 45*, 470–479.

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

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

Dunn, J. C. (2008). The dimensionality of the remember-know task: a state-trace analysis. *Psychological Review, 2008*(115), 426–446.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: the principle of reversed association. *Psychological Review, 95*, 91–101.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex, 39*, 1–7.

Herbranson, W. T., & Shimp, C. P. (2008). Artificial grammar learning in pigeons. *Learning and Behavior, 36*, 116–137.

Higham, P. A., & Vokey, J. R. (1994). Recourse to stored exemplars is not necessarily explicit: a comment on Knowlton, Ramus, and Squire (1992). *Psychological Science, 5*, 59–60.

Jamieson, R. K., Crump, M. J. C., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning and Behavior, 40*, 61–82.

Jamieson, R. K., & Hauri, B. R. (2012). An exemplar model of performance in the artificial grammar task: holographic representation. *Canadian Journal of Experimental Psychology, 66*, 98–105.

Jamieson, R. K., Holmes, S., & Mewhort, D. J. K. (2010). Global similarity predicts dissociation of classification and recognition: evidence questioning the implicit-explicit learning distinction in amnesia. *Journal of Experimental Psychology Learning Memory and Cognition, 36*, 1529–1535.

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology, 62*, 550–575.

Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial-grammar task: string-completion and performance on individual items. *Quarterly Journal of Experimental Psychology, 63*, 1014–1039.

Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: a reply to Kinder (2010). *Quarterly Journal of Experimental Psychology, 64*, 209–216.

Jamieson, R. K., Nevzorova, U., Lee, G., & Mewhort, D. J. K. (2015). Information theory and artificial grammar learning: Inferring grammaticality from redundancy. *Psychological Research.* [Epub ahead of print]

Johns, E. E., & Mewhort, D. J. K. (2002). What information underlies correct rejections in recognition from episodic memory? *Memory and Cognition, 30*, 46–59.

Johns, E. E., & Mewhort, D. J. K. (2003). The effect of feature frequency on short-term recognition memory. *Memory and Cognition, 31*, 285–296.

Johns, E. E., & Mewhort, D. J. K. (2009). Test sequence priming in recognition memory. *Journal of Experimental Psychology Learning Memory and Cognition, 35*, 1162–1174.

Johns, E. E., & Mewhort, D. J. K. (2011). Serial-position effects for lures in short-term recognition memory. *Psychonomic Bulletin and Review, 18*, 1126–1132.

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42*, 61–112.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kinder, A. (2010). Is grammaticality inferred from global similarity? Comment on Jamieson and Mewhort (2009). *Quarterly Journal of Experimental Psychology, 63*, 1049–1056.

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, 15*, 648–669.

Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: a single-system connectionist account. *Psychological Review, 110*, 728–744.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances. *Psychological Science, 3*, 172–179.

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

Manza, L., & Reber, A. S. (1997). Representing artificial grammars: transfer across stimulus forms and modalities. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford: Oxford University Press.

Mewhort, D. J. K., & Johns, E. E. (2000). The extralist-feature effect: a test of item matching in short-term recognition memory. *Journal of Experimental Psychology: General, 129*, 262–284.

Mewhort, D. J. K., & Johns, E. E. (2005). Sharpening the echo: an iterative-resonance model for short-term recognition memory. *Memory, 13*, 300–307.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology, 56*, 433–491.

Miller, R. R. (2006). Challenges facing contemporary associative approaches to acquired behavior. *Comparative Cognition and Behavior Reviews, 1*, 77–93.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology General, 115*, 39–57.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. *Psychological Science, 9*, 247–255.

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

Perruchet, P., & Vinter, A. (1998). PARSER: a model for word segmentation. *Journal of Memory and Language, 39*, 246–263.

Poletiek, F. H., & Lai, J. (2012). How semantic biases in simple adjacencies affect learning a complex structure with non-adjacencies in AGL: a statistical account. *Philosophical Transactions of the Royal Society, 367*, 2046–2054.

Poletiek, F. H., & van Schijndel, T. J. P. (2009). Stimulus set size and grammar coverage in artificial grammar learning. *Psychonomic Bulletin and Review, 16*, 1058–1064.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Science, 28*, 1–49.

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.

Ratcliff, R. (1990). Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological Review, 97*, 285–308.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour, 5*, 855–863.

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

Reber, P. J. (2002). Attempting to model dissociations of memory. *Trends in Cognitive Sciences, 6*, 192–194.

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology Human Learning and Memory, 6*, 492–502.

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *The Quarterly Journal of Experimental Psychology, 56*, 97–113.

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

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition memory: evidence from ROC curves. *Journal of Memory and Language, 43*, 67–88.

Schiff, R., & Katan, P. (2014). Does complexity matter? Meta analysis of learner performance in artificial grammar tasks. *Frontiers in Psychology, 5*, 1084.

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

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology, 37*, 1–21.

Shanks, D. R. (2010). Learning: from association to cognition. *Annual Review of Psychology, 61*, 273–301.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367–395.

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

van den Bos, E., & Poletiek, F. H. (2008). Intentional artificial grammar learning: when does it work? *European Journal of Cognitive Psychology, 4*, 793–806.

Van Orden, G. C., & Kloos, H. (2003). The module mistake. *Cortex, 39*, 164–166.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science, 25*, 111–172.

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.

Vokey, J. R., & Higham, P. A. (1999). Implicit knowledge as automatic, latent knowledge. *Behavioral and Brain Sciences, 22*, 787–788.

Vokey, J. R., & Higham, P. A. (2004). Opposition logic and neural network models in artificial grammar learning. *Consciousness and Cognition, 13*, 565–578.

Vokey, J. R., & Higham, P. A. (2005). Abstract analogies and positive transfer in artificial grammar learning. *Canadian Journal of Experimental Psychology, 59*, 54–61.

Vokey, J. R., & Jamieson, R. K. (2014). A visual familiarity account of evidence for orthographic processing in baboons (Papio papio). *Psychological Science, 25*, 991–996.

Whittlesea, B. W., & Wright, R. L. (1997). Implicit (and explicit) learning: acting adaptively without knowing the consequences. *Journal of Experimental Psychology Learning Memory and Cognition, 23*, 181–200.

Zaki, S. R., & Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective, & Behavioral Neuroscience, 1*, 344–359.