Two presentations of a mathematical system: Each has advantages and combined they are better

Andrew K. Lampinen and James L. McClelland

Department of Psychology, Stanford University

Abstract

Previous research has examined how different presentations of a mathematical concept affect learning of that concept (e.g. (Kaminski, Sloutsky, & Heckler, 2008)). However, mathematics is a richly structured field of study, where concepts are explained in terms previously learned concepts, and includes distinct aspects of understanding of a single mathematical object, such as processes for manipulating it vs. formal ways to represent it. Previous research has generally been limited to exploring the effects of presentation only on the single concept being presented. Here, we explore how these presentations can affect subsequent concepts and aspects of understanding which build upon the original concept. Using the domain of elementary group theory, we teach subjects a group operation using a visuospatial or a numerical presentation, or both. We then teach them concepts like inverses and generators which build upon this operation. We demonstrate that the presentations produce differential understanding of various aspects of reasoning, and that rather than one presentation being clearly superior, each has advantages and disadvantages. Instead of pursuing a single ideal presentation, we show that presenting both presentations and encouraging subjects to integrate them leads to better performance, at least for some participants.

Two presentations of a mathematical system: Each has advantages and combined they are better

Introduction

What is the purpose of a pedagogical presentation of a mathematical concept? How do features of the presentation affect understanding of the concept being presented? Given that mathematics is highly structured and concepts are connected in a variety of ways, how does the presentation of one concept affect understanding of related concepts?

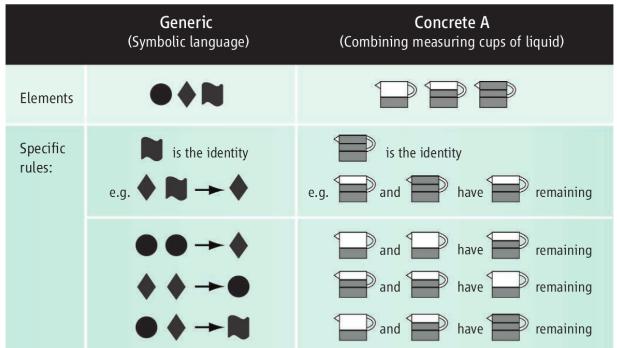
As the name suggests, pedagogical presentations are generally used to present a broader concept, category, or idea, and to link it to other related concepts. However, usually the presentation will not be perfect, in the sense that only some of its features will be category-general. In addition, students may have some preconceptions about the objects included in the presentation. Both of these factors may bias the inferences students make about the concept being explained. Thus changing the way a concept is presented may alter what students learn. Kaminski et al. have demonstrated this using different presentations of a cyclic group (Kaminski et al., 2008).

How do these presentations interact with other concepts? Nothing in mathematics is taught in isolation, there are multifarious relationships among mathematical concepts. The fact that concepts are organized and intricately related, and that this affects learning, has been considered for a long time within cognitive psychology, (e.g. Fischer, 1980), and more specifically within math cognition, (e.g. Hazzan, 1999). In particular, teachers often rely on previously learned concepts to teach a new idea, and students rely on previously learned concepts to understand. Thus we expect that the presentations used to teach a concept can also affect students' understanding of later concepts which are related to it.

In this project, we explored these issues and found evidence that two different presentations of a concept each have relative advantages, supporting different aspects of students' understanding. Building on this, we explored the possibility that exposure to both might allow students to benefit from the advantages of both presentations. We found evidence that some students were able to achieve the benefits of both presentations, and that the benefit increased as students practiced answering questions about the concepts that had been described to them.

We examined these issues within the area of elementary group theory, specifically cyclic groups and some group theoretic concepts relevant to them (a brief introduction to the relevant concepts is provided in Appendix A). Our use of this particular task domain was inspired by previous work on presentations in mathematics education.

(a) Group presentations



Generic and concrete instantiations of a mathematical group.

(b) Transfer domain

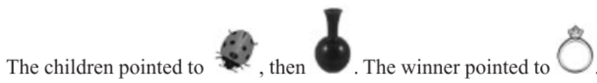


Figure 1. Group presentations from (Kaminski, Sloutsky, and Heckler, 2008) (permission pending)

Background: The Advantage(?) of Abstract Examples

Kaminski and colleagues (Kaminski et al., 2008) explored the effects of presentations in abstract algebra. They presented subjects with either a "generic" instantiation of a cyclic group of order 3, or a "concrete" one. Their presentations are illustrated in figure 1a. The generic representation consists of some arbitrary geometric symbols, with enforced rules for combining them, and the concrete representation consisted of an example with a narrative about combining cups of liquid, and finding the amount left over. There were two other concrete representations (not shown) that were also used in some experimental sessions (using pizza slices and tennis balls, respectively, as the concrete objects.) They trained subjects to perform the operation in either the generic representation or one to three concrete representations. They then showed subjects the isomorphic transfer domain shown in figure 1b, where the objects of the group are replaced by toys in a children's game. (The transfer domain item shown is analogous to the rule on the last line of 1a.) The subjects were explicitly told that this followed the same rules as the earlier examples, and that they should try to use their knowledge to predict the correct answers. Kaminski and colleagues found that the subjects who learned the generic representation performed better at this transfer than the subjects who learned the concrete representation(s). From this, they concluded that "instantiating an abstract concept in a concrete, contextualized manner appears to constrain that knowledge and hinder the ability to recognize the same concept elsewhere" (Kaminski et al., 2008).

However, it is possible to offer alternative interpretations of Kaminski and colleagues' results. For example, Jones (2009) pointed out that in the concrete representations “the feature in question ... is the physical objects that behave like quantities” and the problems can be solved by adding and subtracting, whereas in the generic representation “the symbols used do not appear to represent *quantities*, and are not combined,” and the transfer task, similarly “does not exhibit a quantitative feature; instead it is another version of the generic instatiation with a different contextualization.” Thus he concludes that “The transfer task is more similar to the generic instantiation than to the concrete ones.” In a response to this interpretation, Kaminski, Sloutsky, and Heckler (2009) asserted that the generic and transfer domains were not more similar, because after describing the domains to a set of subjects (without teaching them the rules for combinations), and asking them to rate the similarity between domains, they did not find any significant differences in rated similarity. However, not presenting the rules makes it difficult to be certain that this comparison truly captures the similarity between the domains.

For example, one aspect of the representations which is different is the asymmetry that subjects previous arithmetic knowledge will introduce between the elements which are represented as 1/3 of a cup or 2/3 of a cup in the concrete instantiation. Although in the abstract sense, it is clear that the generic domain and the concrete are isomorphic, in the generic domain the symmetry between the two non-identity elements is clear, circle circle = diamond, and diamond diamond = circle. While the rules that $1 + 1 = 2$ and $2 + 2 = 1$ do follow from the presentation in the numeric case, there is a fundamental asymmetry to the arithmetic interpretations of them (i.e. $1 + 1 = 2$ because 1/3 cup two times makes 2/3 cups, but $2 + 2 = 1$ because 2/3 cup two times makes 1 and 1/3 cups, and we throw away the full cup to get back to 1/3). We suspect this asymmetry may be to blame for the worse transfer performance, since students looking for a cue to map one object to a unit quantity would not find any such specific cue. Similarly, if the notion of generators (see appendix) had been discussed in the study, students would probably have been biased to choose 1 as a generator, even though 2 is an equally good choice, whereas in the generic case there would be no such bias. The presentations provide an easy way to map one of the concrete examples onto another (numerical relationships), which is not present either in the generic or transfer examples. The numerical content of the concrete presentations may change what subjects learn from them.

This idea that what is learned is changed by the presentation is supported by De Bock et al., in their replication of Kaminski’s study (De Bock, Deprez, Van Dooren, Roelens, & Verschaffel, 2011). In this study, they compared the transfer from the generic domain to the concrete, and found that it was worse than the transfer from the concrete domain to a new concrete domain, or from an abstract to an abstract. Thus, each representation was better for transferring to representations similar to it. Furthermore, they asked subjects to give a free response justifying their answer to a difficult problem, and rated it on the ideas that it contained. They found that generic-presentation group subjects were mentioning more group-theoretic ideas (although they still appeared to attain very little understanding of them), but that concrete-presentation group subjects were mentioning the ideas of modular arithmetic as well as some group-theoretic ideas. Thus, the choice of representation had an effect not just on transfer, but on the more abstract concepts being inferred. However, De Bock et al. did not thoroughly explore this idea. They asked only one question of subjects, and were only able to grade on the concepts the subjects explicitly mentioned, so they may have missed understanding which the subjects did not choose to explain, perhaps because it seemed obvious or because they were not comfortable enough with the concept. Furthermore, they did not explore how these inferences subjects made would affect their ability to actually learn the relevant concepts later, as they would in a real pedagogical setting.

Indeed, both Kaminski et al. (and De Bock et al.) omitted many essential features of real educational settings. They did not include much pedagogical explanation of the concepts in question, instead presenting the concepts as a set of rules that only had meaning by their relation to the subjects previous knowledge. They tested only on transfer to a mathematically isomorphic concept, whereas most examples in math instruction are intended to illustrate something more general (a teacher does not show students that $5 + 6 = 11$ just so they can add 5 and 6 in the future, but rather to illustrate the more general principles of addition, carrying, etc.) Furthermore, they only explored subjects procedural ability with the group operation. They did not evaluate how presentations affected subjects ability to learn other related concepts, or more formal ways of understanding the group in question.

We believe that examining the effects of presentations on other concepts is vital, because mathematical concepts are generally not presented in isolation, but rather within a richly structured web of previously learned concepts. In the next section, we explore these ideas in more detail, by examining some of the ways in which concepts are related to one another in mathematics instruction, and consider some of the aspects of understanding which we might hope for a presentation to support.

Relationships Among Mathematical Concepts

How are concepts related to each other in mathematics, and how does this affect mathematical cognition? This is a large topic worthy of a more general investigation, as there are many kinds of relationships between concepts, some of which underlie abstract fields of mathematics such as category theory. Here, we focus specifically on the relationships that are introduced to students when a concept is explained in terms of previously learned concepts.

For example, consider arithmetic. Multiplication is often explained as repeated addition; division may be explained as “undoing” multiplication. These are pedagogically useful relationships between one arithmetic concept and another. Examples that demonstrate arithmetic concepts also often make connections to students experiences and intuitive ideas ("Jane has twelve apples, and wants to share them evenly with her three friends..."). Furthermore, once students understand the arithmetic operations, concepts like primality can be explained in terms of conditions on how numbers behave under them.

When students move on to algebra, they learn more powerful formal ways of manipulating numerical concepts, but they learn them as extensions of the rules of arithmetic they already know. Thus concepts also support later formalisms and other aspects of understanding. For example, the concept of variables as unknowns can be introduced by just substituting a variable in as the solution of a problem the students can already solve (e.g. "$5 + 6 =?$" to "$5 + 6 = x$, solve for x"). Concepts in mathematics are not presented in isolation, but are explained in terms of the related concepts that students have previously learned.

How does this affect learning? Orit Hazzan has suggested that students learning a new concept in abstract algebra reduce the level of abstraction by relying on properties of more concrete examples which they understand (Hazzan, 1999), i.e. the concepts they have previously learned. For example, a student learning a theorem about which elements generate a cyclic group of order n may think about specific examples, such as a cyclic group of order 6.

Because students rely on earlier concepts to understand new ones, presentations of these earlier concepts may have an effect on later learning. The free response results from De Bock et al. (2011) support this idea, by suggesting that the concepts subjects were inferring depended on the presentation. Thus it is important not only that a presentation convey a concept clearly, but also that it provide a foundation for understanding related concepts that will be learned later.

For example, consider cyclic groups, a specific example of which was the domain used by Kaminski et al. (2008) and De Bock et al. (2011). Cyclic groups consist of some elements and a way of combining two of them to get another element. In the cyclic group with n elements, the elements can be essentially thought of as the numbers 0 to $n - 1$, and the way of combining them as addition modulo n (i.e. addition where if you get a result greater than $n - 1$, you just subtract n from it). Kaminski et al. only evaluated subjects understanding of this operation.

However, group theory is a rich domain and in a real math class there are many other concepts that students would learn along with the operation of a cyclic group. They might learn about the identity of the group (the element which leaves every other element unchanged under the operation, i.e. 0), or inverses (the inverse of an element is the thing you combine with it to get the identity), or generators (an element x is a generator of a group if you can make every other element of the group by adding x to itself some number of times). They might also be asked to generalize this understanding to non-isomorphic cyclic groups, or to make general and possibly formal statements about the family of all cyclic groups. These related concepts and more formal aspects of understanding might also be affected by the presentation of the group operation.

We propose the following set of questions for evaluating a presentation's impact on various aspects of understanding (with examples from cyclic groups):

- Does it allow students to understand that particular instance of the concept? (Does it allow students to learn the group operation in group of a specific size?)
- Does it allow them to understand secondary concepts explained in that instance? (Does it allow them to understand inverses and generators?)
- Does it allow them to transfer this knowledge to new, non-isomorphic instances? (Can they do the operation in a cyclic group of different size, and can they work with the concepts of inverses and generators in the new group?)
- Does it allow them to generalize about a class of instances? (Can they explain in words how a procedure like finding the inverse works across groups of any order?)
- Does it allow them to express and understand these generalizations using formal mathematical expressions and language? (Can they write a formula for the inverse in a generic group of order n, or correctly assess formal theorems about which elements are generators?)

Obviously, a single presentation may not address all of these points adequately, but it is important to consider all of them when evaluating a presentation. As mentioned above Kaminski et al. (2008) (and the follow-up work discussed above) focused primarily on the first question. Here, we move beyond this to ask which presentations are better for advancing which of these aspects of understanding. We demonstrate that two presentations which give similar performance on understanding the concept of a cyclic group can each have advantages and disadvantages for learning later group theoretic concepts.

Given multiple presentations, each with unique advantages, which presentation should we teach? Instead of forcing ourselves to choose one and lose the benefits of the other, we propose presenting both to students and explaining the connections between them. In this way, students may be able to achieve the benefits of both. We demonstrate that at least some of our subjects are able to achieve "best of both worlds" results from seeing two presentations.

General Experimental Overview

We conducted a series of experiments investigating the effects of presentations, using two isomorphic presentations of a cyclic group. One presentation is based on a visuospatial manipulation involving counting around the verticies of a polygon, and the other is based on arithmetic and is closely related to modular arithmetic. We refer to these as the polygon and modular presentations, respectively. (See the Materials & Methods section below for more detail.) We used the group theoretic concepts of identities, inverses, and generators, as well as generalization from specific examples of cyclic groups to a generic cyclic group of order n, to investigate the effects of these presentations on different aspects of understanding, and found that while they each were very successful at conveying the group operation, they produced differential understanding of later related concepts. Furthermore, neither was clearly superior, each had advantages and disadvantages relative to the other. Thus we explored combining these presentations to produce a hybrid presentation. We found that this hybrid presentation was beneficial for at least some subjects.

Experiments

Introduction

In this paper, we present the results from three closely related experiments. (These experiments were performed sequentially in order to explore new hypotheses and replicate previous results, but the results are interleaved here for the sake of brevity and coherence.) The goals of the experiments were as follows:

Experiment 1: In our first experiment we explored whether the polygon and modular presentations produced differential performance, and if so, for which aspects of understanding.

Experiment 2: In our second experiment, we had two goals. First, we wished to replicate the results of our first experiment with a planned analysis (to ensure that the effects were not just chance variation, since we didn't have *a priori* hypotheses about which presentation would be superior for which types of questions). Second, we wished to explore whether we could improve overall performance by teaching the subjects the hybrid presentation which included both the polygon and modular presentations, and encouraging the subjects to integrate them (while keeping total instruction time approximately the same).

Experiment 3: In our third experiment, we wished to replicate the results of our previous experiments, and to further explore the thought processes of hybrid-group subjects. In order to examine this, we added questions for the hybrid group (presented after the main experiment had been completed), in which they described the extent to which they had used each representation when answering a question.

Materials & Methods

All materials are available on our github[1], including complete versions of our experiments, which can be downloaded and run, or viewed using github's html preview.

The experimental layout was as follows:

1. Training on group operation (order 6 group)

[1] `https://github.com/lampinen/fyp_cyclic_groups`

2. Training on concepts of identity, inverses, and generators

3. Test of ability to transfer concepts to a new cyclic group (order 9)

4. Test of ability to formulate concepts at a general level about a family of groups (order n)

5. (Representation-use questions, only for hybrid group and only in experiment 3.)

6. Demographic and background questions

We taught the subjects to perform the group operation of a cyclic group of order 6 (using the polygon, modular, or hybrid presentation, between subjects), and then taught them the concepts of identities, inverses, and generators using this operation. The explanations of identities, inverses, and generators were the same between experimental groups (we did not need to refer to the specifics of the underlying operation), to ensure that any effects we observed were due to the different presentations of the underlying concept. For example, for inverses we explained that "the inverse of a number is the element that you combine with it to produce the identity."

We then tested subjects' transfer of these concepts to a cyclic group of order 9. (These group orders were chosen in order to have enough elements for demonstrations of concepts like inverses, and to have sufficiently many generating and non-generating elements to make the generator questions interesting). Finally, we tested subjects for understanding of the generic case by using a cyclic group with an unspecified order n.

This design addresses a variety of concepts and aspects of understanding. The learning of each group operation corresponds to a process-level understanding of a specific group. The concepts of identities, inverses and generators, are built upon this operation. The transfer of these concepts to a cyclic group of a different order requires transfer of process-level understanding of how to find inverses, generators, etc. The subsequent questions about the generic cyclic group of order n require the ability to understand and formulate generic (and usually formal) statements about the processes and concepts learned.

Group Presentations. Each experimental group received a different presentation of the cyclic groups. We chose these to compare a presentation based on modular arithmetic (which is easily explained as a slight variation on regular arithmetic), with a more visually concrete presentation based on counting around a polygon (which allows subjects to develop a visual intuition, but which is not as directly familiar as standard arithmetic, although subjects may find analogies, e.g. to clocks). For experiments 2 and 3, we added a hybrid group, where subjects were presented with both presentations and asked to integrate them.

For the modular arithmetic presentation, we presented the group operation as $+_6$, and we explained to subjects that to compute $+_6$ you add the two numbers, and then subtract 6 if your result is 6 or larger.

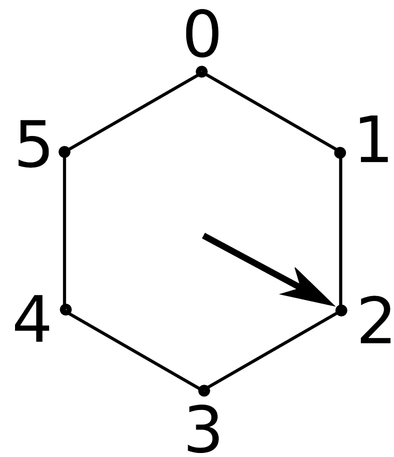


Figure 2. Order 6 polygon figure

For the polygon presentation, we presented the group operation in the form of rotating an arrow around a polygon. We wrote the group operation as ⑥, a hexagon containing the numeral 6, and provided the subjects with a diagram like figure 2. The diagram that subjects were provided was interactive, so that they could click or click and drag to move the arrow around the polygon. The arrow would "snap" to the nearest vertex when released. (The diagram for the currently relevant group order was provided on each problem in the experiment.) We explained to subjects that to compute ⑥ you point the arrow in the hexagon to the first number, and then move it the second number of spaces clockwise. The number that the arrow points at is your result. In our training presentation, we used ⑥, and gave examples such as 4⑥4 = 2, because 4 steps clockwise from 4 makes the arrow point at 2.

After seeing several examples, subjects practiced the operation on 10 problems, and if their accuracy was below 80%, they were given an additional 10 practice problems. On all of these problems, the subjects received feedback on their answers and an explanation of the correct answer.

For the hybrid group subjects, we presented both presentations, calling them respectively the "arithmetic method" and "polygon method." (We used "arithmetic method" because we felt that some subjects would find the term "modular method" to be confusing.) We alternated asking the subjects to use the polygon method and arithmetic methods on six of the initial operation practice problems, to encourage them to develop a familiarity with both presentations. The answer explanations on these questions were presented in accordance with the operation we had asked them to use; on the questions where we did not specify an operation we provided both types of feedback. Like participants in the other groups, hybrid group subjects did 10 practice problems, plus an additional 10 if their accuracy on the first 10 was below 80%.

In experiments 2 & 3, we added one additional page after presentation of the operation (but before

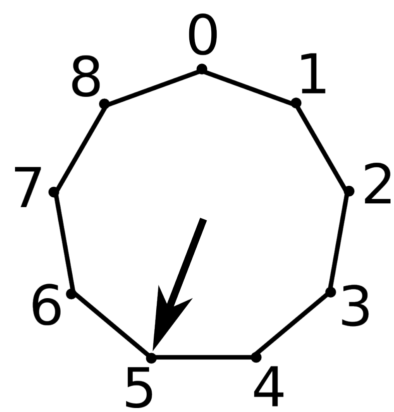


Figure 3. Order 9 polygon figure

the practice) asking the hybrid group subjects to reflect on how the different methods corresponded, and analogous questions asking the other subjects to reflect on how the operation worked.

Indentities & Inverses. Next, we explained the concept of identity by stating that 0 is the identity because when you combine it with anything, you get the same thing back. We gave two examples to illustrate this. (This, and all subsequent concepts, were explained to the different experimental groups using exactly the same text, except for the differences in the operation symbols used. For the remainder of the paper, when presenting material that both experimental groups saw, we will use either of the operation symbols.)

Similarly, we explained the concept of inverses by saying something's inverse is what you need to combine with that thing to get to the identity. For example, the inverse of 1 is 5, because 1 ⑥ $5 = 0$ and 5 ⑥ $1 = 0$. We then allowed subjects to find inverses for all other group elements as practice, and subjects received feedback on their answers and an explanation of the correct answer.

Generators. Finally, we taught the subjects the idea of generators, by explaining that a generator can make every other element of the group by combining with itself. For example, 1 is a generator under $+_6$, because $1 = 1$, $2 = +_6 1$, etc. However, 2 is not a generator under $+_6$, because $2 = 2$, $4 = 2 +_6 2$, $0 = 2 +_6 2 +_6 2$, but there is no way to make 1, 3, or 5. We then asked subjects to find whether each of the remaining elements generates the group as practice, and provided them with feedback on their answers and an explanation of the correct answer.

Transfer Test. We first tested the subjects transfer of concepts to the cyclic group of order 9, presented to the modular group as $+_9$, or to the polygon group as ⑨ with the visual aid in figure 3. We allowed the subjects one practice problem (with feedback) on the new operation, to ensure that they understood it. We then asked the subjects questions to test their knowledge of the concepts outlined in each section above, namely:

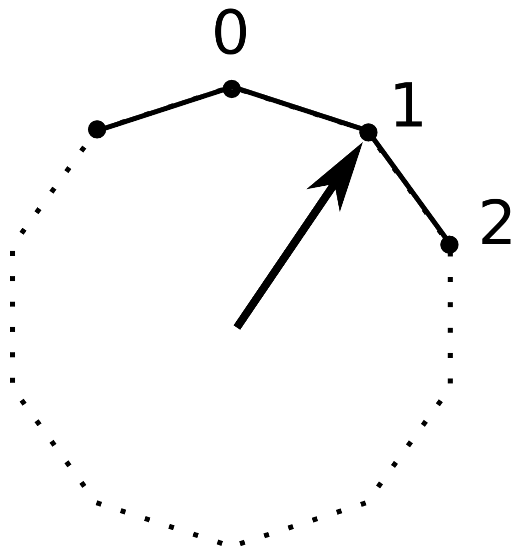


Figure 4. Order n polygon figure

- A set of seven problems with the group operation, e.g. $6 \circledcirc 4 =?$, with subjects asked to provide an explanation of their answers for two of them.
- One problem on the identity under the operation, with explanation.
- Three inverse problems for the group, with explanation for one of them.
- Four generators questions (two generators, two non-generators), with explanation for one generator and one non-generator.

Generalization Test. Finally, we told subjects we were now considering an order n cyclic group, presented to the modular group subjects as $+_n$, and to the polygon group subjects as $\circledcirc$ with the visual aid shown in figure 4. (Unlike the other visual aids, in this one the arrow would rotate freely, and would not "snap" to the vertices, to avoid implicitly indicating a specific number of vertices to subjects.) We then asked them the following questions:

- What is the identity under $+_n$?
- Two questions on giving formulas for inverses under $\circledcirc$, for 1 and for an arbitrary element x.
- Two free-response questions on which elements are generators.
- Four true/false questions on which elements are generators, successively narrowing in on a correct statement about non-generators (If an element x is not a generator under $+_n$, x must be a multiple of a divisor of n.)

- Three always/sometimes/never questions about generators. (E.g. If an element x is a generator under $\textcircled{n}$, is its inverse a generator always, sometimes, or, never?)

(The exact T/F and A/S/N questions are listed along with the presentation of the results from them in Appendix B.)

Representation-Use Questions (Experiment 3). In experiment 3, we added four questions for the hybrid subjects in which they answered a question analogous to one earlier in the experiment, and then subsequently indicated on 5-item Likert scales ("Not at all" to "Very much") for each representation the degree to which they had used it on that question. After this, they were presented a text box and asked to describe in as much detail as possible how they had used each representation in solving the question. We added one question for each of the four question types where we previously observed an effect, inverse of zero, inverse of non-zero elements, identifying generators, and answering T/F questions about generators.

Hypotheses

For experiment 1, our hypothesis was that there would be a difference in learning between the subject groups in several of the aspects of undrstanding, and a presentation that is beneficial for one concept or aspect may be deleterious for another. (We had no a priori theory to predict which concepts would be more easily learned from which presentation, so part of the purpose of experiments 2 & 3 was to verify our results.)

For experiments 2 & 3, we hypothesized that we would replicate the differences we found in our first experiment, namely:

- The modular and polygon groups would not differ significantly in their learning of the operation.
- The modular group would be significantly better than the polygon group at finding the inverse of non-zero elements.
- The polygon group would be significantly better than the modular group at finding the inverse of zero.
- The polygon group would be significantly better than the modular group at identifying elements that are generators in the specific groups.
- The modular group would be significantly better than the polygon group at answering T/F questions about generators in the order n group.

Furthermore, we hypothesized that the hybrid group would achieve approximately the maximum performance of the two groups, i.e.:

- The hybrid group would be significantly better than the modular group whenever the polygon group performed better.
- The hybrid group would not significantly differ from the modular group when the modular group performed better.

(Note that the asymmetry in the statement of the hybrid group hypotheses is due to the fact that our regressions were designed so that all comparisons were to the modular group.) This can be contrasted with other possible predictions for hybrid group performance. One possibility is that seeing both representations would simply confuse or overload the subjects, and they would perform worse on every type of question, resulting in them being significantly worse at every question type. Another possibility is that subjects would just pick one representation and use it exclusively, and perform as though they were subjects in that representation group. This, and possibilities such as subjects randomly picking a representation to use on each question, would result in patterns of data where the hybrid group appeared to perform at the average of the other two groups. (Of course, there may be individual differences, and some subjects may achieve maximal performance while others are simply confused, which could also produce a similar effect.)

Finally, for the experiment 3 questions where we had the hybrid subjects describe which representation they used, we hypothesized that where the polygon subjects performed better, using the polygon representation would be significantly predictive of success or using the modular representation would be significantly predictive of failure, and vice versa for the questions where the modular subjects performed better.

Analysis. For experiment 1, we chose to analyse the data via a mixed-effects linear regression on the question-by-question scores of the subjects, with the fixed effects being question type, including the group order (6, 9, or n) in which it was presented; representation, polygon or modular; the interaction of those two; the effect of having a high math background, defined as algebra II, trigonometry, statistics, or above; and a random effect of subject. We excluded subjects who reported in the background section that they had used modular arithmetic or mathematical groups before. The results presented are taken from this analysis. (We did not compute multiple comparisons correction in our analyses for experiment 1, we instead validated them in the subsequent experiments. These results must be interpreted with this in mind.)

For experiment 2, we used the same analysis as in experiment 1, except that we added the hybrid group, and our comparisons were specified *a priori* in accordance with the above hypotheses.

For experiment 3, we decided to alter our analyses (because we were concerned about violating the normality assumptions of the standard linear regression), and analyzed the data via a planned logistic regression on the question-by-question scores of the subjects, which we bootstrapped across 10,000 resamples of the subjects, with the predictors being as in experiments 1 and 2. We used the inclusion of zero in the percentile bootstrap 95% confidence intervals for the predictors to test the significance of our results. This analysis for experiment 3 was pre-registered on the Open Science Framework (`https://osf.io/5gthx/`). (We also retroactively ran this bootstrapped logistic regression on the data from experiments 1 & 2, in order to have a uniform set of results for our meta-analysis.) For the hypotheses about the representation-use questions, we used logistic regression predicting score on the question by the ratings of representation used.

Implementation details. We performed the experiments on Amazon's Mechanical Turk, using high-reputation subjects (over 85% approval rate), and using subject tracking (so we could run follow-up and replication studies on Mechanical Turk without having the same subjects participate and contaminate the results). Experiment 1 had $n = 50$ subjects per group, $N = 100$ total; experiment 2 had $n = 50$, $N = 150$; and experiment 3 had $n = 100$, $N = 300$. The tasks were developed using JSPsych framework with a custom plugin to integrate the interactive polygon diagrams where necessary, hosted on Stanford's servers, and embedded in the Mechanical Turk page. We made small alterations and typo fixes between experiments that we did not think would affect the results. (The only significant change was that in experiment 1 we included for half the subjects in each condition a prompt after each section to reflect on the results. This did not have any significant effect, so we collapsed across it in our experiment 1 analyses, and removed it from experiments 2 and 3.) The final versions of the experiments can be compared on our github, http://www.github.com/lampinen/fyp_cyclic_groups/.

Results

Overall, performance was quite high on the basic operation questions and declined on the questions about inverses and generators. Performance was similar across the order 6 and order 9 groups, but declined substantially in the order n group. This suggests that, while most subjects were able to transfer their procedures for solving the questions to a different group order, only some subjects were able to reason generically or express formal statements about generic cyclic groups.

The subjects in the polygon and modular groups differed significantly on a number of question types, with the polygon group consistently performing better at identifying elements that were generators and finding the inverse of zero, while the modular group performed significantly better at finding the inverse of non-zero elements. See fig. 5 for a summary of the results aggregated across experiments and group orders. However, note that this may understate the hybrid group's performance, because the hybrid group showed an improvement on a number of aspects of understanding between the order 6 and order 9

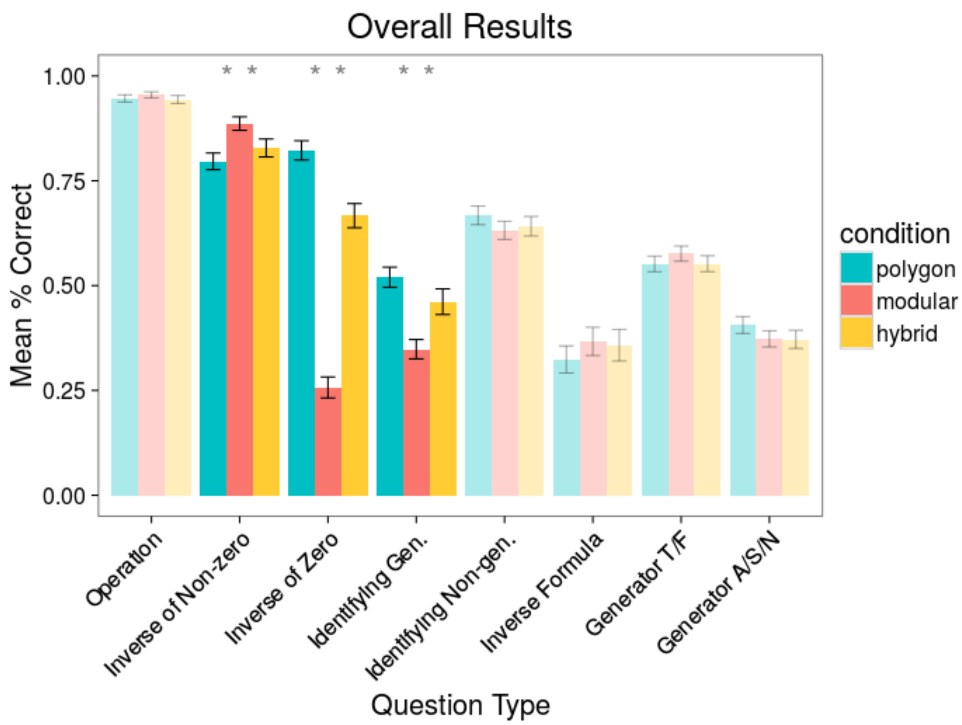


Figure 5. Results aggregated across group orders and experiments 2 & 3 (Exp. 1 had no hybrid condition, so we omitted it from this graph). Highlighted results are the main findings relevant to our hypothesis, stars mark comparisons where the meta-analysis 95% confidence interval did not overlap zero (statistics include experiment 1 data for polygon vs. modular comparisons).

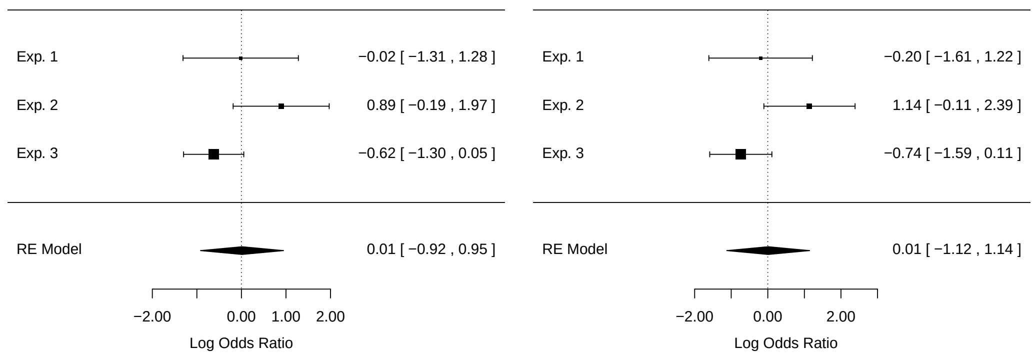
questions (as evidenced the decreased difference between the hybrid and modular conditions on inverse of non-zero element questions, the increased difference on identifying generators, etc.). Specifically, although the hybrid group appears significantly worse than the modular group at the inverse of non-zero questions aggregated across both orders, the difference was not significant on the later questions. For a full presentation of the results, see below and the supplemental figures appendix.

Full Results

Operation.

Experiment 1: There was no significant difference in the performance on the basic operation questions between the experimental groups (see Figure 22) in either the order 6 group ($\beta = 0.006$, $t(235) = 0.18$, $p = 0.86$) or the order 9 group ($\beta = -0.009$, $t(4178) = -0.26$, $p = 0.79$). These are the questions where the wording varied in accordance with the different representations taught to the experimental groups. Both groups performed quite well, with over 90% accuracy.

Experiment 2: We replicated our result that there was no significant difference in the performance



(a) Order 6 (b) Order 9

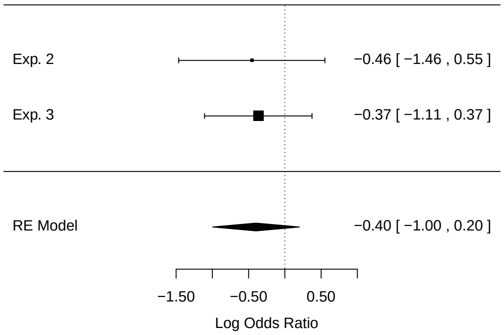
Figure 6. Meta analysis – polygon vs. modular, operation

on the basic operation questions between the modular and polygon experimental groups (see Figure 23) in either group order (order 6: $\beta = 0.02$, $t(480) = 0.735$, $p = 0.46$; order 9: $\beta = 0.02$, $t(6030) = 0.63$, $p = 0.53$). Furthermore, there was no significant difference between the hybrid and modular groups (order 6: $\beta = -0.02$, $t(449) = -0.72$, $p = 0.47$; order 9: $\beta = 0.04$, $t(6042) = 1.10$, $p = 0.27$).

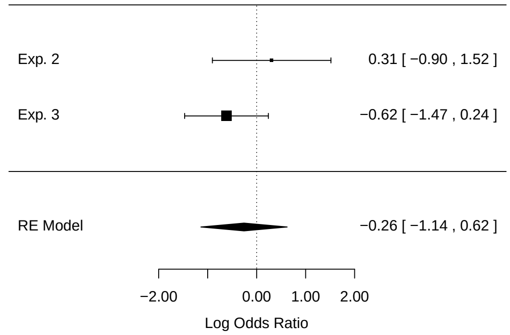
Experiment 3: We replicated our result that there was no significant difference in the performance on the basic operation questions between the modular and polygon experimental groups (see Figure 24) in either group order (order 6: β 95%-CI $= [-2.55, 0.15]$, $p > 0.05$; order 9: β 95%-CI $= [-1.59, 0.11]$, $p > 0.05$), and our result that there was no significant difference between the hybrid and modular groups (order 6: β 95%-CI $= [-2.13, 0.13]$, $p > 0.05$; order 9: β 95%-CI $= [-1.90, 0.17]$, $p > 0.05$).

Meta-Analysis: We estimate any effect of the polygon presentation on the ability to perform the group operation to be small (order 6: log OR = 0.01; order 9: log OR = 0.01; see fig. 6). We estimate any effect of the hybrid presentation on the ability to perform the group operation to be small (order 6: log OR = -0.40; order 9: log OR = -0.26; see fig. 7).

Summary: Despite learning different methods for performing the group operation, the experimental groups do not differ substantially in their ability to perform it.



(a) Order 6



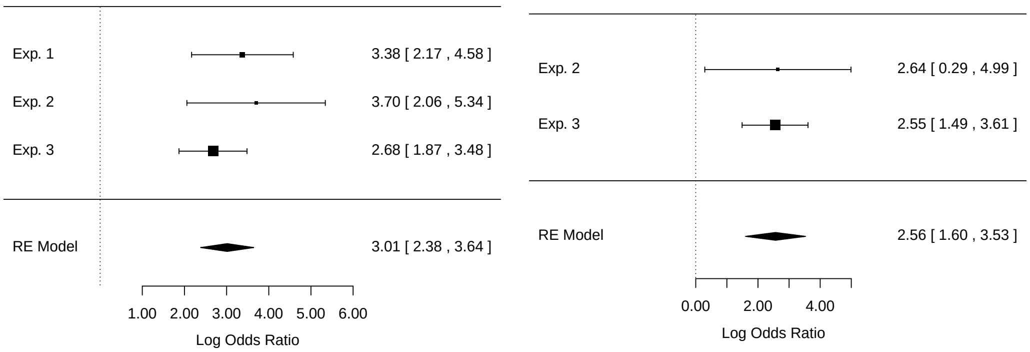
(b) Order 9

Figure 7. Meta analysis – hybrid vs. modular, operation

Inverses.

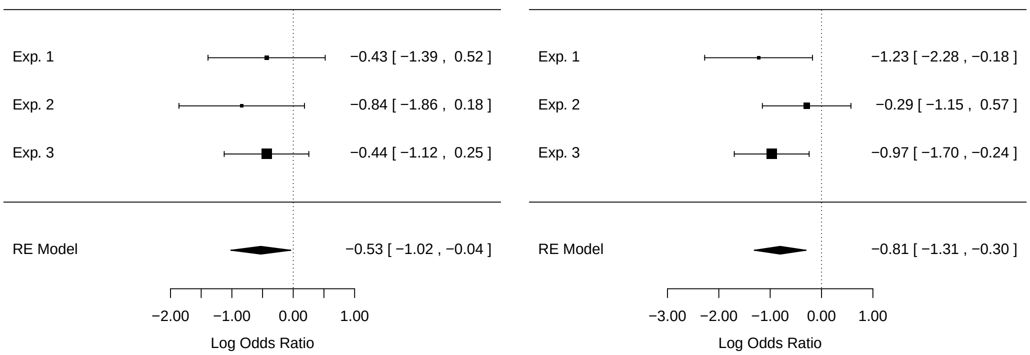
Experiment 1: For inverses the polygon group was significantly worse at finding the inverse of non-zero elements in the order 9 group ($\beta = -0.169$, $t(4178) = -3.73$, $p < 0.001$), but not significantly worse in the order 6 group ($\beta = -0.04$, $t(4178) = -1.01$, $p = 0.31$), and was significantly better at finding the inverse of zero (see Figure 25) in the order 6 group ($\beta = 0.66$, $t(4178) = 9.13$, $p < 0.001$). (Unfortunately we did not include an inverse of zero question in the order 9 group in this experiment, so we only have data from the order 6 group.) The modular group was significantly better at writing the generic formulas for the inverse of 1 and an arbitrary element x in the group of order n ($\beta = -0.11$, $t(4178) = -1.99$, $p = 0.047$).

Experiment 2: We replicated our results that the modular group performed better at inverses of non-zero elements (order 6: $\beta = -0.11$, $t(6026) = -2.59$, $p = 0.009$; order 9: $\beta = -0.07$, $t(6026) = -1.48$, $p = 0.13$), and the polygon group performed better at inverse of zero questions (order 6: $\beta = 0.68$, $t(6024) = 8.842$, $p < 0.001$; order 9: $\beta = 0.44$, $t(6024) = 5.80$, $p < 0.001$). (Note that the modular group performed worse than the polygon group even in the order 9 group, once they had already seen an example of an inverse of zero with feedback.) The polygon group did not perform significantly differently from the modular group at the inverse formula questions ($\beta = -0.05$, $t(6025) = -0.94$, $p = 0.35$). The hybrid group did not perform significantly worse than the modular group at inverses of non-zero elements (order 6: $\beta = -0.007$, $t(6033) = -0.173$, $p = 0.86$; order 9: $\beta = 0.08$, $t(6031) = 1.56$, $p = 0.12$), and did perform significantly better at inverses of zero (order 6: $\beta = 0.46$, $t(6027) = 5.95$, $p < 0.001$; order 9: $\beta = 0.38$, $t(6027) = 4.91$, $p < 0.001$). However, it is clear that the hybrid group is not initially performing as well as the polygon group on this question, so there is still room for improvement. The hybrid group did not perform significantly differently from the modular group at the inverse formula questions ($\beta = 0.01$, $t(6029) = 0.125$, $p = 0.90$). (See Figure 26.)



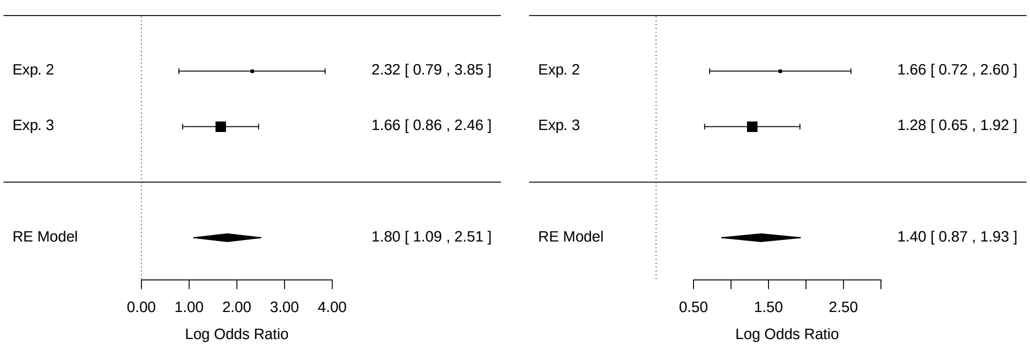
(a) Order 6 (b) Order 9

Figure 8. Meta analysis – polygon vs. modular, inverse of zero



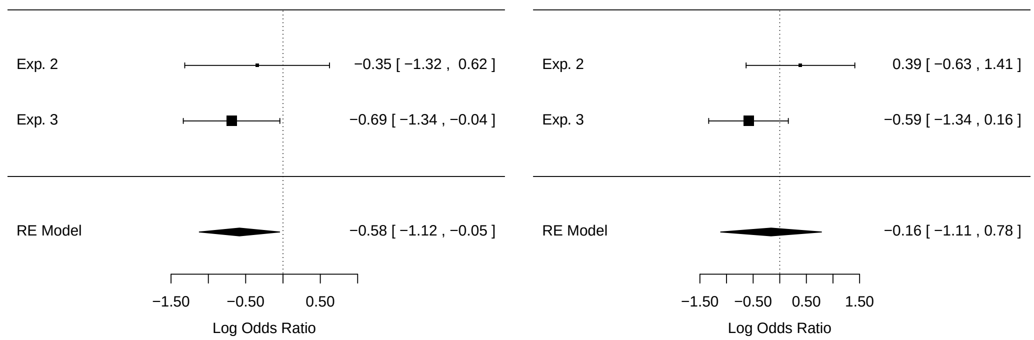
(a) Order 6 (b) Order 9

Figure 9. Meta analysis – polygon vs. modular, inverse of non-zero



(a) Order 6 (b) Order 9

Figure 10. Meta analysis – hybrid vs. modular, inverse of zero

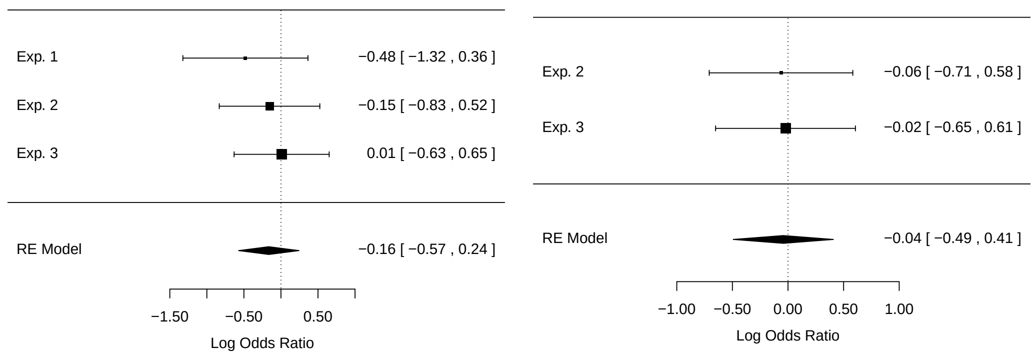


(a) Order 6 (b) Order 9

Figure 11. Meta analysis – hybrid vs. modular, inverse of non-zero

Experiment 3: We replicated our results that the modular group performed better than the polygon group at inverses of non-zero elements only in the order 9 group, although the effect was in the correct direction in the order 6 group (order 6: β 95%-CI $= [-1.14, 0.24]$, $p > 0.05$; order 9: β 95%-CI $= [-1.74, -0.27]$, $p < 0.05$). We replicated our result that the polygon group performed better at inverse of zero questions (order 6: β 95%-CI $= [1.93, 3.55]$, $p < 0.05$; order 9: β 95%-CI $= [1.72, 3.60]$, $p < 0.05$). The polygon group did not perform significantly differently from the modular group at the inverse formula questions (β 95%-CI $= [-0.64, 0.65]$, $p > 0.05$). The hybrid group was not significantly worse than the modular group at inverses of non-zero elements in either group, although it was trending worse in the order 6 group (order 6: β 95%-CI $= [-1.92, 0.01]$, $p > 0.05$; order 9: β 95%-CI $= [-1.96, 0.24]$, $p > 0.05$). We replicated our result that the hybrid group was significantly better at inverses of zero only in the order 6 group, although the effect was trending in the order 9 group (order 6: β 95%-CI $= [0.21, 2.65]$, $p < 0.05$; order 9: β 95%-CI $= [-0.01, 2.09]$, $p > 0.05$).The polygon group did not perform significantly differently from the modular group at the inverse formula questions (β 95%-CI $= [-1.32, 0.78]$, $p > 0.05$) (See Figure 27.)

Meta-Analysis: We estimated the positive effect of the polygon condition on inverse of zero questions to be large for both group orders, although the effect is smaller for order 9, consistent with some learning in the modular group (order 6: log OR = 3.01; order 9: log OR = 2.56; see fig. 8). We estimated the negative effect of the polygon condition on inverse of non-zero questions to be moderately sized (order 6: log OR = -0.53; order 9: log OR = -0.81; see fig. 9). We estimated the positive effect of the hybrid condition on inverse of zero questions to be large for both group orders, although it is not as large as that of the polygon condition, and the effect is smaller for order 9, consistent with some learning in the modular group (order 6: log OR = 1.80; order 9: log OR = 1.40; see fig. 10). We estimated the negative effect of the hybrid condition on inverse of non-zero questions to be moderately sized and possibly vanishing after



(a) Polygon vs. modular

(b) Hybrid vs. modular

Figure 12. Meta analysis – inverse formula

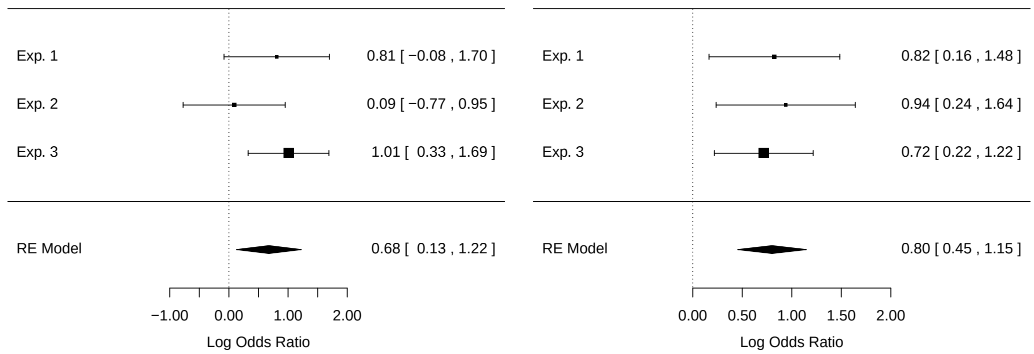
further practice in the order 9 group (order 6: log OR = -0.58; order 9: log OR = -0.16; see fig. 11). We estimated the effect of the polygon condition on the inverse formula questions to be neglible (log OR = -0.16, see fig. 12), and similarly for the effect of the hybrid condition on the inverse formula questions (log OR = -0.04, see fig. 12).

Summary: Overall, it seems that the modular presentation is generally beneficial for finding inverses, except in the case of zero, where the polygon presentation subjects perform much better. (See discussion for a possible explanation of this result.)

Generators.

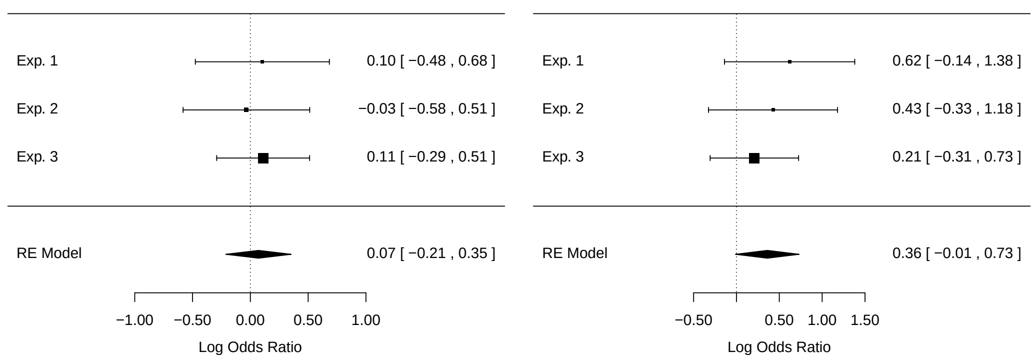
Experiment 1: The polygon group performed significantly better at identifying elements that were generators (order 6: $\beta = 0.164$, $t(4178) = 2.27$, $p = 0.02$; order 9: $\beta = 0.184$, $t(4178) = 3.45$, $p < 0.001$). The groups did not differ significantly at identifying non-generators in the order 6 group, but the polygon group performed significantly better in the order 9 group (order 6: $\beta = 0.01$, $t(4178) = 0.235$, $p = 0.81$; order 9: $\beta = 0.12$, $t(4178) = 2.32$, $p = 0.02$). The modular group performed better at the T/F questions about generators in the order n group ($\beta = -0.11$, $t(4178) = -2.60$, $p = 0.009$). However, performance on these questions was not too far above chance. The groups did not differ significantly on the A/S/N questions ($\beta = -0.03$, $t(4178) = -0.5$, $p = 0.62$). (See Figure 28.)

Experiment 2: We replicated our result that the polygon group is better at identifying generators only in the order 9 group (order 6: $\beta = 0.00$, $t(6024) = -0.05$, $p = 1.00$; order 9: $\beta = 0.19$, $t(6025) = 3.37$, $p < 0.001$), and the polygon group did not perform better at identifying non-generators in either group (order 6: $\beta = -0.03$, $t(6026) = -0.60$, $p = 0.55$; order 9: $\beta = 0.07$, $t(6025) = 1.31$, $p = 0.19$). We replicated our result that the modular group performed better at T/F questions about generators ($\beta = -0.12$, $t(6026) = -2.71$, $p = 0.007$). We replicated our result that the polygon and modular groups did not



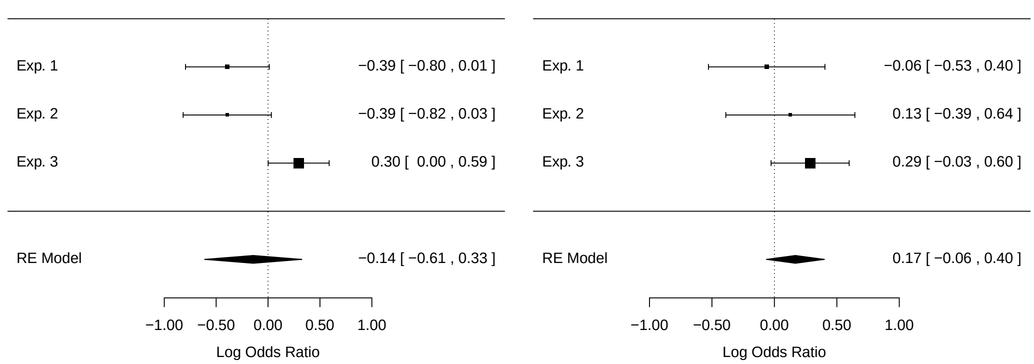
(a) Order 6 (b) Order 9

Figure 13. Meta analysis – polygon vs. modular, identifying generators



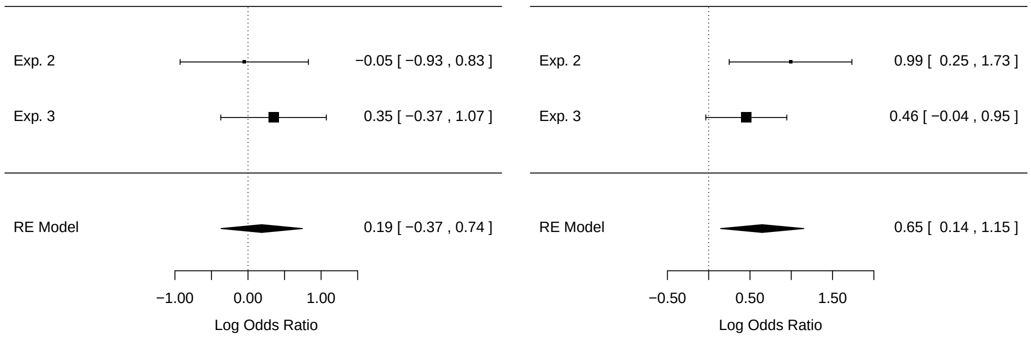
(a) Order 6 (b) Order 9

Figure 14. Meta analysis – polygon vs. modular, identifying non-generators



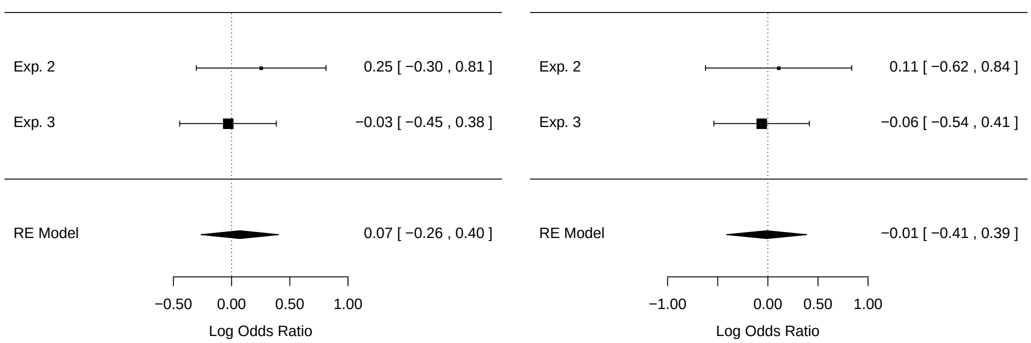
(a) T/F (b) A/S/N

Figure 15. Meta analysis – polygon vs. modular, generator t/f & a/s/n



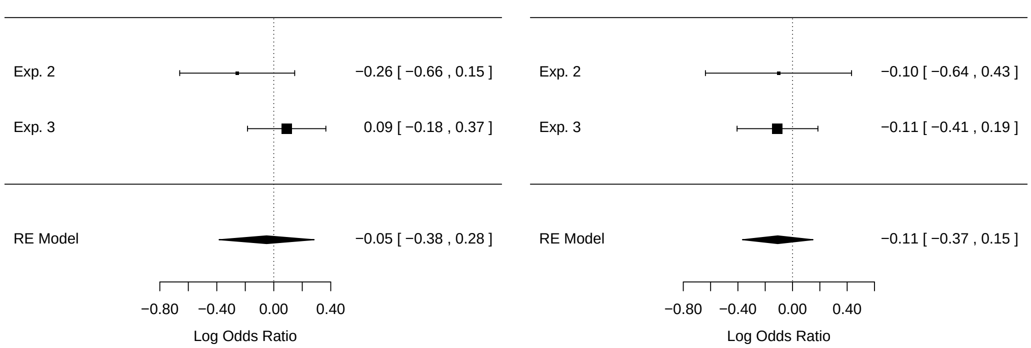
(a) Order 6 (b) Order 9

Figure 16. Meta analysis – hybrid vs. modular, identifying generators



(a) Order 6 (b) Order 9

Figure 17. Meta analysis – hybrid vs. modular, identifying non-generators



(a) T/F (b) A/S/N

Figure 18. Meta analysis – hybrid vs. modular, generator T/F & A/S/N

significantly differ at the A/S/N questions ($\beta = 0.01$, $t(6026) = 0.139$, $p = 0.89$). The hybrid group perfomed significantly better than the modular group at identifying generators in the group order where our previous finding replicated ($\beta = 0.25$, $t(6029) = 4.31$, $p < 0.001$), did not significantly differ at identify non-generators (order 6: $\beta = 0.07$, $t(6031) = 1.51$, $p = 0.13$; order 9: $\beta = 0.05$, $t(6029) = 0.80$, $p = 0.42$), and did not perform significantly worse at the T/F questions ($\beta = -0.042$, $t(6033) = -0.98$, $p = 0.33$). However, it appears that the hybrid group did not be achieve performance completely on par with the modular group on the T/F questions. The hybrid and modular groups did not differ significantly on the A/S/N questions ($\beta = -0.00$, $t(6031) = -0.01$, $p = 0.99$). (See Figure 30.)

Experiment 3: We replicated our result that the polygon group is better at identifying generators (order 6: β 95%-CI $= [0.35, 1.72]$, $p < 0.05$; order 9: β 95%-CI $= [0.22, 1.22]$, $p < 0.05$), and that the polygon group did not significantly differ from the modular at identifying non-generators (order 6: β 95%-CI $= [-0.29, 0.50]$, $p > 0.05$; order 9: β 95%-CI $= [-0.32, 0.73]$, $p > 0.05$). We failed to replicate our result that the modular group is better than the polygon group at T/F questions about generators, in fact the polygon group performed significantly better in this experiment (β 95%-CI $= [0.004, 0.59]$, $p < 0.05$). We replicated our result that the polygon and modular groups did not differ significantly at answering A/S/N questions (β 95%-CI $= [-0.03, 0.60]$, $p > 0.05$). We failed to replicate our result that the hybrid group is significantly better than the modular group at identifying generators in either group order, although both effects were slightly in that direction (order 6: β 95%-CI $= [-1.11, 1.27]$, $p > 0.05$; order 9: β 95%-CI $= [-0.78, 1.17]$, $p > 0.05$). We replicated our result that the hybrid group did not significantly differ at identifying non-generators (order 6: β 95%-CI $= [-1.17, 0.60]$, $p > 0.05$; order 9: β 95%-CI $= [-1.25, 0.61]$, $p > 0.05$). We replicated our result that the modular and hybrid groups did not significantly differ at the T/F questions (β 95%-CI $= [-1.05, 0.71]$, $p > 0.05$). We replicated our result that the modular and hybrid groups did not significantly differ at the A/S/N questions (β 95%-CI $= [-1.26, 0.53]$, $p > 0.05$). (See Figure 32.)

Meta-Analysis: We estimated the positive effect of the polygon condition on identifying generators to be moderately sized (order 6: log OR = 0.68; order 9: log OR = 0.80; see fig. 13). We estimated the effect of the polygon condition on identifying non-generators to be small, but maybe slightly positive in the group of order 9 (order 6: log OR = 0.07; order 9: log OR = 0.36; see fig. 14). We estimated that any effect of the polygon condition on answering True/False questions about generators is small (log OR = -0.14; see fig. 15). We estimated that any effect of the polygon condition on answering Always/Sometimes/Never questions about generators is small (log OR = 0.17; see fig. 15). We estimated the positive effect of the hybrid condition on identifying generators to be small in the order 6 group, but increasing in the order 9 group, (order 6: log OR = 0.19; order 9: log OR = 0.65; see fig. 16). We estimated the effect of the hybrid condition on identifying non-generators to be small (order 6: log OR = 0.07; order 9: log OR = -0.01; see fig. 17). We estimated the effect of the hybrid condition on answering True/False questions about generators to be small (log OR = -0.05; see fig. 18). We estimated the effect of the hybrid condition on answering Always/Sometimes/Never questions about generators to be small (log OR = -0.11; see fig. 18).

Summary: Overall, it seems that the polygon presentation is beneficial for identifying generators, and the hybrid presentation seems to be similarly beneficial for identifying generators in the order 9 group, once they have had some practice. None of the presentations seem particularly beneficial for identifying non-generators, or for answering T/F or A/S/N questions about generators, performance was quite low on these questions.

Other Analyses

We conducted several other analyses to further elucidate the differences in performance between the groups, and the cognitive factors underlying them.

Diagram use. We hypothesized that the polygon group's superior performance on identifying generators might be due to the ability to use the spatial structure of the polygon to more easily visualize the elements generated by an element (see discussion). One possible prediction of this hypothesis would be that within the polygon group, interaction with the diagram might be predictive of success on these questions. (Of course, we could only record the interactions with the mouse, while many subjects may have just gazed or pointed at the diagram to use it in their thinking. Furthermore, the use of the diagram may be confounded with overall engagement. Our results must be interpreted with these qualifications in mind.)

We performed a mixed-model logistic regression on data from the polygon and hybrid subjects from Experiments 2 and 3, predicting correct answers by whether or not they used the diagram (and a random effect of subject). We found that using the diagram was significantly predictive of success on the questions (Exp. 2: $\beta = 1.90$, $z = 2.76$, $p = 0.006$; Exp. 3: $\beta = 2.24$, $z = 5.12$, $p < 0.001$). Furthermore, this effect was present even when controlling for reaction time (Exp. 2: $\beta = 1.62$, $z = 2.26$, $p = 0.024$; Exp. 3: $\beta = 1.64$, $z = 3.51$, $p < 0.001$), which might suggest that engagement alone wasn't the driving factor, and the effect was significant or trending within the polygon and hybrid conditions individually, suggesting that both benefitted.

Using analogous mixed-model logistic regressions across the full data from the hybrid and polygon groups, we found that on all questions in the experiment (not just generator questions) that using the diagram was significantly predictive of success (Exp. 2: $\beta = 1.26$, $z = 6.93$, $p < 0.001$; Exp. 3: $\beta = 1.19$, $z = 10.28$, $p < 0.001$), even when controlling for reaction time (Exp. 2: $\beta = 1.42$, $z = 7.53$, $p < 0.001$; Exp. 3: $\beta = 1.25$, $z = 10.71$, $p < 0.001$). However, the estimated effect sizes were smaller than for the generator

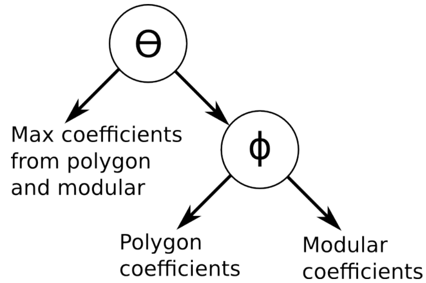


Figure 19. Hierarchical model

questions. This suggests that the diagram may have been especially helpful on these generator questions, as we hypothesized.

Hierarchical modeling of hybrid subjects. Although we found the hybrid group did perform better than either the polygon or modular group individually, it did not seem to achieve truly best-of-both-worlds performance. One explanation for this might be that some subjects were just picking one representation and using it consistently, while others were really receiving the benefits of both and performing optimally (at the max level of the two). We attempted to model this with a hierarchical model which assumed that the data were generated by the following process (depicted schematically in fig. 19):

1. With probability θ, the subject would benefit from both representations, and would perform optimally in the sense that their data would be best fit by assuming they picked the optimal representation on each question (or equivalently, that their regression coefficients were the element-wise maximum of the regression coefficients of the two other groups).

2. If the subjects did not benefit from both representations (probability $1-\theta$), they would pick the polygon representation with probability ϕ, and the modular representation with probability $1-\phi$, and use it for the entire experiment, thus their data would be best fit by the coefficients for the respective group.

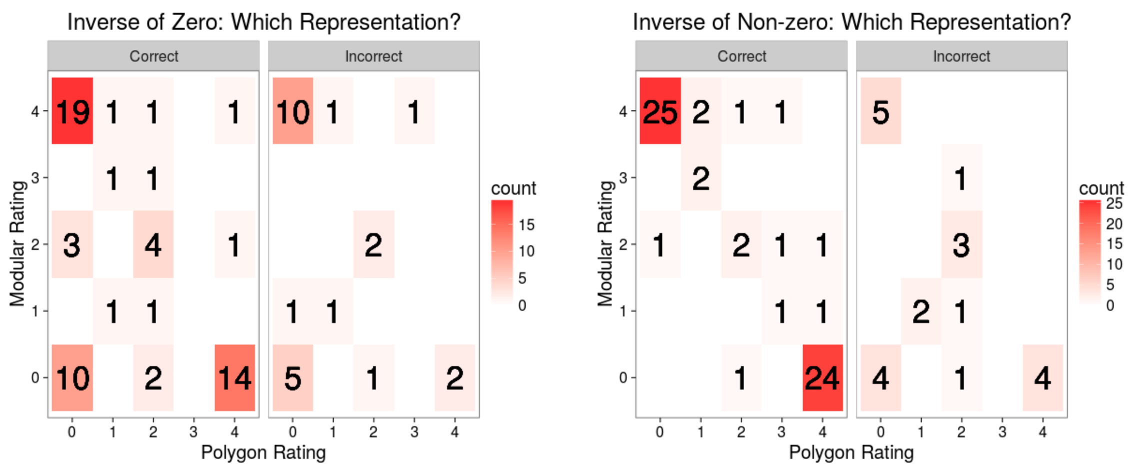
We used maximum likelihood to fit this model to the experiment 2 data, and estimated that $\theta = 0.41, \phi = 0.49$ (log-likelihood $LL = -822.4$), so the data are best fit under this model by assuming that about 40% the subjects are integrating their knowledge, and those that aren't are choosing the modular representation and polygon almost equally. We used the Aikake Information Criterion (AIC) to compare this model ($AIC = 1648.8$) to models where all subjects chose modular ($AIC = 1829.0$), all chose polygon ($AIC = 1720.5$), where no subjects integrated i.e. a fixed $\theta = 0$ and fit $\phi = 0.56$ ($AIC = 1679.2$), and a model where all subjects integrated i.e. $\theta = 1.0$ ($AIC = 1689.9$). The full model is significantly better than any of these comparison models. However, there are many other possible ways people could use the two representations (such as picking arbitrarily on each question), so further investigation is needed.

Similarly, with the experiment 3 data we estimated that $\theta = 0.39, \phi = 0.56$ (log-likelihood $LL = -1758.4$), so the data are best fit under this model by assuming that a little less than 40% of the subjects are integrating, and those that aren't are choosing the polygon representation slightly more frequently than the modular. We used the Aikake Information Criterion (AIC) to compare this model ($AIC = 3520.8$) to models where all subjects chose modular ($AIC = 3826.5$), all chose polygon ($AIC = 3734.6$), where no subjects integrated, i.e. a fixed $\theta = 0$ and fit $\phi = 0.56$ ($AIC = 3582.2$), and a model where all subjects integrated, i.e. $\theta = 1.0$ ($AIC = 3700.2$). The full model is again significantly better than any of these comparison models. However, as above there are other possible ways that the subjects could use both representations, so there remain questions to be answered. Still, the consistent estimates of about 40% integration suggest that the hybrid group is increasing the understanding of some subjects.

After noting the increase in hybrid group performance between the order 6 and order 9 groups, we decided to run these models again on the subsets of the data from order 6 and order 9 separately to see if we saw evidence of greater integration later on. Indeed, in exeriment 2 we estimated the proportion integrating in the order 6 section to be $\theta_6 = 0.30$, and the proportion integrating by the order 9 section to be $\theta_9 = 0.58$. If we compare this model where the subjects follow their order 6 best model (of polygon, modular, and the max of the two) for the order 6 questions and their order 9 best model for the order 9 and n questions, we find it improves substantially on the best earlier model ($AIC = 1432.1$). In experiment 3, we estimated $\theta_6 = 0.22$, whereas $\theta_9 = 0.50$. As above, this substantially improves on the earlier model ($AIC = 3191.5$). This corroborates the idea that integration may have increased as the experiment went on, with only about 20-30% of subjects appearing to integrate early on, but 50-60% integrating by the time they reached the order 9 material.

Representation-Use Question Results. For the experiment 3 representation-use questions, we performed logistic regressions predicting score on each representation-use question by the ratings ("Not at all" - "Very much", 5 point Likert scale) of representation used. We found that neither modular nor polygon rating was significantly predictive of success on the inverse of zero questions ($\beta_{mod} = -0.02$, $z = -0.15$, $p = 0.88$; $\beta_{poly} = 0.25$, $z = 1.37$, $p = 0.17$). We suspect this may have been due in part to the fact that this was the third presentation of an inverse of zero question, so subjects may have simply recalled the answer. Performance was very high in the hybrid group overall on this question, the majority (71%) of the subjects got the question right in the representation-use section, so it had the only positive intercept of any of the representation-use regressions.

Intriguingly, we found that both modular and polygon rating were significantly predictive of success



(a) Inverse of zero

(b) Inverse of non-zero

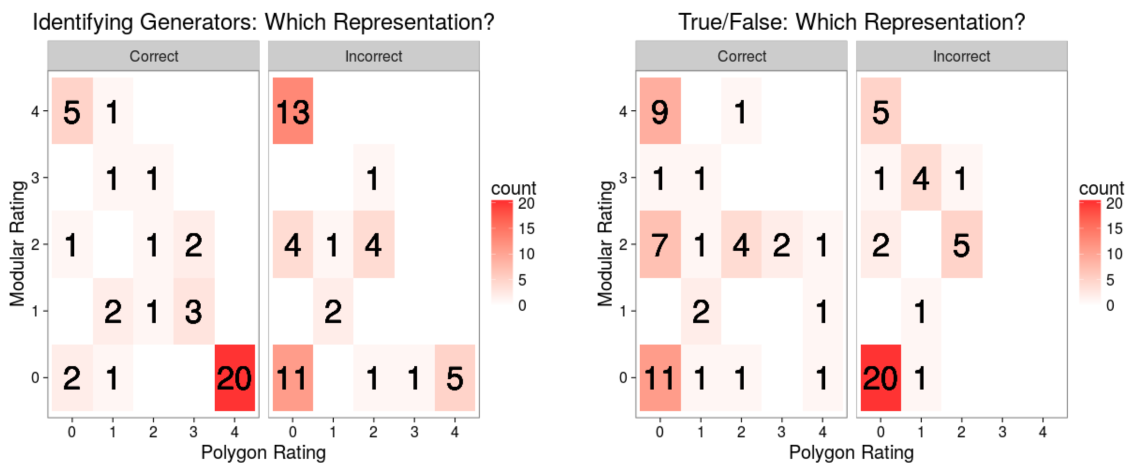
Figure 20. Experiment 3 – representation-use responses on inverse questions (counts of subjects giving each rating, split by whether answer was correct)

on inverse of non-zero questions ($\beta_{mod} = 1.09$, $z = 2.95$, $p = 0.003$; $\beta_{poly} = 1.11$, $z = 2.94$, $p = 0.003$). This, together with the previous finding, may suggest some integration occuring in the hybrid condition, such that the advantages of each representation are to some extent shared even when the other representation is used. We found that subjects polygon rating, but not modular, was significantly predictive of success on identifying generators ($\beta_{mod} = 0.19$, $z = 0.95$, $p = 0.34$; $\beta_{poly} = 0.75$, $z = 3.86$, $p < 0.001$). This corroborates our other data supporting the superiority of the polygon representation for these question, but suggests (as much of our earlier data did) that the integration in the hybrid condition is far from complete. We found that neither rating was significantly preditive of success on the generator True/False questions ($\beta_{mod} = 0.19$, $z = 1.28$, $p = 0.20$; $\beta_{poly} = 0.31$, $z = 1.38$, $p = 0.17$). This is unsurprising, since we did not observe any significant differences between the polygon and modular groups on these questions in the third experiment.

Discussion

Polygon vs. Modular Representations

Despite the fact that the polygon and modular groups did not significantly differ at learning the initial operation, they did differ in their ability to understand the subsequent concepts built upon it. Furthermore, one representation was not generally "better" than the other; they both had strengths and weaknesses. The polygon group performed better at identifying generators and finding the inverse of zero, the modular group performed better at finding the inverse of non-zero elements. Thus our initial



(a) Identifying generators (b) Generator T/F

Figure 21. Experiment 3 – representation-use responses on generator questions (counts of subjects giving each rating, split by score)

hypothesis that there would be differences in performance between the groups was confirmed.

More formally, how do the presentations measure up to the questions we proposed in the introduction?

- Does they allow students to understand that particular instance of the concept? Yes, both the polygon and modular presentations produced excellent performance on the operation in the order 6 group.

- Does it allow them to understand secondary concepts explained in that instance? Yes, both the polygon and modular presentations produced relatively high performance on certain secondary concepts, but they each had advantages and disadvantages for particular concepts.

- Does it allow them to transfer this knowledge to new, non-isomorphic instances? Yes, both groups were able to transfer to a cyclic group of order 9, and their relative advantages and disadvantages carried over.

- Does it allow them to generalize about a class of instances? Does it allow them to express and understand these generalizations using formal mathematical expressions and language? Both groups had fairly low success on these portions of the experiment, and their did not appear to be many differences between the groups on these questions.

Process Differences Underlying the Performance Differences. Although we have not fully explored these factors, we do have some hypotheses for the pattern of results we observed, based on responses to problems where we asked the subjects to explain their answers, and our post-hoc analyses of things like explicit use of the diagram:

Inverses: It appears that the modular representation may have cued the subjects to recognize an algorithm for finding most of the inverses, namely that the inverse of x can be found by subtracting x from the group order. For example, under $+_6$, the inverse of 2 is 4, and $6 - 2 = 4$. This would not be as obvious for the polygon group, which mostly found inverses by counting, which is less reliable. However, the algorithm breaks down when $x = 0$, because the inverse of 0 is 0, not the group order. We suspect that subjects in the polygon group made more mistakes counting than the modular group on inverses of non-zero elements, but the modular group subjects were misled by this algorithm when computing the inverse of zero.

Generators: There is a spatial structure to the generator questions in the polygon case which may assist in solving them, for example, when finding if 5 is a generator on the nonagon, we get the sequence $5 \to 1 \to 4 \to 2 \to \cdots$. It might be more clear to someone seeing the polygon how precisely this sequence would fill in the gaps to generate all the numbers. This might even become apparent to some participants without stepping through all of the cases; the fact that the pattern would eventually step through all the cases may be apparent after a few steps. This hypothesis is corroborated by our post-hoc analysis demonstrating that diagram use was predictive of success on these questions (more so than in the experiment overall). By contrast, for the formal questions in the order n case, it is difficult or impossible to use these spatial intuitions, which would explain why the polygon group's advantage disappears.

Hybrid Group

The results from our meta-analysis suggest that by the time they reached the order 9 group, some of the hybrid group subjects were able to to achieve approximately "best of both worlds" performance, in the sense that they outperformed the modular group where the polygon group was superior, and performed comparably to the modular group where it was superior. However, they did not appear to be achieving the full advantages of each group. Our hierarchical modeling results suggest that this imperfect performance may be explained by some individual variation, with some subjects picking just one representation, whil others achieved the benefits of both. The results of this analysis also give the promising suggestion that the number of subjects who achieved the benefits of both presentations was increasing quite substantially over time (from 20-30% in the order 6 portion of the experiment up to 50-60% in the order 9 portion). Thus, given enough practice with it, the hybrid representation might be beneficial for most subjects.

In our analysis of the representation-use question responses, we also observed some interesting heterogeneity in the effects of representation on performance. On some question types, such as finding the inverse of non-zero elements, it appears that the most hybrid group subjects have integrated their understanding of both representations sufficiently so that using either representation is equally beneficial. On others, such as identifying generators, one representation is still much more beneficial than the other for the hybrid group. This may reflect the underlying nature of the knowledge we think is being used for each of these scenarios. We hypothesized above that the modular formulation cues a process for finding the inverse of non-zero elements based on subtraction, which might be accessible using either representation. However, we posited that the advantages of the polygon group on the identifying generator questions were based on a visuospatial reasoning process, which could not as easily be transferred to the modular representation.

It remains a question for future research how the strategy could be altered to encourage more uniform improvement across all question types and subjects. Nevertheless, these results at least begin to suggest that teaching multiple presentations may be beneficial to students overall understanding.

Polygon Presentation as Intrinsically Hybrid. Despite the fact that it still has some disadvantages, between the polygon and modular presentation, the polygon representation seems overall more advantageous. It is possible that this is due to the hybrid elements inherent in the polygon presentation in our design. By including both the visuospatial representation and numbers as symbols, it may cue subjects to recognize some of the arithmetical patterns that are more explicitly explained in the modular presentation. This might explain why the polygon presentation did not perform too much worse than the modular presentation even when thinking in an arithmetic way seemed more useful, and why the hybrid presentation did not improve substantially on the polygon presentaton in these aspects of understanding.

Formalization & Generalization

None of the presentations seemed to encourage formal or general understanding particularly well, as evidenced by the low overall performance on the order n questions. (E.g., despite the fact that performance on the inverse questions was around 75% on average, only about one third of the subjects were able to articulate the formula for computing an inverse in a generic cyclic group.) This may be representative of the more general separation between these aspects of understanding, or it may simply reflect the fact that this study was too short for subjects understanding of specific groups to develop into a more general understanding.

It is very likely that more explicit practice with formalizing concepts within the experiment would lead to better performance on these types of questions, but are there more implicit aspects of the presentations that could be manipulated to encourage formalization? For example, attaching labels to concepts like the group order might better prepare subjects to think of them as variables (as in the generic order n group case). Perhaps using more formulas in the presentation of the operation, rather than the procedural description we gave, would help subjects to produce formulas on their own later on. There is already some work addressing formalization and ways to encourage it, e.g (Nathan, 2012). However, there is ample room for further development, and for research which examines how formalizations interact with presentations.

One question such research will have to confront more closely is the relationship between formalization and explicit general understanding. Although in this study these factors were confounded in many of the order n questions, it is possible to disentangle them. For example, we might ask subjects to formalize their understanding of inverses in the cyclic group of order 6, and then later ask subjects to give a generic explanation in words of how to find inverses in a general cyclic group of arbitrary order, before asking them to unite generalization and formalization in a single formula for the inverse of an element in an arbitrary cyclic group. (Indeed, (Nathan, 2012) suggests that plain language descriptions may be very beneficial in encouraging understanding of more formal representations of an idea.) Further research should explore the relationship between these different forms of abstraction, and how presentations may affect each of them.

Time & Practice

As alluded to above, there is at least one other important pedagogical element lacking in this study (and the work by Kaminski and others): time and repeated practice. Because the progression from concept introduction to final assessment of understanding occurs in about an hour, we may be short-changing the presentations by not allowing the subjects sufficient practice to develop a sufficiently elaborated understanding. Indeed the hybrid group seemed to achieve much better performance by the order 9 section of the experiment than earlier on, and its possible that the hybrid group would continue to improve faster than the other groups with further practice. In an abstract algebra class, these concepts would probably be encountered repeatedly across the course of a semester, and the students would only have a thorough understanding of them at the end. (In addition, students taking such a course would have greater mathematical literacy and a set of relevant examples to build upon.)

It is interesting to ask whether simple practice with a concept can lead to formalization, and if so, under what circumstances. Do students acquire sudden insights after computing inverses for some time, such that they can articulate the formula for the inverse? How does this depend on prior mathematical experience? How does it depend on the presentation of the concepts in question? Can hybrid presentations help by encouraging more abstract thought about the concept? These questions provide a fascinating direction for future research.

Conclusion

We explored the way presentation of concepts in math instruction affects understanding of the concept being exemplified, and of concepts related to it, using elementary group theory as our test domain. We found that even if two presentations produce equal performance on a basic concept, they can produce differential understanding of related concepts. Furthermore, it does not appear that there is always a clear advantage of one presentation over another, instead a presentation may be more useful for some concepts or some aspects of reasoning, and less useful for others. We think that these findings provide a contribution to the discussion of presentation in math cognition, illustrating that presentations have different advantages and disadvantages. This suggests that the pursuit of a single best type of presentation may be futile.

Instead, we have identified an alternative strategy for improving performance: teaching with multiple presentations while encouraging subjects to use both of them. Preliminary data suggests that this may have positive effects even if the total instruction time is approximately the same. By the end of the experiment, subjects in our hybrid condition appeared to have achieved a more complete understanding, and to perform better overall than those instructed with one presentation alone. However, it remains to be seen whether they could truly achieve best-of-both-worlds performance over a longer time period, and whether this or another approach could encourage subjects to generalize and formalize their understanding. These questions provide an exciting new direction for research in both math cognition and math pedagogy.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-114747.

References

De Bock, D., Deprez, J., Van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Abstract or Concrete Examples in Learning Mathematics? A Replication and Elaboration of Kaminskli, Sloutsky, and Heckler's Study. *Journal for Research in Mathematics Education*, *42*(2), 109. Retrieved from `http://www.staff.science.uu.nl/{~}savel101/as-sec410/literature/Bock2011.pdf`

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, *87*(6), 477–531. Retrieved from `http://psycnet.apa.org/journals/rev/87/6/477.pdf` doi: 10.1037/0033-295x.87.6.477

Hazzan, O. (1999). Reducing Abstraction Level When Learning Abstract Algebra Concepts. *Educational Studies in Mathematics*, *40*, 71–90. Retrieved from `http://download.springer.com/static/pdf/400/art:10.1023/A:1003780613628.pdf?originUrl=http://link.springer.com/article/10.1023/A:1003780613628{\&}token2=exp=1444342117{~}acl=/static/pdf/400/art:10.1023/A:1003780613628.pdf?originUrl=http:/`

Jones, M. G. (2009). Transfer, Abstraction, and Context. *Journal for Research in Mathematics Education*, *40*(2), 80–89.

Kaminski, J. a., Sloutsky, V. M., & Heckler, A. F. (2008). The Advantage of Abstract Examples in Learning Math. *Science*, *320*(5875), 454–455. Retrieved from `http://www.sciencemag.org/content/320/5875/454{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/18436760` doi: 10.1126/science.1154659

Kaminski, J. a., Sloutsky, V. M., & Heckler, A. F. (2009). Concrete Instantiations of Mathematics: A Double-Edged Sword. *Journal for Research in Mathematics Education*, *40*, 90–93. Retrieved from `http://www.jstor.org/stable/10.1086/431026`

Lang, S. (2002). *Algebra (revised third edition)* (Rev. 3rd ed.). Springer Graduate Texts in Mathematics.

Nathan, M. J. (2012). Rethinking Formalisms in Formal Education. *Educational Psychologist*, *47*(2), 125–148. Retrieved from `http://2015.laschool4education.com/wp-content/uploads/2014/10/2012{\_}Nathan-EP-Rethinking-Formalisms-FF.pdf` doi: 10.1080/00461520.2012.667063

Appendix A: A Brief, Selective Introduction to Group Theory

Groups are mathematical structures that provide us with a nice way of doing something like arithmetic with objects besides the ordinary numbers, like symmetries of an object or permutations, or with smaller sets of ordinary numbers (as in the experiments presented in this paper). They have applications throughout mathematics, physics, chemistry, and computer science. Here I present the formal definition of a group with informal intuitions in italics. A **group** consists of a set G (*some objects*) and a binary operation $* : G \times G \to G$ (*a way of combining two objects to get another object, analogous to addition or multiplication*) such that:

- G is **closed** under $*$, that is $a * b \in G$ for all $a, b \in G$. (*Combining two of the objects you started with gives you another of the objects you started with.*)
- $*$ is **associative**, $a * (b * c) = (a * b) * c$ for all $a, b, c \in G$. (*It doesn't matter how you parenthesize the operation, just like addition or multiplication.*)
- There is an **identity** element $e \in G$ such that $\forall x \in G, e * x = x * e = x$. (*There's something that when you combine it with anything else has no effect, just like multiplying by one gives you the same number back.*)
- Each element $x \in G$ has an **inverse** element $x^{-1} \in G$ such that $x * x^{-1} = x^{-1} * x = e$. (*There's something you can combine with each element to get back to the identity, just like* $2 \times 0.5 = 1$.)

For example, if we take G to be the numbers less than 4, $G = \{0, 1, 2, 3\}$, and define a new operation $*$ by

$$a * b = \begin{cases} a + b & \text{if } a + b < 4 \\ a + b - 4 & \text{if } a + b \geq 4 \end{cases}$$

G and $*$ form a group, called the **cyclic group of order 4** (the **order** of a group is the number of elements in it). For example, in this group $1 * 1 = 2$, $2 * 3 = 5 - 4 = 1$ because $5 \geq 4$, $3 * 1 = 4 - 4 = 0$, etc. 0 is the identity in this group, because $0 * x = x * 0 = x$ for any of $0, 1, 2, 3$. Furthermore, the inverse of 1 in the group is 3, because $1 * 3 = 4 - 4 = 0$, the inverse of 2 is 2, and so on.

There is a great deal of structure to groups, far more than there is space to explain here. The only topic of interest for us beyond these simple properties will be the concept of **generators**. An element x generates a group if every other element of the group can be written as $x * x * \cdots * x$ for some number of xs. For example, in our cyclic group of order 4, defined above, 1 is a generator of the group because $1 = 1, 2 = 1 * 1, 3 = 1 * 1 * 1, 0 = 1 * 1 * 1 * 1$. Similarly, 3 is a generator because

$3 = 3, 2 = 3 * 3, 1 = 3 * 3 * 3, 0 = 3 * 3 * 3 * 3$. However, 2 is not a generator because $2 = 2, 0 = 2 * 2$, but there is no way to generate 1 or 3 using 2. This illustrates the only theorem we will give here:

Cyclic Group Generators Theorem: In a cyclic group of order n, written as the integers 0 to $n - 1$, $x < n$ generates the group if and only if x and n are relatively prime (i.e. have no common factors except 1).

For more information on groups and group theory, see e.g. (Lang, 2002).

Appendix B: Full Result Plots

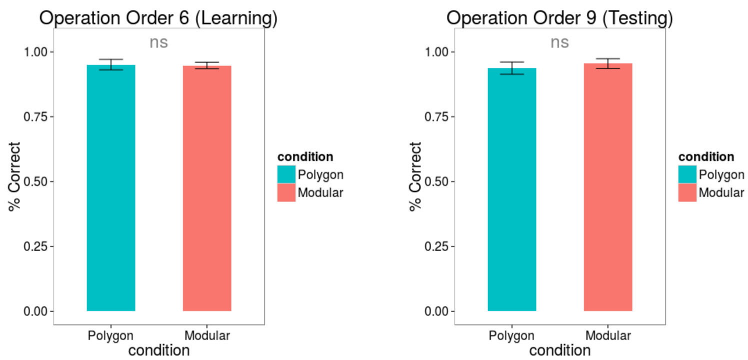


Figure 22. Experiment 1 – operation results

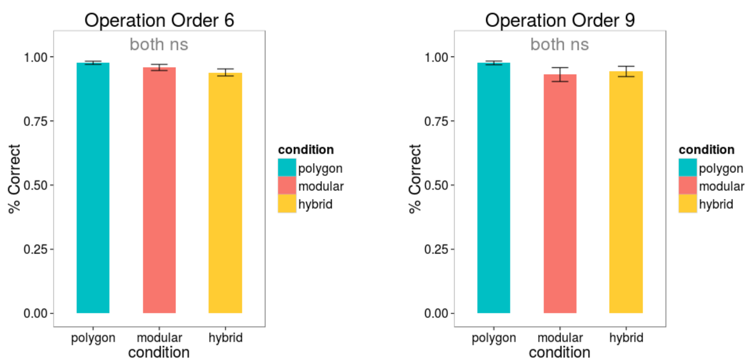


Figure 23. Experiment 2 – operation results

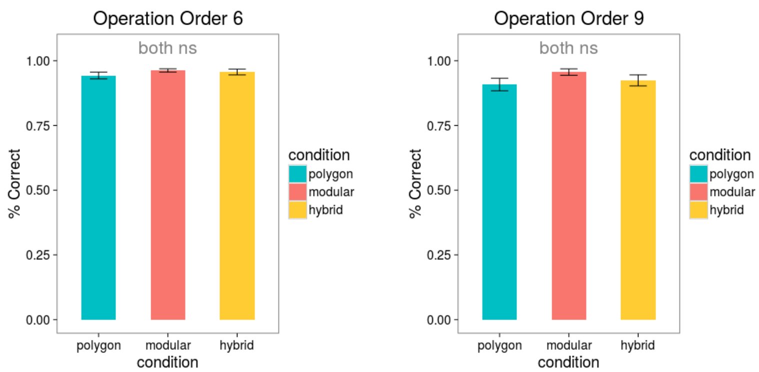


Figure 24. Experiment 3 – operation results

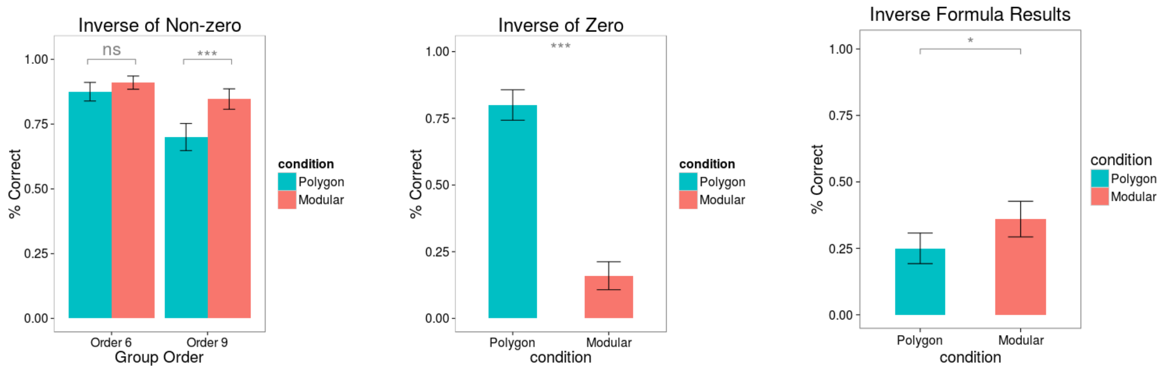


Figure 25. Experiment 1 – inverse results

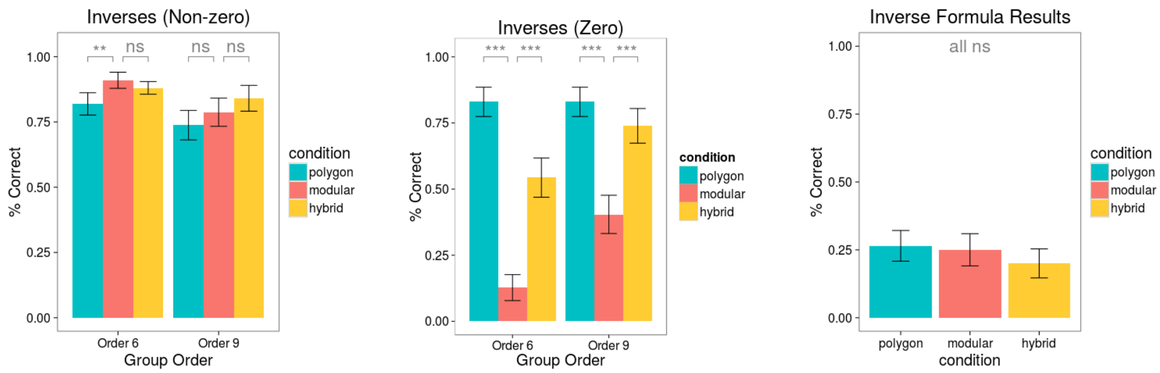


Figure 26. Experiment 2 – inverse results

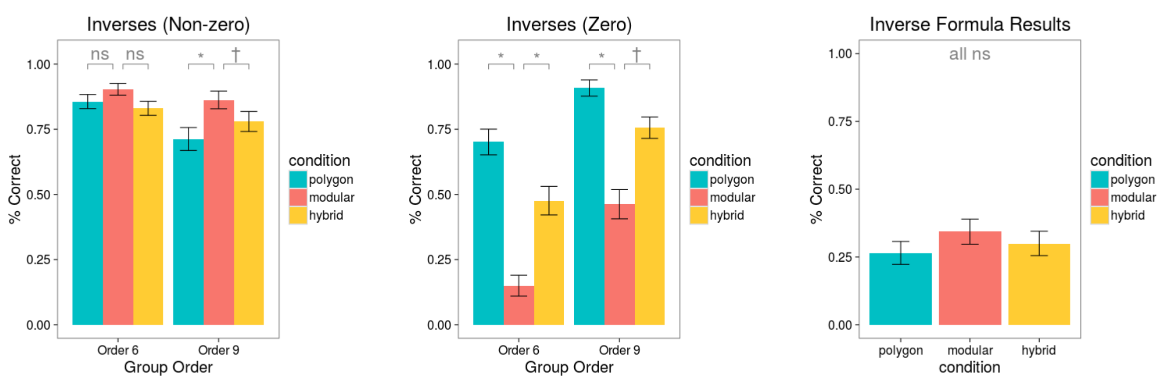


Figure 27. Experiment 3 – inverse results

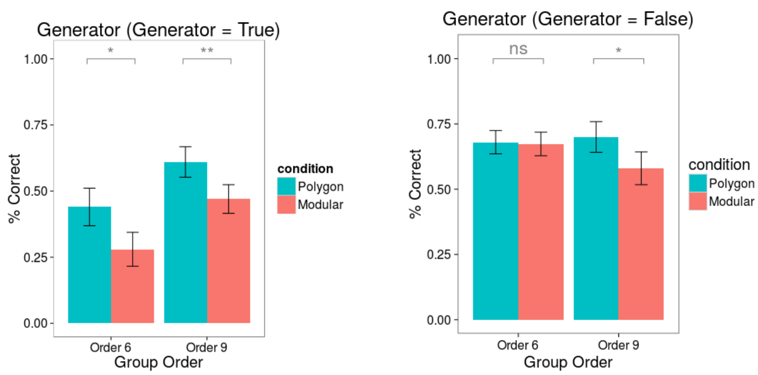


Figure 28. Experiment 1 – generator results

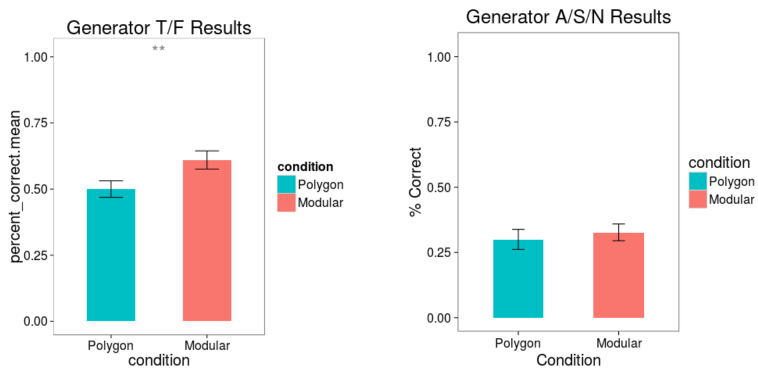


Figure 29. Experiment 1 – generator T/F & A/S/N results

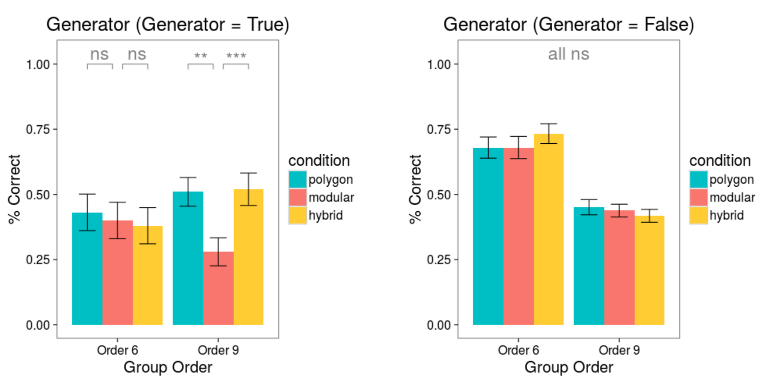


Figure 30. Experiment 2 – generator results

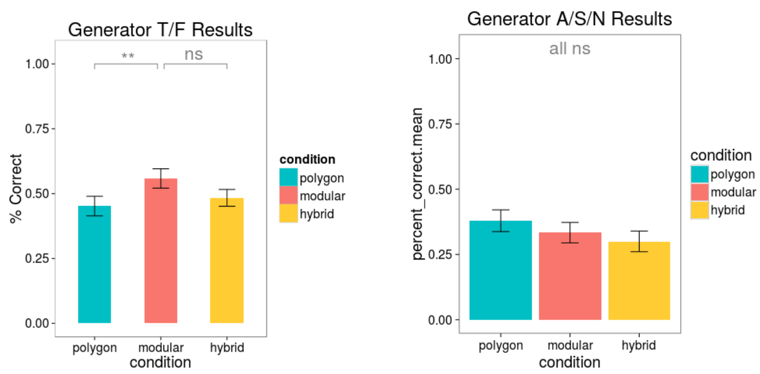


Figure 31. Experiment 2 – generator T/F & A/S/N results

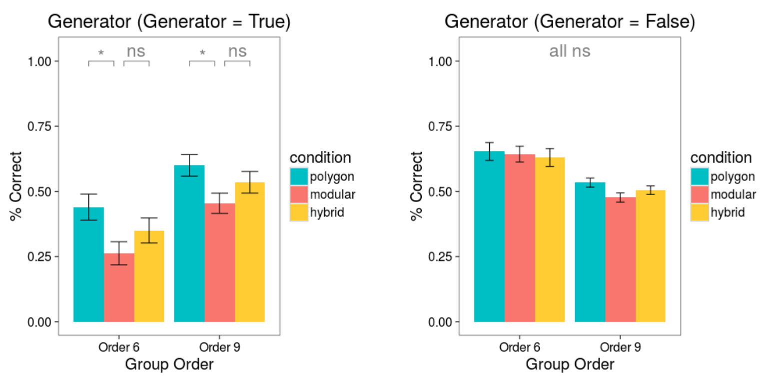


Figure 32. Experiment 3 – generator results

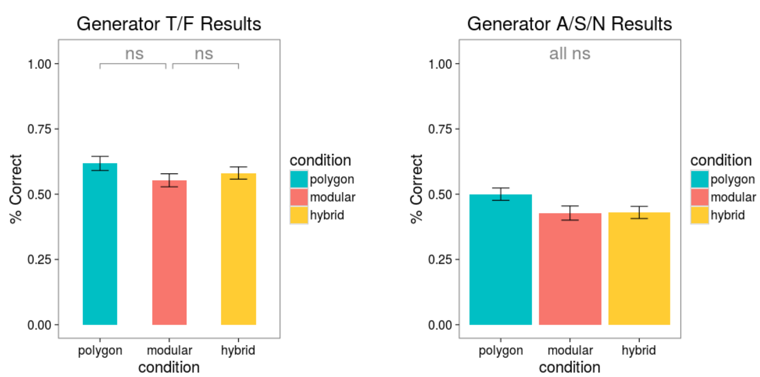


Figure 33. Experiment 3 – generator T/F & A/S/N results

The apparent chance performance on the T/F and A/S/N questions is misleading, since there were interesting patterns in the responses to individual questions, so below we have plotted the question by question responses for each experiment. For the T/F questions, Q1 was "true or false: if x is an odd number, it must be a generator under $+_n$." Q2 was "true or false: if x is an even number, it must not be a generator under $+_n$." Q3 was "true or false: if x is not a generator under $+_n$, x must be a divisor of n, that is x must divide n evenly with no remainder." Q4 was "true or false: if x is not a generator under $+_n$, x must be a multiple of a divisor of n." For the A/S/N questions, Q1 was "if an element x is a generator under $+_n$, is its inverse a generator under $+_n$ always, sometimes, or never?" Q2 was "if an element x is a generator under $+_n$, is $x +_n x$ a generator always, sometimes, or never?" and Q3 was "if an element x is not a generator under $+_n$, is $x +_n x$ a generator always, sometimes, or never?"

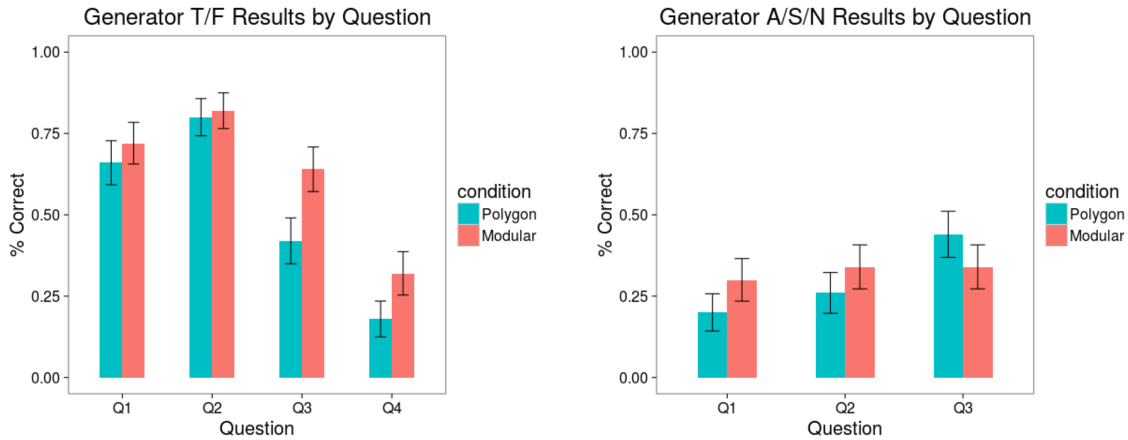


Figure 34. Experiment 1 – T/F & A/S/N results by question

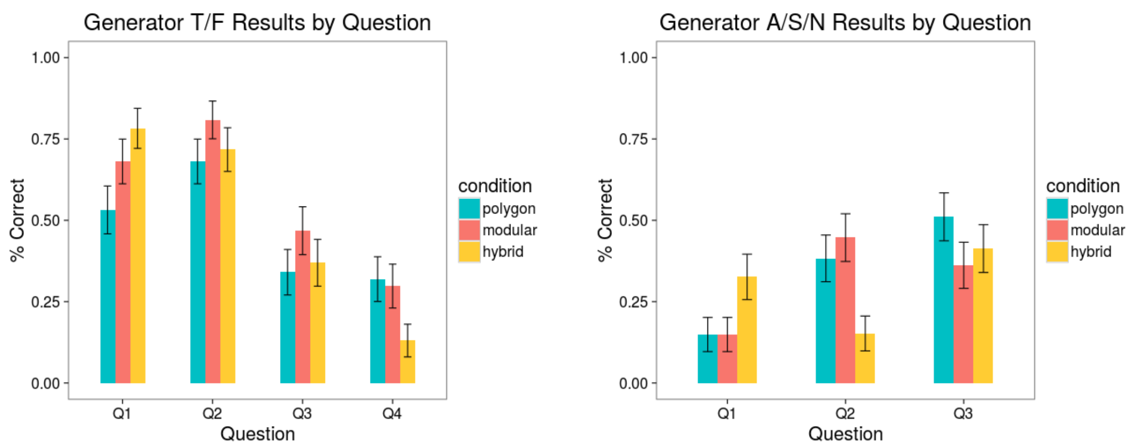


Figure 35. Experiment 2 – T/F & A/S/N results by question

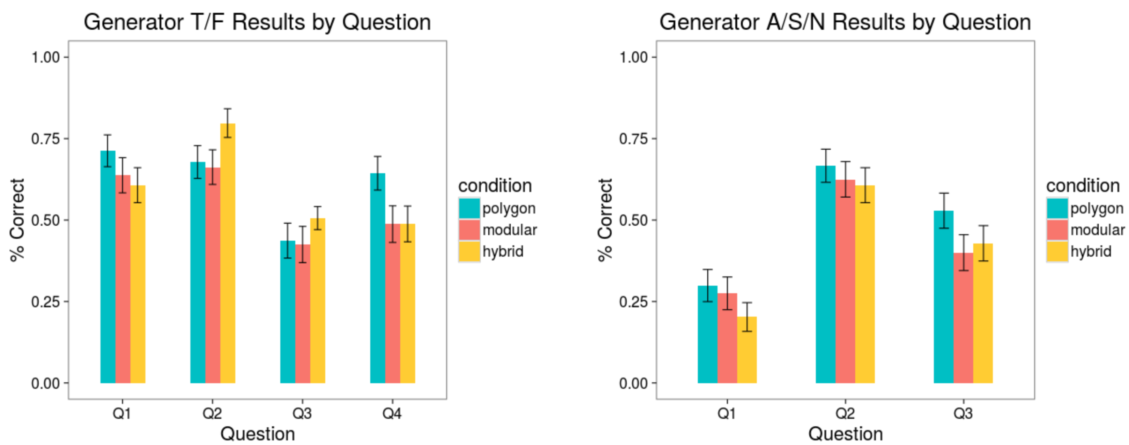


Figure 36. Experiment 3 – T/F & A/S/N results by question