

Learning a novel rule-based conceptual system

Joshua S. Rule,¹ Steven T. Piantadosi,² & Joshua B. Tenenbaum¹

¹Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

²Department of Psychology, University of California, Berkeley

Abstract

Humans have developed complex rule-based systems to explain and exploit the world around them. When a learner has already mastered a system’s core dynamics—identifying its primitives and their interrelations—further learning can be effectively modeled as discovering useful compositions of these primitives. It nevertheless remains unclear how the dynamics themselves might initially be acquired. Composing primitives is no longer a viable strategy, as the primitives themselves are what must be explained. To explore this problem, we introduce and assess a novel concept learning paradigm in which participants use a two-alternative forced-choice task to learn an unfamiliar rule-based conceptual system: the MUI system (Hofstadter, 1980). We show that participants reliably learn this system given a few dozen examples of the system’s rules, leaving open the mechanism by which novel conceptual systems are acquired but providing a useful paradigm for further study.

Keywords: Concept learning; Induction; Rule learning

Introduction

The puzzle of human learning and cognitive development—how humans learn so much from so little (data) so quickly—is central to cognitive science (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Part of the solution appears to be that humans develop systems in which concepts can be represented and reasoned about productively (Carey, 2009; Fodor, 1975; Smith & Medin, 1981). These conceptual systems support many of humanity’s great cultural achievements: law, science, government & social cooperation, mathematics, language & literature, etc.

Despite concepts’ diverse forms and contents (e.g. objects, agents, magnitudes, categories and kinds, relationships, and events), one striking feature of conceptual systems is that they often appear to be well described as rule-based systems (Table 1). By rule-based, we mean that the behavior of the system can be described as a set of conditional transformations (i.e. If X , then Y) applied over a set of objects. That so many and such widely varying domains should be appropriately described using rules is remarkable, even surprising (e.g. Hamming, 1980; Wigner, 1960). Nonetheless, computationally precise, rule-based descriptions—i.e. programs—remain a powerful tool for representing knowledge.

The exact substrate of rule-based thinking remains unclear. Human learning and cognition may, for example,

Domain	Examples
logic	first-order, modal, & deontic logic
mathematics	counting, algebra, topology
scientific theories	Mendelian genetics, mechanics
productivity systems	Inbox Zero, Getting Things Done
operating procedures	Robert’s Rules, codes of conduct
games & sports	Go, football, 8 queens
norms & mores	class systems, social cliques
legal codes	constitutions, contracts, civil code
religious systems	monastic orders, vows, rituals
mundane chores	tying shoes, mowing lawns
intuitive theories	theory of mind, intuitive physics
kinship	family trees, clan systems
domain theories	cooking, lockpicking, carpentry

Table 1: A sampling of domains well described using rule-based systems, with motivating examples.

rely primarily on distributed or subsymbolic representations (LeCun, Bengio, & Hinton, 2015; Rumelhart, McClelland, & PDP Research Group, 1987), with rules being a recent and difficult-to-use innovation for storing, transmitting, and applying knowledge. Alternatively, rules may be basic mental representations but difficult to acquire. In that case, rules may be stored and applied primarily through innate, domain-specific resources developed on an evolutionary timescale (Fodor, 1975, 1981; Laurence & Margolis, 2002). A third option is that rules are a basic representational tool and that mechanisms for acquiring new rules are essential to human-like learning.

Whatever the case, recent progress has been made in modeling human learning and cognition by treating concepts as programs in a mental programming language such that learning new concepts maps onto discovering new and useful programs (Anderson, 2009; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Klahr, Langley, & Neches, 1987; Lake, Salakhutdinov, & Tenenbaum, 2015; Lenat, 1983; Newell, Shaw, & Simon, 1959; Piantadosi, Tenenbaum, & Goodman, 2016; Sussman, 1973). These models learn programs from observations, a technique known as inductive programming (Flener & Schmid, 2008; Muggleton & De Raedt, 1994), part of the broader field of program synthesis (Gulwani, Polozov, & Singh, 2017). The collective success of these models in explaining human performance on multiple tasks across many domains is strong evidence for program-like, and

Rule	Example
"Mx" \rightarrow "Mxx"	"MIU" \rightarrow "MIUIU"
III \rightarrow U	"UIIIM" \rightarrow "UUM"
UU \rightarrow ϵ	"IUUMI" \rightarrow "IMI"
I" \rightarrow IU"	"MUMI" \rightarrow "MUMIU"

Table 2: Douglas Hofstadter’s MUI system transforms strings composed of three symbols M, U, and I according to the four rules above. In our paradigm, these symbols were mapped to colored dots (M \mapsto ●, U \mapsto ●, I \mapsto ●).

thus rule-based, conceptual representations.

This prior work, however, assumes the set of rules affecting objects is pre-established by a fixed grammar or programming language. Learning is primarily searching through the space to find a specific program or combination of rules to explain observations. This differs from a problem children (and professional scientists, see Gopnik, 1996), often face when learning: searching to explain observations while simultaneously developing the language or set of rules through which to search (Carey, 2009). In kinship, for example, children must first develop concepts of gender, marriage, and parent-child relationships before they can understand terms like grandparent or in-law. In number, children must discover the natural numbers themselves to explain the usefulness of previously learned routines like counting. This sort of conceptual change is what we seek to study.

In this work, we use an unfamiliar rule-based system explicitly constructed to require mastering new rules in a new domain and test naive participants’ ability to learn its dynamics. This means that to succeed, participants must induce the grammar and rules themselves rather than simply composing pre-existing primitives. We specifically use the MUI system introduced by Douglas Hofstadter (1980). The concepts in MUI are strings of three symbols, M, U, and I, which can be transformed using four simple rules (Table 2). While Hofstadter originally used the system to illustrate certain points about provability, we use it here for its balance of novelty and simplicity. Its rules are novel and not obviously drawn from other systems with which participants are likely to be familiar (e.g. number, kinship), but simple enough that they could conceivably be learned from examples in a single sitting (i.e. not rocket science).

We introduce a novel concept learning paradigm built around a game in which participants help scientists to unlock doors in an alien labyrinth. Each lock face presents a Two-Alternative Forced-Choice (2AFC) triad which can be correctly solved by applying the MUI rules to select one of the two alternatives given the prompt. Using this paradigm we find that: **1)** participants reliably learned this system given a few dozen examples of the system’s rules; **2)** there was significant variation in how easily

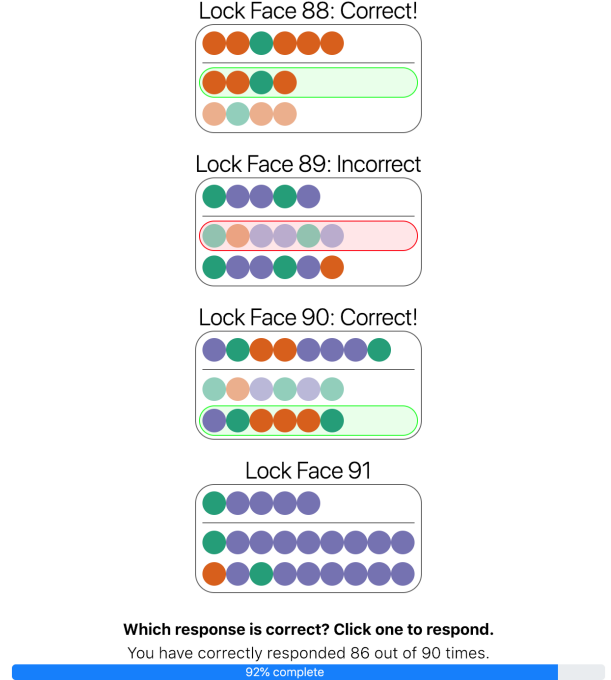


Figure 1: Task overview: Each lock face has a challenge (top row) and two responses (one correct and one incorrect, randomly ordered). Correct (green) and incorrect (red) feedback is shown for previous trials and the current trial (trial 91) is at the bottom of the display; From top to bottom, the four lock faces use rules $UU \rightarrow \epsilon$, $I" \rightarrow IU"$, $III \rightarrow U$, and $"Mx" \rightarrow "Mxx"$, respectively.

individual rules were acquired; **3)** there was significant variation in both how quickly and how reliably individual participants learned; and **4)** participants spontaneously used rule-based language to describe what they learned.

Experiment

This experiment studied how people learn a system of rule-based concepts from examples. Participants repeatedly predicted how the system would transform an input sequence into an output sequence, choosing between one of two alternatives. Because our focus was on testing for learnability, we introduced participants to the conceptual system gradually, introducing one new rule at a time until the entire system had been presented.

Participants and Design We recruited 100 participants (49 female; 48 male; 2 other; 1 NA; age mean: 36.77yrs, sd: 10.49yrs) from Amazon Mechanical Turk. Participants were paid a flat fee of \$3 and up to \$4.14 in bonuses (losing \$0.01 for each incorrect response; mean: \$6.29, min: \$4.83, max: \$7.12, std: \$0.55). The experiment took 36.50 minutes on average to complete (min: 15.07min, max: 126.15min, std: 19.37min).

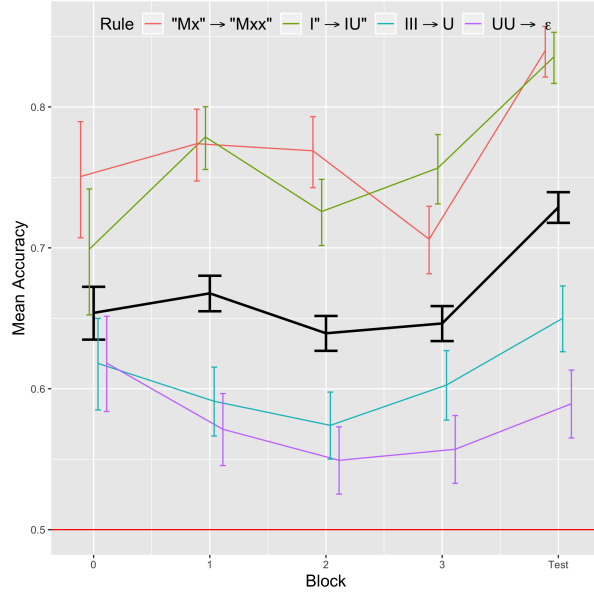


Figure 2: Mean performance (y-axis) by block (x-axis) by rule (individual curves). The black line aggregates across rules. CIs are 95% binomial confidence intervals.

Materials and Procedure Participants played a 2AFC game in which they were asked to help a team of scientists study locked doors in an alien labyrinth. Each lock contained three rows of colorful dots. Dot sequences were used in lieu of strings of M, U, and I to further accentuate the novelty and avoid associations with character-based string manipulations used in natural language or mathematics ($M \mapsto \text{green dot}$, $U \mapsto \text{orange dot}$, $I \mapsto \text{blue dot}$). The top row was called the challenge, and the bottom rows were called responses. The exact sequences of dots changed on each attempt to unlock the door; a specific combination of challenge and responses was called a lock face. Each lock face always had one correct response and one incorrect response. Choosing the correct response would unlock the door, but an incorrect response would leave the door locked and advance the lock to a new lock face.

On each trial, participants saw a unique lock face and attempted to choose the correct response. The challenge contained one to nine dots. The correct response was the result of a single application of one of the four rules of the MUI system to the challenge. The incorrect response was created by selecting a dot sequence which matched the correct response both in length and in edit distance from the challenge. Participants were told that locks contained several mechanisms, and for each lock face, one of these mechanisms created the correct response to the challenge. Crucially, participants were never given example descriptions of specific mechanisms nor told that each mechanism could be described using simple rules.

After choosing a response, participants received visual and written feedback. Incorrect responses also incurred a

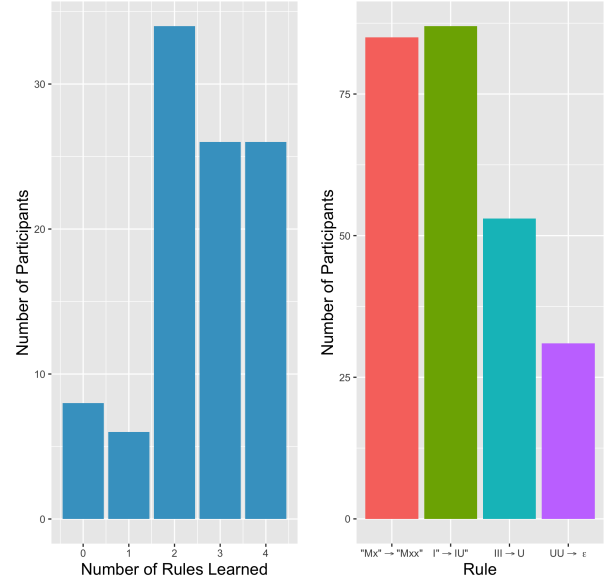


Figure 3: Testing block performance assessing participant rule-learning. Rules are considered learned if χ^2 test over performance on a rule in the testing block is significant at $p < 0.05$ level. Number of participants (y-axis) who learned: **(Left)** each of N rules (x-axis); **(Right)** each rule (x-axis).

four-second delay. Past trials and feedback accumulated on screen with the current trial always displayed at the bottom of the screen along with a performance summary.

Participants completed five blocks of trials: four training blocks followed by a testing block. Each participant training block introduced a single new rule from the MUI system; we refer to this as the target rule and trials using this rule as target trials. The order in which rules were introduced was randomized for each participant. Each block contained 50 target trials and 50 other trials drawn uniformly from the previously learned rules. The first block, having no previously learned rules thus contained just 50 trials, while the other training blocks contained 100 trials. The order of trials within the block was randomized, save that the final trial was always a target trial. Participants could end a block early by correctly responding to 9 out of 10 consecutive target trials. The testing block contained 16 trials for each rule for a total of 64 trials, randomly ordered. Participants were not explicitly informed of this structure but were told that a lock's mechanisms would activate gradually. They were informed each time a new block started, being told that a new mechanism had activated and reminded of the total number of active mechanisms. They were never told which mechanism a particular lock face used. After completing all five blocks, participants were asked to complete a post-task survey reporting how they thought each mechanism worked.

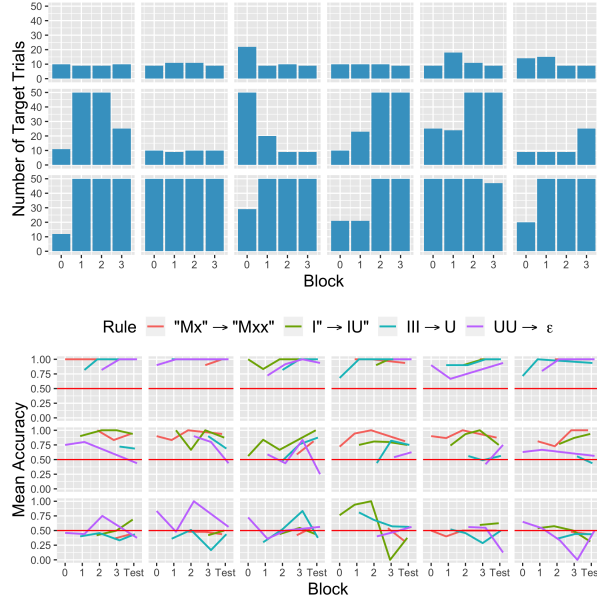


Figure 4: Individual variation for 18 participants. The top, middle, and bottom rows of each subplot contain the top, middle, and bottom 6 participants, respectively, by testing block performance. **(Top)** Number of target trials (y-axis) for each training block (x-axis). There are a maximum of 50 target trials per block; **(Bottom)** Mean performance (y-axis) by rule (individual curves) by block (x-axis).

Results Participants completed a 2AFC task in which they were asked to predict which of two sequences of dots could be derived from a prompt sequence. Participants found the task difficult (self-reported difficulty on 7-point Likert Scale, 1: not difficult, 7: very difficult, mean: 5.29, sd: 1.34) but extremely engaging (self-reported engagement on 7-point Likert Scale, 1: not engaging, 7: very engaging, mean: 6.22, sd: 1.25).

Participants reliably learned the system. Mean accuracy across all rules and all training blocks was 0.651, significantly above chance ($\chi^2(1)=1763.7$, $p<0.001$, 95% CI=[0.644, 0.658]). Mean accuracy in the testing block was also significantly above chance at .729 ($\chi^2(1)=1338.6$, $p<0.001$, 95% CI=[.718, .740]). This level of performance was significantly higher than in the training blocks (Figure 2; $\chi^2(1)=130.62$, $p<0.001$, 95% CI=[0.065, 0.090]). 77 individual participants performed significantly above chance in the testing block, which is unlikely to occur by chance (Figure 5; $\chi^2(100)=1833.9$, $p<0.001$). Performance was also significantly better than chance in the testing block for all four rules (Table 3). The mean (sd) participant performed better than chance on 2.56 (1.17) rules in the testing block, with 26 individual participants learning all 4 rules, 26 learning 3 rules, 34 learning 2 rules, 6 learning 1 rule, and just

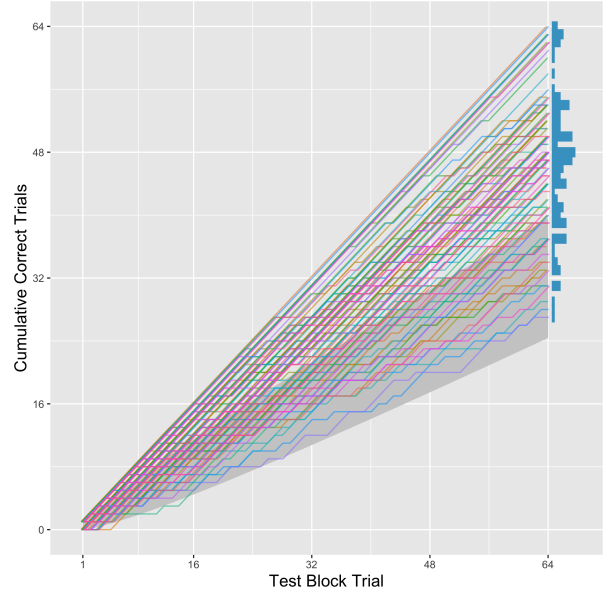


Figure 5: Testing block cumulative performance (y-axis) for each trial (x-axis) by participant (individual curves). The gray region is a 95% binomial confidence interval around chance performance. The histogram shows how many participants performed at each level.

8 failing to learn any rules (Figure 3, Left). 85 people learned "Mx" → "Mxx", 87 learned "I" → "IU", 53 learned "III" → "U", and 31 learned "UU" → ε (Figure 3, Right; Table 3). Looking just at target trials (i.e. trials for the rule being learned during a training block), participants completed significantly less than the ceiling of 50 trials in each block ($ps<0.001$; mean [95% CI] for block 0: 24.7 [21.5, 27.8]; 1: 26.8 [23.6, 30.1]; 2: 29.0 [25.4, 32.6]; 3: 28.5 [25.0, 32.0]). Learning the dynamics of each new rule is what allowed them to end blocks early.

There was significant variation in how easily specific rules were acquired. Performance in the training blocks differed significantly by rule ($\chi^2(3)=586.39$, $p<0.001$), being significantly better for trials using the rules "Mx" → "Mxx" and "I" → "IU" than those using "III" → "U" or "UU" → ε ($\chi^2(1)>200$, $ps<0.001$), and for trials using "III" → "U" than those using "UU" → ε, although the difference was small ($\chi^2(1)=7.686$, $p=0.006$, 95% CI=[0.008, 0.0447]). Performance between "Mx" → "Mxx" and "I" → "IU" was not significantly different ($\chi^2(1)=0.037$, $p=0.848$). Participant performance in the testing block similarly differed ($\chi^2(3)=400.06$, $p<0.001$), being significantly better for trials using the rules "Mx" → "Mxx" and "I" → "IU" than "III" → "U" or "UU" → ε ($\chi^2(1)>140$, $ps<0.001$) and for trials using "III" → "U" than those using "UU" → ε ($\chi^2(1)=12.22$, $p<0.001$, 95% CI=[0.026, 0.095]). Performance between "Mx" → "Mxx" and "I" → "IU" was not significantly different ($\chi^2(1)=0.083$, $p=0.774$).

A two-way repeated-measures ANOVA showed a main effect of rule ($F(3)=12.976$, $p<0.001$) and of block ($F(1)=6.713$, $p=0.010$) but no interaction ($F(3)=2.090$, $p=0.101$). These results suggest that the number of trials required to learn a new rule increased as the complexity of the task increased and, critically, that the number of trials required to learn a new rule also differed significantly by rule. Pairwise comparisons show that "Mx" \rightarrow "Mxx" and "I" \rightarrow "IU" both require significantly fewer trials than $UU \rightarrow \epsilon$ or $III \rightarrow U$ ($ps<0.001$), and $III \rightarrow U$ requires fewer trials than $UU \rightarrow \epsilon$ ($p=0.037$), but that the number of trials needed for "Mx" \rightarrow "Mxx" and "I" \rightarrow "IU" did not significantly differ ($p=0.972$).

There was significant variation in how quickly and reliably participants learned Because participants could end blocks early through accurate performance, difference in block length reflect differences in learning. Participants completed 52–350 training trials and 64 testing trials with a mean \pm sd of 192.98 ± 79.18 trials (block 0: 24.7 ± 15.6 ; 1: 53.6 ± 33.5 ; 2: 57.8 ± 36.6 ; 3: 56.9 ± 35.6). The number of target trials per block also varied significantly by participant ($F(3)=1.988$, $p<0.001$). This effect can be seen visually by examining the number of target trials required in each training block for different participants (Figure 4, Top). High-performers (as judged by test performance) require very few target trials to master each rule and complete each training block, mid-range performers sometimes learn quickly and other times slowly, and low-performers almost invariably hit the ceiling of 50 target trials. Performance in the testing block also varied significantly by participant ($\chi^2(99)=625.17$, $p<0.001$). As with block length, this effect can be seen by comparing the behavior of individual participants (Figure 4, Bottom). High-performers always perform well above chance and having learned a rule once, tend to perform well on all future trials using that rule. Mid-range performers sometimes learn and remember a rule and other times appear to forget. Low-performers rarely perform above chance for more than a single block and frequently perform below chance.

Participants spontaneously used rule-based language. While our main focus in this work was assessing learnability through performance, we asked participants to verbally report at the close of the experiment how they thought each mechanism worked. Participants were neither shown descriptions of specific mechanisms nor told participants that each mechanism could be described using simple rules. Their freeform descriptions, then, provide some insight into the structure of the mental representations they used to solve the task. Participants described the lock mechanisms in various ways, but consistently used rule-like language (Table 4). Not all participants, however, were able to articulate their

Rule	Mean	95% CI	Learned
"Mx" \rightarrow "Mxx"	0.84	[0.821, 0.857]	85
"I" \rightarrow "IU"	0.84	[0.817, 0.853]	87
$III \rightarrow U$	0.65	[0.626, 0.673]	53
$UU \rightarrow \epsilon$	0.59	[0.565, 0.613]	31

Table 3: Test performance by rule, reporting the mean accuracy (out of 1600 trials), a 95% χ^2 confidence interval, and the number of participants who performed above chance for this rule. Crucially, performance across participants is above chance for all rules ($ps<0.001$).

understanding in rule-based language. One participant in the top quartile, for example, said that, "you have to memorize which sequence opens which lock" (i.e. there was no clear pattern or rule). Given that each trial was unique, however, memorization would not have proven helpful. Another claimed, "The mechanisms seemed to be generated at random by a binary sequence."

Discussion & Conclusion

We explored the question of how human learners acquire a novel conceptual system. To do so, we introduced and assessed a gamed-based paradigm in which participants made repeated 2AFC predictions with feedback to acquire an unfamiliar rule-based conceptual system: the MUI system. We found that: **1)** participants, as a whole, reliably learned the system; **2)** there was significant variation in how easily individual rules were acquired; **3)** there was significant variation in both how quickly and how reliably individual participants learned; and **4)** participants spontaneously used rule-based language to describe what they had learned, whether or not their understanding matched the underlying MUI system. While this is a preliminary investigation, these results suggest that humans are able to rapidly explain observations of novel rule-based systems while simultaneously learning the underlying rules of the system.

This work is limited in several important ways. First, rule-based conceptual systems are often difficult to acquire. Many participants, even after extensive practice, did not learn 1 or more rules and/or performed at chance in the testing block. Studies in the history of science and developmental psychology further suggest that many important theories and conceptual systems (e.g. number, classical mechanics) were painfully constructed over centuries and are acquired only with difficulty over the course of many years (see e.g. Carey, 2009). While the difficulty of discovering and/or acquiring a conceptual system like number can be partially explained by the scope of the problem, MUI is a relatively simple system. More work is needed to explain the variability in performance on this task.

Second, we have not explained why rules of seem-

"Mx" → "Mxx"
<ol style="list-style-type: none"> 1. The pattern of dots following the starting color is to be repeated exactly. 2. The original string + the original string minus the first colored circle. 3. Another of the mechanisms mirrored the sequence on the end of it.
I" → IU"
<ol style="list-style-type: none"> 1. Add an orange dot to the end. 2. Copies the original challenge then ends with an orange dot. 3. One of the mechanisms adds a orange circle at the end.
III → U
<ol style="list-style-type: none"> 1. 3 purple circles were converted into 1 orange circle. 2. The final mechanism combined three blue dots to make one orange dot. 3. Choose the answer that as is the same as the face, only 3 purple can = 1 orange.
UU → ε
<ol style="list-style-type: none"> 1. Subtracted two orange dots from the sequence. 2. Minus two reds. 3. sometime 2 orange ones are took away and thats the right answer.
Other
<ol style="list-style-type: none"> 1. Pick the one that looked most like the original. 2. Choose one where the first circle and the last circle are each the same as on the lock face. 3. I thought it depended on where the green one was in relation to the orange one.

Table 4: Representative participant descriptions of the mechanisms of participants. Correct or incorrect, participants described looking for patterns or rules and frequently used rule-based language.

ingly similar syntactic complexity differ dramatically in learnability. One explanation may be that MUI contains multiple types of rules. "Mx" → "Mxx" and I" → IU" are both additive—they add characters to the string without removing characters—and UU → ε is subtractive—removing one or more characters from the string. III → U is a hybrid, replacing one set of characters with another. At the same time, "Mx" → "Mxx" and I" → IU" both apply in only a single place within a string. "Mx" → "Mxx" applies to an entire string, and I" → IU" applies only at the end of a string. III → U and UU → ε, by contrast, can both apply anywhere within a string and perhaps in multiple places. While we started our exploration of rule-based learning with a known system, future work should more carefully tease apart the many classes of rules to understand how rule structure affects learnability.

Third, MUI is a string transformation system, and strings are simple structures compared even to other formal structures like trees or graphs, much less to naturalistic structures like natural kinds or events. Adapting this paradigm to work over many kinds of structures, including these more complex varieties, could significantly

strengthen the approach taken in this paper.

Fourth, most conceptual systems are acquired in environments affording significantly richer interaction and exploration than our tightly controlled paradigm. We are exploring alternatives in which participants can better control their own learning through self-supervision (choosing from a set of possible next problems) and active learning (creating their own problems to solve).

Perhaps the most significant limitation, however, is that we lack a computationally precise model of human performance on this and similar tasks. We have not yet developed an adequate model of how humans learn novel conceptual systems. We are actively working toward such a model based on the term rewriting formalism (Bezem, Klop, & de Vrijer, 2003) and building on previous work (Rule, Schulz, Piantadosi, & Tenenbaum, 2018; Schmid, Hofmann, & Kitzelmann, 2009).

References

- Anderson, J. R. (2009). *How can the human mind occur in the physical universe?* Oxford U. Press.
- Bezem, M., Klop, J. W., & de Vrijer, R. (2003). *Term rewriting systems*. Cambridge U. Press.
- Carey, S. (2009). *The origin of concepts*. Oxford U. Press.
- Flener, P., & Schmid, U. (2008). An introduction to inductive programming. *AI Review*, 29(1), 45–62.
- Fodor, J. A. (1975). *The language of thought*. Harvard U. Press.
- Fodor, J. A. (1981). The present status of the innateness controversy. In *Representations: philosophical essays on the foundations of cognitive science*. MIT Press.
- Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, 32(1), 108–154.
- Gopnik, A. (1996). The scientist as child. *Philosophy of Science*, 63(4), 485–514.
- Gulwani, S., Polozov, O., & Singh, R. (2017). Program synthesis. *Foundations and Trends in Programming Languages*, 4(1-2).
- Hamming, R. W. (1980). The unreasonable effectiveness of mathematics. *The American Mathematical Monthly*, 87(2), 81–90.
- Hofstadter, D. R. (1980). *Gödel, Escher, Bach*. Vintage Books New York.
- Klahr, D., Langley, P., & Neches, R. (1987). *Production system models of learning and development*. MIT press.
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332–1338.
- Laurence, S., & Margolis, E. (2002). Radical concept nativism. *Cognition*, 86(1), 25–55.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436.
- Lenat, D. B. (1983). EURISKO: A program that learns new heuristics and domain concepts—the nature of heuristics III: Program design and results. *AI*, 21, 61–98.
- Muggleton, S., & De Raedt, L. (1994). Inductive logic programming: Theory and methods. *J. of Logic Programming*, 19, 629–679.
- Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a general problem solving program. In *IFIP congress* (Vol. 256, p. 64).
- Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psych. Review*, 123(4), 392.
- Rule, J., Schulz, E., Piantadosi, S. T., & Tenenbaum, J. B. (2018). Learning list concepts through program induction. In *The 40th Annual Conference of the Cognitive Science Society*.
- Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1987). *Parallel distributed processing* (Vol. 1). MIT press Cambridge, MA.
- Schmid, U., Hofmann, M., & Kitzelmann, E. (2009). Analytical inductive programming as a cognitive rule acquisition devise. In *Proceedings of the 2nd conference on artificial general intelligence* (2009).
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard U. Press.
- Sussman, G. (1973). *A computational model of skill acquisition* (Unpublished doctoral dissertation). MIT.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022), 1279–1285.
- Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications on pure and applied mathematics*, 13, 1–14.