

Order effects in learning relational structures

Baxter S. Eaves Jr. (B0eave01@Louisville.Edu)

Department of Brain Psychological Sciences
University of Louisville

Patrick Shafto (p.shafto@Louisville.Edu)

Department of Brain Psychological Sciences
University of Louisville

Abstract

Much of the knowledge people acquire is structured: number systems, taxonomies; chemical structures. Learning using the individual components that compose a structured theory may be difficult due to the memory load induced by remembering the entities and their relations. Though much research has demonstrated the effects of ordering on category learning, to our knowledge, none has been conducted on the learning of relational structures. In three experiments we explore the effects of different orderings in learning different relational structures, finding that ordering affects learning, only orderings that tend to eliminate simpler alternative structures are better, and that the complexity of learning appears to be driven by the number of relations, as opposed to the number of nodes.

The effects of data ordering on incremental learning are well-documented. Ordering is of obvious importance in sequence learning in which a human or a machine must learn the sequence of actions that produce a desired effect (e.g., language, planning, skill acquisition, etc) (Clegg, DiGirolamo, & Keele, 1998; Sutton & Barto, 1998). Ordering is studied in instructional design in which students must learn multiple interdependent topics. These topics could be presented in a variety of orders that may lead to different learning outcomes in different contexts (Ritter, 2007, p.19-39). Category learning researches have attempted to formalize methods for presenting data in optimal sequences. Elio and Anderson (1984) found that to facilitate learning of novel categories, it is best to start with low-variance exemplars and gradually increase the variance, Medin and Bettger (1994) showed that it is best to show successive examples that maximize the similarity between exemplars, and Mathy and Feldman (2009) suggest that it is better still to present exemplars in rule-based order in which categories are further divided into subclasses and members of subclasses are shown in succession.

Category learning has been a target for fine-tuned order analysis because it is a well-formalized problem that can be readily adapted for the lab setting. However, in both intuitive experience and educational endeavors, people learn about relations among concepts. For instance, people learn about the relations among categories of living things that compose a taxonomy, the relations among elements that compose the periodic table, and sequences of events that form a causal chain. In each of these cases, the information is not just relational, but can be characterized by an abstract pattern: trees, periods, or chains (Kemp & Tenenbaum, 2008). Recent work by Kemp, Goodman, and Tenenbaum (2008b) has formalized relational theories. Kemp (2008) investigated learning rela-

tional structures based on randomly sampled examples, finding that simpler structures were faster to learn.

While there has been a considerable amount of research into learning of concepts and categories (see Murphy, 2004; Smith & Medin, 1981, for reviews), considerably less work has been done to understand how people learning more rich, relational structures (see Kemp, Goodman, & Tenenbaum, 2008a, 2008b) and as a result, ordering effects in theory learning are not well understood. Models of theory acquisition predict biases toward structures that are compactly-represented in predicate notation (see Kemp et al., 2008a, 2008b)—biased toward simpler structures—but do not explicitly explore the implications of different orderings.

Consider, for example, the relational structure in Figure 1a. The overall structure is a line. Each node (with the exception of the two end nodes) has a single incoming and a single outgoing link. If one attempted to learn this structure, it would require tracking 11 different names, one for each node, and 10 relations among the nodes. If these were independent bits of information, remembering them might be quite difficult (Miller, 1956). Because the relations are structured, it may be possible to learn quite quickly. For example, one might organize the information based on the structure—learning the relations from left to right. In other structures it may not be so obvious which ordering is best. When teaching relations that form a tree (see Figure 1b) is it best to order examples from root to leaves or from level to level? A more thorough analysis of which order is best is needed for these non-obvious cases.

Motivated by previous research formalizing relational theories, we investigate the hypothesis that better orderings are those that rule out other abstract forms; orders that demonstrate the underlying structure of the relations or do not suggest other structures. First, to demonstrate order effects in theory learning, in Experiment 1 we teach a linear structure with linearly-ordered and random examples. In Experiment 2, we begin to speak to which orderings facilitate learning. We teach a binary tree and compare learning outcomes under different orderings inspired by graph-traversal algorithms. Lastly, in Experiment 3 we teach a more complex structure based on electron orbitals. We contrast orders based on how electron orbitals are often presented in textbooks. Across experiments we find that order affects learning but more importantly that different, intuitively-sound orderings can have vastly different effects on learning and that some orders are

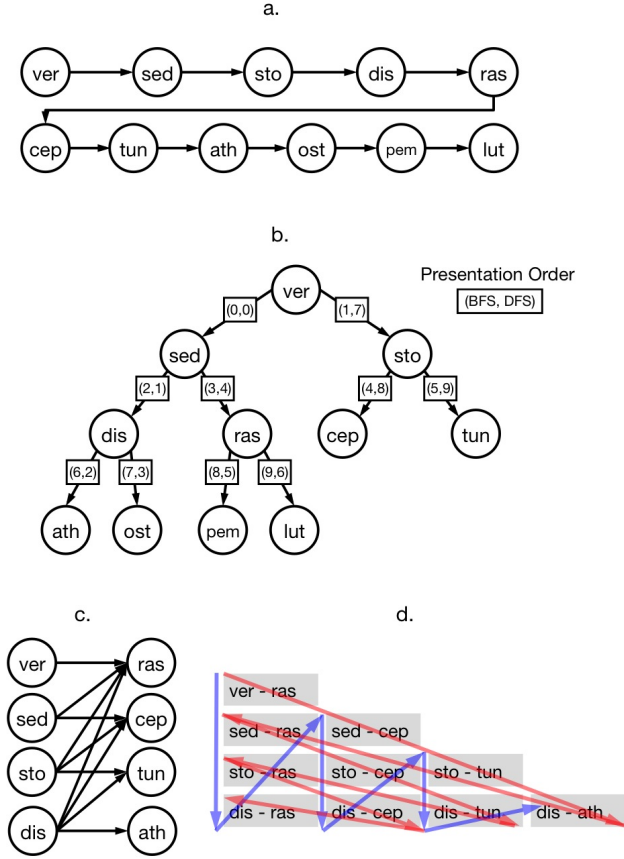


Figure 1: The target structures. a) Linear structure for Experiment 1. b) Binary tree structure for Experiment 2. The order in which the relationships were shown are in the boxes. The first entry in each box is the order in which that relationship appeared in the breadth-first condition; the second entry is the order in the depth-first condition. c) Graphical representation of the structure in Experiment 3. d) Table representation of the structure in Experiment 3. Blue arrows trace the ordering in which each cell was shown in the vertical order condition and the red arrows trace the order of the diagonal condition.

as bad as no order at all.

Experiment 1

To establish the importance of ordering in relational learning, we first selected a simple linear structure consisting of 11 symbols (three-letter nonsense words) connected by 10 links. We contrast performance given two different orderings: a linear order that presents links in order from left to right (see Figure 1a), and a fixed random order.¹ Providing examples in sequence maintains the linear structure, essentially ruling out other structures. With each new example, the learner who receives sequenced examples incorporates new examples sim-

¹Pilot studies showed that, when the random order varied from trial to trial, structures of this complexity were frequently not learnable by participants. For this reason, throughout the paper we use randomly chosen, fixed orders as a baseline.

ply by extending their theory. A learner who receives disjoint examples must concurrently assemble several smaller disjoint substructures of the full structure. We hypothesize that providing examples in order will lead to faster learning than providing examples randomly. We test this hypothesis by comparing the proportion of participants who pass as well as the time and number of repetitions (trials) to successful recall of the examples under random and linear ordering.

Methods

Participants Participants were 44 University of Louisville undergraduates who completed the study for course credit.

Design Participants were randomly assigned to one of two conditions: linear or random. In the linear condition, the relationships were presented in order from left to right, as depicted in Figure 1a. In the random condition, participants were presented examples in an order that was randomly determined at the beginning of the experiment by shuffling the ordered examples, and which was held constant across trials.

Procedure Participants completed the study on computers. To begin, participants were told “You will be shown a series of relationships. Your job is to remember as many as you can.” Participants then progressed through each example one-at-a-time, at their own pace, until they had seen each example. Examples were presented as sets of two symbols separated by a right-directional arrow, e.g. **SED** → **VER**. They were then tested. Participants were presented with a single answer box composed of two blank fields separated by a right directional arrow (____ → ____). Participants could add additional answer boxes. They were asked to fill in as many relationships as they could remember. The study and test phases were repeated until the participant had passed—had recalled each of, and only, the 10 relationships—or had not passed after 25 minutes. The assignment of symbols to nodes in each was constant across participants.

Results

In the ordered condition, 20 of 22 participants passed and in the random condition, 10 of 22 passed. The proportion of passing participants varied significantly between conditions ($\chi^2(1,44) = 7.33, p = 0.007$; see Figure 2a). The difference in the number of trials until completion between passing participants in each condition was not significant ($M_{ord} = 7.25, S_{ord} = 4.24; M_{rand} = 10.5, S_{rand} = 4.97; t(28) = -1.869, p = 0.072$; see Figure 2b). An independent samples t-test revealed that of the participants who passed, those in the ordered condition completed the study in less time than those in the random condition ($M_{ord} = 625 \text{ sec}, S_{ord} = 339 \text{ sec}; M_{rand} = 935 \text{ sec}, S_{rand} = 370 \text{ sec}; t(28) = -2.289, p = 0.030$; see Figure 2c). Taken together, systematic ordering of relations led to a marked decrease in difficulty of learning.

We also investigated omission errors and the learning trajectories to characterize why the ordered condition led to improved performance. Omissions of relations and symbols were calculated by dividing the number of unique (cor-

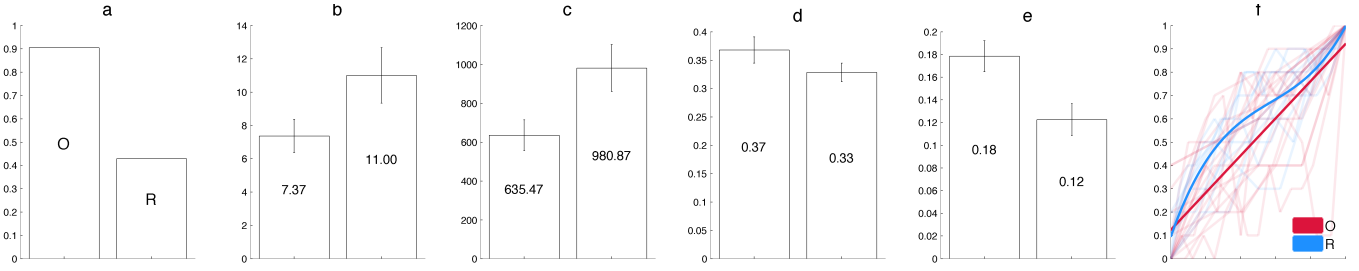


Figure 2: Experiment 1 results. a) Proportion of participants who passed for ordered (left) and random (right). b) Mean and standard error number of trials completed by passing participants. c) Mean and standard error time in seconds taken by passing participants. d) Mean proportion of relationship omissions over trials. e) Mean proportion of symbol omissions over trials. f) Proportion correct (Y-axis) at proportion complete for passing participants in the ordered (red) and random (blue) conditions with first- and third-order polynomial fit lines, respectively.

rect) relations or symbols recalled by the participant by the total number of relations or symbols. Figure 2d and e shows omission errors for relations and symbols. Independent samples t-tests reveal no effect of condition on mean omission of relationships ($M_{ord} = 0.37, S_{ord} = 0.10, M_{rand} = 0.33, S_{rand} = 0.05, t(28) = 1.141, p = 0.264$) or symbols ($M_{ord} = 0.18, S_{ord} = 0.06, M_{rand} = 0.12, S_{rand} = 0.04; t(28) = 2.561, p = 0.016$), suggesting no marked differences in errors across conditions².

To compare the progression of learning under different orderings, in Figure 2f we plot the scaled learning curves. We constructed scaled learning curves for each participants by dividing the number of correct responses at each trial by the total number of relationships in the sequence; we scaled time to completion by dividing a participant's trial numbers by the number of trials taken by that participants. For example, if a participant correctly recalled 2, 4, 8, and 10 relationships, their y-values would be [.2, .4, .8, 1] and their x-values would be [0, .334, .667, 1]. Polynomial lines were fit to the data in each condition. The order of the fit line was that which minimized the Bayesian Information Criterion (BIC): the quantity $(n \ln SSE / n + p \ln n)$, where n is the number of data points, SSE is the sum of squared error between data and the polynomial line, and p is the order of the polynomial. A first-order polynomial (i.e. a straight line) provided the best fit in the ordered condition, suggesting a steady learning progression. In contrast, a third-order polynomial best fit the random condition, consistent with an uneven learning progression.

Presenting data in linear order may help because it maintains the structure. Presenting relations disjointly may lead learners to infer a disjoint structure. Disjoint structures are not as compactly represented because disjoint sets essentially correspond to special cases. Instead of having to remember **VER** → **SED** and **SED** → **STO** individually, the learner directly observes **VER** → **SED** → **STO**. If the learner remembers the sequence of the nodes, she is able to reconstruct the structure.

²Errors of commission were rare, generally one every ten trials, and were therefore not analyzed.

The linear structure is almost trivially simple. In order to speak to which orderings are best we must turn to richer structures that are more representative of real-world structures and which can be presented in several reasonable orders.

Experiment 2

In Experiment 2, we turn our attention to a more interesting case of relational learning, inspired by the problem of learning biological taxonomies. Specifically, we explore ordering effects in a binary tree and propose two non-random orderings inspired by graph-searching algorithms: depth-first (DFS) and breadth-first search (BFS). DFS traverses a graph by traveling down a path until it reaches a dead end (a leaf in our tree) and then back-tracks; BFS visits each node adjacent to the current node before proceeding deeper (see Figure 1b). In this case, BFS better represents and maintains the binary tree structure with each example, therefore we hypothesize that BFS will lead to faster learning.

Methods

Participants were 62 University of Louisville undergraduates who completed the study for course credit. The procedure was identical to that of Experiment 1 barring the different data structure and the additional ordering condition. The tree structure to be learned was composed of 11 symbols and 10 links: the same number of symbols and links as the linear structure in Experiment 1. The same nonsense words used in Experiment 1 were used in Experiment 2. The random order was individually determined for each participant by shuffling the examples before beginning the experiment, during which the order remained constant.

Results

In the BFS condition, 12 of 20 participants passed; in the DFS condition 15 of 21 participants passed; and in the random condition, 13 of 21 participants passed. The proportion of passing participants did not vary between conditions ($\chi^2(2, 62) = 0.68, p = 0.71$). Independent samples T-test revealed that of the participants who passed, those in the

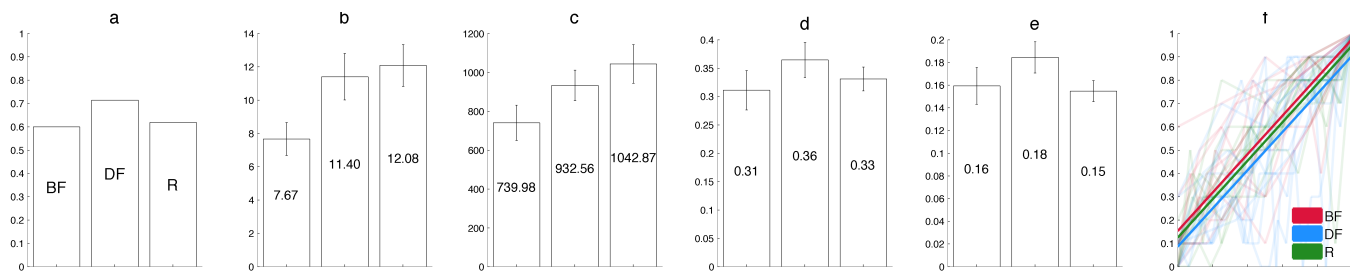


Figure 3: Experiment 2 results. a) Proportion of participants who passed for breadth-first search (left), depth-first search (center), and random (right). b) Mean and standard error number of trials completed by passing participants. c) Mean and standard error time in seconds taken by passing participants. d) Mean proportion of relationship omissions over trials. e) Mean proportion of symbol omissions over trials. f) Proportion correct (Y-axis) at proportion complete for passing participants in the breadth-first (red), depth-first (blue) and random (green) conditions with first-order polynomial fit lines, respectively.

BFS condition passed in fewer trials than those in the DFS condition ($M_{BFS} = 7.67, S_{BFS} = 3.42, M_{DFS} = 11.4, S_{DFS} = 5.4; t(25) = -2.08, p = 0.048$) and those in the random condition ($M_{rand} = 12.08, S_{rand} = 4.52; t(23) = -2.73, p = 0.012$). Of the participants who passed, those in the BFS condition passed more quickly than those in the Random condition ($M_{rand} = 1042sec, S_{rand} = 357sec; t(23) = -2.25, p = 0.035$). While there was a trend toward faster completion in the BFS condition versus the DFS condition, the difference was not significant ($M_{BFS} = 734sec, S_{BFS} = 313sec, M_{DFS} = 932sec, S_{DFS} = 304sec; t(25) = -1.61, p = 0.119$).

Omission errors for relations and symbols are shown in Figure 3d and e. As in Experiment 1, one-way ANOVA's reveal no affect of condition on mean omission of symbols ($M_{BFS} = 0.16, S_{BFS} = 0.06, M_{DFS} = 0.18, S_{DFS} = 0.05, M_{rand} = 0.15, S_{rand} = 0.03; F(2, 39) = 1.533, p = 0.229$) or relationships ($M_{BFS} = 0.31, S_{BFS} = 0.12, M_{DFS} = 0.36, S_{DFS} = 0.12, M_{rand} = 0.33, S_{rand} = 0.08; F(2, 39) = 0.859, p = 0.432$).

A first-order polynomial provided the best fit in each condition, suggesting a steady learning progression.

We suspect that the absence of an effect of condition on the proportion of participants who passed is related to the difficulty of representing the tree compactly. That is, the tree structure inherently places a higher memory load on participants. The effect of ordering is only apparent in the rate of learning. Those in the BFS condition learned more quickly than those in the DFS and random conditions. To speculate, perhaps to learn the tree in the time allotted requires a high working-memory capacity and that these high-working-memory-capacity individuals were helped by the ordering. That is, these individuals were able to hold the individual relations in memory but were best aided by the BFS ordering in finding their arrangement.

Experiment 3

In Experiment 3, we consider a real-world problem from the domain of chemistry: learning atomic orbitals. Two alternative orderings stand out. In textbooks orbitals are often listed

on a grid which can be traversed horizontally, vertically, or diagonally. We contrast vertical and diagonal ordering. Diagonal ordering introduces symbols quickly but presents the relationships disjointly; vertical ordering introduced symbols more slowly while exposing learners to the relationship between symbols. The diagonal ordering also places a higher early memory load on learners by presenting relationships in an order that does not eliminate alternative structures expeditiously. For this reason, we hypothesized that diagonal ordering would be more difficult to learn from than vertical ordering and at least as difficult, if not more so, than random ordering.

Methods

Participants were 59 University of Louisville undergraduates who completed the study for course credit.

The procedure was identical to that of Experiment 1 barring the different data structure and the additional ordering condition. The orbital structure was composed of 8 symbols—a subset of the symbols from Experiments 1 and 2—and 10 links. The random order was individually determined for each participant by shuffling the examples at the beginning of the experiment after which the order remained constant.

Results

In the vertical ordering condition, 17 of 19 participants passed; in the diagonal ordering condition, 10 of 20 participants passed; and in the random ordering condition, 12 of 20 participants passed. The proportion of passing participants varied significantly between conditions ($\chi^2(2, 59) = 7.28, p = 0.026$). An independent samples t-test showed that, of the participants who passed, participants in the vertical ordering condition passed in fewer trials than those in the diagonal-ordering condition ($M_{vert} = 6.75, S_{vert} = 2.42, M_{diag} = 11.9, S_{diag} = 4.28; t(25) = -4.05, p < 0.001$) and those in the random ordering condition ($M_{rand} = 12.25, S_{rand} = 4.56; t(27) = -4.26, p < 0.001$). Participants in the vertical-ordering condition passed in less time than participants in both the diagonal ordering condition

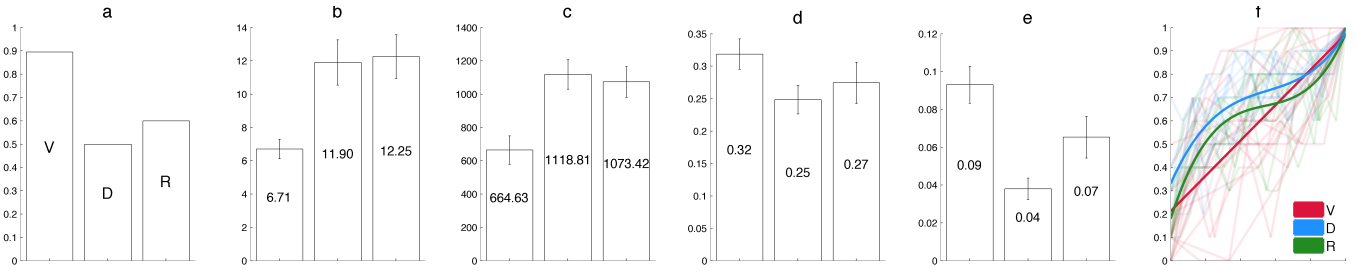


Figure 4: Experiment 3 results. a) Proportion of participants who passed for vertical (left), diagonal (center), and random (right) orderings. b) Mean and standard error number of trials completed by passing participants. c) Mean and standard error time in seconds taken by passing participants. d) Mean proportion of relationship omissions over trials. e) Mean proportion of symbol omissions over trials. f) Proportion correct (Y-axis) at proportion complete for passing participants in the vertical (red), diagonal (blue), and random (green) conditions with with, first-, third-, and third-order polynomial fit lines, respectively.

($M_{vert} = 664sec, S_{vert} = 351sec, M_{diag} = 1118sec, S_{diag} = 282sec; t(25) = -3.48, p = 0.002$) and the random ordering condition ($M_{rand} = 1073sec, S_{rand} = 320sec; t(27) = -3.20, p = 0.003$).

Omission errors for symbols and relations are shown in Figure 3d and e. One-way ANOVA's reveal an effect of condition on mean omission of symbols ($M_{vert} = 0.09, S_{vert} = 0.04, M_{diag} = 0.04, S_{diag} = 0.02, M_{rand} = 0.07, S_{rand} = 0.04; F(2, 39) = 7.711, p = 0.002$) but not relationships ($M_{vert} = 0.32, S_{vert} = 0.10, M_{diag} = 0.25, S_{diag} = 0.07, M_{rand} = 0.27, S_{rand} = 0.11; F(2, 39) = 1.846, p = 0.172$). Bonferroni-corrected post-hoc comparisons revealed that participants in the vertical ordering condition committed significantly more symbol omission errors than participants in the diagonal ordering condition (95% CI [0.019 0.091]).

A first-order polynomial provided the best fit in the vertical-ordering condition, suggesting a steady learning progression. In contrast, a third order polynomial best fit the diagonal and random conditions, consistent with uneven learning progression.

In the case of orbitals, ordering has a clear effect. Presenting relationships vertically leads to quicker learning but causes learners to produce more errors along the way. This results suggests that ordering not only affects the speed of learning but may affect on which parts of the theory learners focus. Here, the diagonal ordering exposes learners to the symbols most quickly. If learners expect teachers to begin with important information, learners may allocate their efforts more heavily to remembering the symbols rather than the relationships among the symbols. Symbol-focused orderings like this may result in fewer errors early on, but because they place such a high emphasis on symbols rather than structure, may damage learners' ability to learn structure, in turn damaging their ability to compress the information and slowing learning.

General Discussion

We investigated the effects of ordering information on relational learning across three different structures: a line, a tree,

and the real-world case of electron orbitals. Based on previous theory, we hypothesized that orderings that eliminated simpler, alternative-domain structures would lead to faster learning. Across all three experiments, the results confirmed this basic hypothesis.

We also attempted to characterize differences in learning across the conditions by analyzing errors of omission and learning curves. Though ordering affected errors of omission in Experiment 3, over experiments, errors of omission provide little evidence for systematic differences. Learning curves provide evidence that learning follows a linear progression when provided with better orderings, while worse orderings lead to non-linearities in learning. These non-linearities capture a temporary leveling that occurs midway through learning. This is consistent with the idea that participants may realize the structure midway through, which requires a reconceptualization of the domain. Systematic investigations of why orderings facilitate learning are an important direction for future work.

Previous results suggest that learning times are dependent on the type of structure (De Soto, 1960). Looking across experiments, our results show no effects of structure. One-way ANOVAs show that among participants who passed, the time and number of trials until completion for participants in the random and poorly-ordered conditions (random, DFS-tree, and diagonal-orbital) did not vary ($F_{time}(4, 58)=0.63, p=0.64; F_{trials}(4, 58)=0.13, p=0.97$), nor did time and number of trials for participants in the better-ordered conditions ($F_{time}(2, 47)=0.35, p=0.74; F_{trials}(2, 47)=0.29, p=0.75$). This suggests that the time to learn was not related to the structure, nor to the number of nodes.

The orderings that best preserve the structure with each subsequent example facilitated learning. This ordering has two related effects: it rules out alternative hypotheses and reduces memory load. For example in Experiment 3, the good ordering reduces memory load early on by minimizing the number of symbols introduced and by allowing for a more compact representation of the structure, which according to a representation-length bias, should reduce inferences over al-

ternative structures (Kemp et al., 2008a).

It is also possible that participants expect orderings that facilitate learning and treat bad orderings as misleading. If participants assumed that they were being *taught* a structure, in the sense that examples had been chosen by a teacher, then bad orderings could have lead them to incorrect hypotheses because teachers should produce examples in a way that optimally teaches the target structure (Shafto & Goodman, 2008; Shafto, Goodman, & Griffiths, 2014).

Here we have focused on one way to facilitate learning: by ordering the information. However, in textbooks, structures, such as atomic orbitals and taxonomic trees, are often presented as a whole, in figures and in tables. This raises two questions. First, how does ordering compare with holistic depictions such as figures? Second, given data in figure format, in what order do learners choose to review the items?

Recent developments in formalizing and modeling learning of structured relations and intuitive theories open the door to investigating knowledge that more closely approximates intuitive theories and the scientific theories taught in educational settings. Understanding how to facilitate learning in the context of these richer structures is a fundamental theoretical question that is of great potential practical importance. For example, constructing a formal framework that chooses pedagogically optimal orderings of arbitrary concepts may take some of the guess work out of instructional design. We have focused on one technique for facilitating learning, finding that ordering information can lead to sharply decreased learning times. There are many other possibilities, and it is important for future work to investigate how to leverage theoretical advances in knowledge representation to practical problems of facilitating learning in more realistic settings.

Acknowledgments

This research was supported by NSF grant DRL-1149116 to Shafto.

References

- Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in cognitive sciences*, 2(8), 275–281.
- De Soto, C. B. (1960). Learning a social structure. *The Journal of Abnormal and Social Psychology*, 60(3), 417.
- Elio, R. & Anderson, J. R. (1984, January). The effects of information order and learning mode on schema abstraction. *Memory & cognition*, 12(1), 20–30.
- Kemp, C. (2008). *The acquisition of inductive constraints* (Doctoral dissertation, Massachusetts Institute of Technology).
- Kemp, C., Goodman, N., & Tenenbaum, J. (2008a). Learning and using relational theories. *Advances in neural information processing systems*, 20, 753–760.
- Kemp, C., Goodman, N., & Tenenbaum, J. (2008b). Theory acquisition and the language of thought. In *Proceedings of thirtieth annual meeting of the cognitive science society*.
- Kemp, C. & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, 105(31), 10687–10692.
- Mathy, F. & Feldman, J. (2009, December). A rule-based presentation order facilitates category learning. *Psychonomic bulletin & review*, 16(6), 1050–7.
- Medin, D. L. & Bettger, J. G. (1994, June). Presentation order and recognition of categorically related examples. *Psychonomic bulletin & review*, 1(2), 250–4.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2), 81.
- Murphy, G. L. (2004). *The big book of concepts*. MIT press.
- Ritter, F. E. (2007). *In order to learn: how the sequence of topics influences learning*. Oxford University Press, USA.
- Shafto, P. & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In *Proceedings of the thirtieth annual conference of the cognitive science society*.
- Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: teaching by, and learning from, examples. *Cognitive psychology*, 71, 55–89.
- Smith, E. E. & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press Cambridge, MA.
- Sutton, R. S. & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT Press.