Two presentations of a mathematical system: complementary advantages that can be combined

Submission to The Journal of Educational Psychology

November 8th, 2016

Abstract

Previous research has examined how different presentations of a mathematical concept affect generalizable learning of that concept (e.g. Kaminski, Sloutsky, & Heckler, 2008). This research has generally been limited to exploring the effects of presentation only on transfer to other instances of the single concept being presented. 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 concept, such as the ability to use it in practice vs. the ability to state formal theorems about it. Here, we explore how these presentations can affect subsequent concepts and aspects of understanding that build upon the original concept. Using the domain of elementary group theory, we teach participants a group operation using a visuospatial or a numerical presentation, or both. We then teach them concepts like inverses and generators that build upon this operation. We demonstrate that the presentations differ in how well they support success in learning and applying different aspects of the concepts presented, with each having advantages and disadvantages. Instead of pursuing a single ideal presentation, we show that presenting both presentations and encouraging participants to recognize the relationship between them leads to better performance, at least for some participants.

Two presentations of a mathematical system: complementary advantages that can be combined

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 affect 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 that 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.

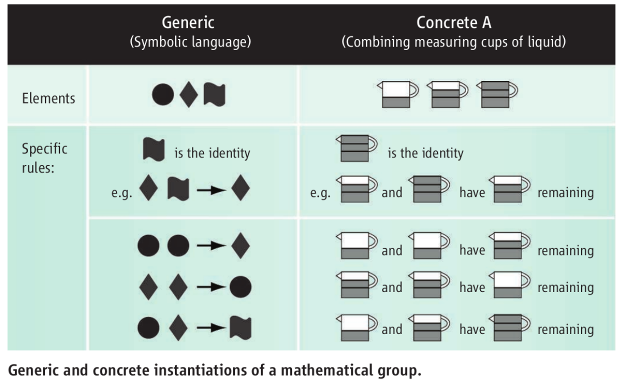
We examined these issues within the area of elementary group theory, specifically cyclic groups and some group theoretic concepts relevant to them. An introduction to the relevant concepts is provided in Appendix A, but we briefly sketch them here. A group consists of a set of elements and an operation that takes any two members of the group and always produces a group member as a result. A cyclic group is a group whose members can be seen as forming a cycle. More specifically, the cyclic group of order n is a group of n elements whose elements can be brought into correspondence with the numbers 0 to n-1. Once a correspondence with the integers 0 to n-1 has been established, the rule for combining two elements can be described as adding the two numbers, and subtracting n if the result is greater than or equal to n. There are many concepts that can be built up from these simple ideas, including the identity element of the group (the element that leaves every other element unchanged under the operation, in this case 0); the concept of the inverse of an element (the thing you combine with an element to get the identity); and the concept of a generator of the group (an element that can make every other element of the group by repeatedly adding the element to itself). Our use of this particular task domain was inspired by previous work on presentations in mathematics education.

Background: The Advantage(?) of Abstract Examples

Kaminski and colleagues (Kaminski et al., 2008) explored the effects of presentations in a cyclic group of order 3. They presented participants with either a "generic" instantiation of the group, or a "concrete" one. Their presentations are illustrated in figure 1a. The generic presentation consists of some arbitrary geometric symbols, with enforced rules for combining them, and the concrete presentation consisted of an example with a narrative about combining fractional cups of liquid, and finding the amount left over. There were two other concrete presentations (not shown) that were also used in some experimental sessions (using fractional slices of a pizza or number of tennis balls in a can as the concrete objects.) They trained participants to perform the operation in either the generic presentation or one to three concrete presentations. They then showed participants 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 participants 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 participants who learned the generic presentation performed better at this transfer than the participants who learned the concrete presentation(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) suggests that in the concrete presentations "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 presentation "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

(a) Group presentations



(b) Transfer domain

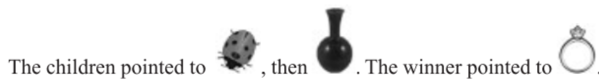


Figure 1. Group presentations from (Kaminski, Sloutsky, and Heckler, 2008). Figures adapted from “The Advantage of Abstract Examples in Learning Math” by J. A. Kaminski, V. M. Sloutsky, & A. F. Heckler, 2008, *Science*, 320 (5875), p. 454 and “Abstract or Concrete Examples in Learning Mathematics? A Replication and Elaboration of Kaminski, Sloutsky, and Heckler’s Study” by D. De Bock, et al., 2011, *Journal for Research in Mathematics Education*, 42 (2), p. 112. Permission pending.

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 participants (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, the possibility remains that there are structural or conceptual differences between the concrete and generic instantiations.

For example, one aspect of the presentations that is different is the asymmetry that participants 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 the unit quantity and the other to twice that quantity would not find any such cue. Similarly, if the notion of generators had been discussed in the study, participants who saw the concrete examples 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 concrete presentations provide a shared basis (number of identified parts, be they tennis balls, slices of pizza, or 1/3s of a cup of liquid) that can be used to map one onto another. This obvious mapping is not present in either the generic or transfer examples. In summary, the numerical content of the concrete presentations may change what participants 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 presentation was better for transferring to presentations that were similar in terms of whether or not they supported a mapping to number. Furthermore, they asked participants to give a free response justifying their answer to a problem of combining four elements of the group, and rated it on the ideas that it contained. They found that generic-presentation group participants mentioned more group-theoretic ideas (although they still appeared to attain very little understanding of them), but that concrete-presentation group participants mentioned the ideas of modular arithmetic as well as some group-theoretic ideas. Thus, the choice of presentation had an effect not just on transfer, but on the more abstract concepts being inferred. These results suggest that in pedagogically more realistic scenarios where students are asked to explore concepts further, the presentation may have effects that propagate.

De Bock et al. and Kaminski et al. did not teach additional concepts to their participants beyond presenting the rules for applying the operation to arbitrary strings of symbols, (although concepts like the identity may have been implicitly communiated by the format of the rules). 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 participants ability to identify the correspondences between the original elements and the transfer elements in order to perform the operation. They did not evaluate how presentations affected participants 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 that concepts are related to one another in mathematics instruction, and consider some of the aspects of understanding that 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 these relationships 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 (at least in abstract algebra) reduce the level of abstraction by relying on properties of more concrete examples that 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 participants 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. The groups Kaminski et al studied all correspond to the cyclic group of order 3. In an educational setting, after learning this operation students might learn about the identity of the group, inverses, generators, etc. 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. We will refer to these as "aspects of understanding," but they are essentially different performance measures that can be used to assess students' mastery of diferent aspects of the mathematical system under consideration in several more specific ways. They may not capture all the aspects of understanding that are pedagogically relevant in all circumstances; we propose these questions more as a starting point for thinking about these issues than a definitive list. The questions we suggest for assessing a presentation (with examples from the case of cyclic groups) are:

- Does it allow students to apply the directly-instructed base concepts correctly? Does it allow them to transfer these base concepts to a non-isomorphic group? (Does it allow students to combine elements using the group operation within the context of the specific example – e.g. a group of order 6 – used to introduce the operation? Once they have learned this in a group of order 6, does it allow them to do similarly in a group of order 9 with minimal additional explanation?)

- Does it allow them to answer questions about further concepts that build upon the base concept, within the original instance? Does it allow them to transfer these concepts to a non-isomorphic group? (Does it allow them to correctly identify inverses and generators in the cyclic group of order 6? Does it allow them to transfer this understanding to a cyclic group of order 9?)

- 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 (or evaluate the truth of) 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 statements 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 transfer between distinct presentations of the same group. Their participants may have developed other aspects of understanding (for example, participants might have discovered the concept of inverses as computationally useful when combining long strings of symbols), but their experiments did not explicitly encourage or assess this. Here, we move beyond this to ask which presentations are better for advancing each of these aspects of understanding. We demonstrate that two presentations which give similar performance on direct application of the group operation can each have advantages and disadvantages for other aspects of understanding of 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.

The idea that this might be beneficial has roots in the work of Gick and Holyoak (1983), who showed among other things that seeing multiple analogs of an idea was more likely to lead to transfer, and dissimilar analogs were beneficial in some cases, although they did not address independent and complementary advantages of distinct presentations. It can also be related to more recent work; for example Schwartz and Goldstone (2015) have argued that rather than "an either-or" problem, we should try to "coordinate learning processes so they can do more together than they can alone." They explicitly refence the abstract-concrete dichotomy, but more generally we can view different presentations from the perspective of different learning processes, especially if one presentation is visuospatial and interactive while another is more symbolic and abstract. Other work has shown the benefits of concreteness fading, the idea that concrete and abstract materials each have advantages, and "fading" from concrete presentations to more abstract ones may enhance learning (Fyfe, McNeil, Son, & Goldstone, 2014). We demonstrate that the benefits of multiple presentations may be more general than just the effects of presenting concrete and abstract. We show that even over the short time of an experimental session, some subjects develop the ability to exploit the advantageous features of two different presentations with complementary advantages.

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 (we call this the "polygon presentation"), and the other is based on arithmetic and is closely related to modular arithmetic (the "modular presentation"). 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 the general case of a cyclic group of unspecified order n, to investigate the effects of these presentations on different aspects of understanding. We found that while both presentations were very successful at allowing students to learn to correctly apply the group operation and extend it to a group of a new order, the presentations produced differential success with related concepts. Furthermore, neither was clearly superior, each had advantages and disadvantages relative to the other, and these advantages and disadvantages transferred to the group of a new order. Thus we explored combining these presentations to produce a hybrid presentation and found that this hybrid presentation was beneficial for at least some participants.

Experiments

Introduction

In this paper, we present the results from three closely related experiments. (These experiments were performed sequentially, but the methods and 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 participants the hybrid presentation that included both the polygon and modular presentations, and encouraging the participants to integrate them (while keeping total instruction time approximately the same).

Experiment 3: In our third experiment, we wished to further increase our confidence in the results from the first two experiments, and to further explore the thought processes of hybrid-group participants. In order to examine this, we added questions for the hybrid group (presented after the main experiment had been completed), which asked them to rate the extent to which they had used each presentation when answering a question.

Participants

We performed the experiments on Amazon's Mechanical Turk (c.f. Buhrmester, Kwang, & Gosling, 2011), using high-reputation participants (over 85% approval rate), and using participant tracking (so we could run follow-up and replication studies on Mechanical Turk without having the same participants participate and contaminate the results). Our experimental design was approved by our IRB, and all subjects gave informed consent to participate in the experiments. Experiment 1 had $n = 50$ participants per group, $N = 100$ total; experiment 2 had $n = 50$, $N = 150$; and experiment 3 had $n = 100$, $N = 300$.

Materials & Methods

All materials are available on our github[1], including complete versions of our experiments that 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)
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 participants to perform the group operation on a cyclic group of order 6 (using the polygon, modular, or hybrid presentation, between participants), 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 participants' 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 participants for understanding of the general 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 general (and usually formal) statements about the processes and concepts learned.

[1] (masked)

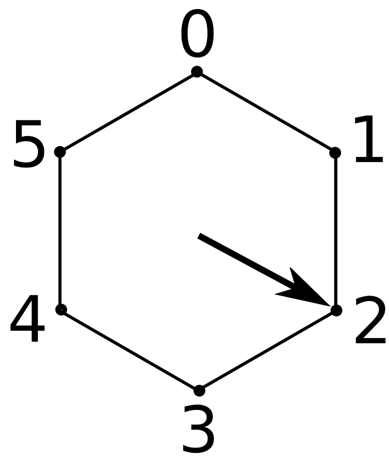


Figure 2. Order 6 polygon figure

Group presentations. In all three experiments, participants in one group received a presentation based on modular arithmetic (which is easily explained as a slight variation on regular arithmetic), while participants in another group received a more visually concrete presentation based on counting around a polygon (which allows participants to develop a visual intuition, but which is not as directly familiar as standard arithmetic, although participants may find analogies, e.g. to clocks). For experiments 2 and 3, we added a hybrid group, where participants were presented with both presentations and asked to integrate them.

For the modular presentation, we presented the group operation as $+_6$, and we explained to participants that to compute $+_6$ you add the two numbers, and then subtract 6 if your result is 6 or larger. We gave examples such as $4 +_6 4 = 2$, because $4 + 4 = 8$ and $8 - 6 = 2$.

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 participants with the diagram shown in Figure 2. The diagram that participants 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 participants 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. We gave examples such as 4⑥4 = 2, because 4 steps clockwise from 4 makes the arrow point at 2.

After seeing several examples, participants 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 participants received feedback on their answers and an explanation of the correct answer.

For the hybrid group participants, we presented both presentations, calling them respectively the "arithmetic method" and "polygon method." (We used "arithmetic method" because we felt that some participants would find the term "modular method" to be confusing, since they could associate it with meanings other than the intended mathematical meaning.) We alternated asking the participants to use the polygon 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 participants 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 the practice) asking the hybrid group participants to reflect on how the different methods corresponded; participants in other groups were asked 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 article, 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 produce the identity. For example, the inverse of 1 is 5, because $1 \text{⑥} 5 = 0$ and $5 \text{⑥} 1 = 0$. We then allowed participants to find inverses for all other group elements as practice, and participants received feedback on their answers and an explanation of the correct answer.

Generators. Finally, we taught the participants 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 $1 = 1$, $2 = 1 +_6 1$, 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 participants to find whether each of the remaining elements generates the group, and provided them with feedback on their answers and an explanation of the correct answer.

Transfer test. We next tested the participants 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 participants one practice problem (with feedback) on the new operation, to ensure that they understood it. We then asked the participants questions to test their knowledge of the concepts outlined in each section above, namely:

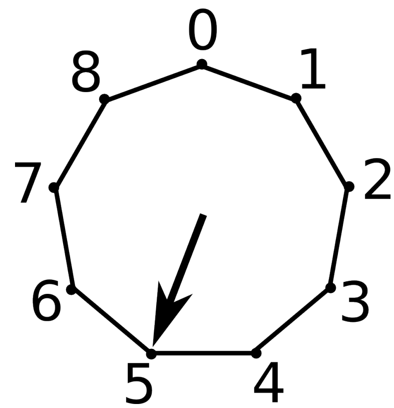


Figure 3. Order 9 polygon figure

- A set of seven problems with the group operation, e.g. $6 \textcircled{9} 4 = ?$, with participants asked to provide an explanation of their answers for two of them.
- One problem asking participants to identify the identity under the operation, and to explain their answer.
- Three problems asking subjects to identify the inverse of an element; one of these also asked them to explain their answer.
- Four problems asking subjects to identify whether an element was or was not a generator. Two generators and two non-generators were presented, and participants were asked to explain their answer for one of each.

Test of reasoning about the general case. Finally, we told participants we were now considering an order n cyclic group, presented to the modular group participants as $+_n$, and to the polygon group participants as $\textcircled{n}$ 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 participants.) We then asked them the following questions:

- What is the identity under $+_n$?
- Two questions on giving formulas for inverses under $\textcircled{n}$, 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.)

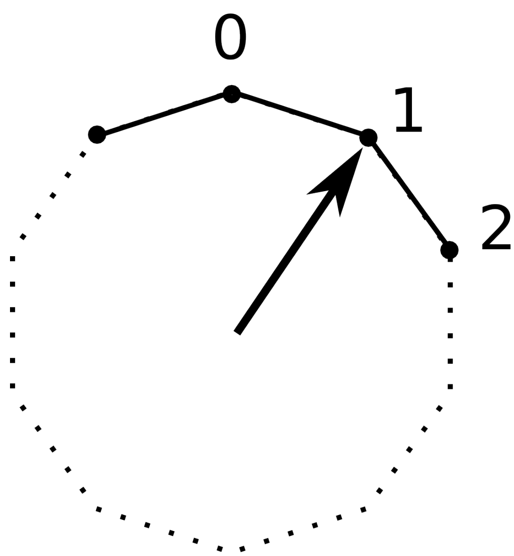


Figure 4. Order n polygon figure

- Three always/sometimes/never questions about generators. (E.g. If an element x is a generator under ⓝ, 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). (We will use the term representation to refer to the method hybrid group participants were using to think about the problem.) In experiment 3, we added four questions for the hybrid participants. On these questions, 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 perform like the better of the polygon and modular groups on each question type.

This can be contrasted with other possible predictions for hybrid group performance. One possibility is that seeing both presentations would simply confuse or overload the participants, and they would perform worse on every type of question, resulting in them being significantly worse at every question type. Another possibility is that participants would just pick one presentation and use it exclusively, and perform as though they were participants in that presentation group. This, and possibilities such as participants randomly picking a presentation 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 participants may achieve maximal performance while others are simply confused. These possibilities could also produce a similar pattern of results.)

Finally, for the experiment 3 questions where we had the hybrid participants describe which representation they used, we hypothesized that where the polygon participants 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 participants 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 participants, with the fixed effects being question type, including the group order (6, 9, or n) where it occurred; presentation, 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 participants 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 participants bootstrapped across 10,000 resamples of the participants, 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. (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. 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 participants 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.

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 participants were able to transfer their procedures for solving the questions to a different group order, only some participants were able to reason about the general case or express formal statements about generic cyclic groups.

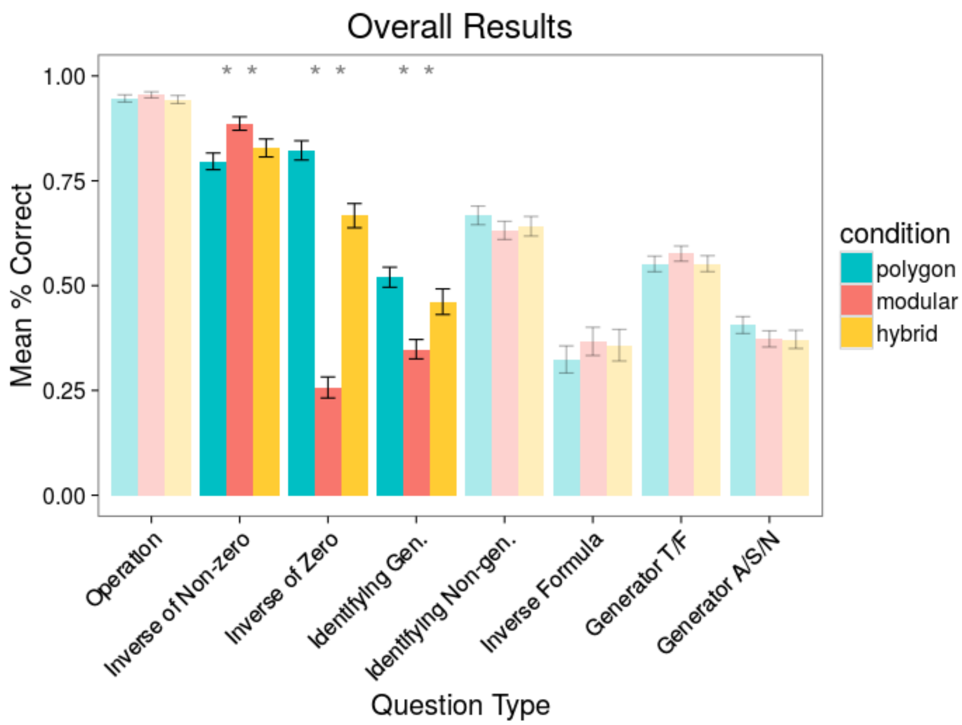
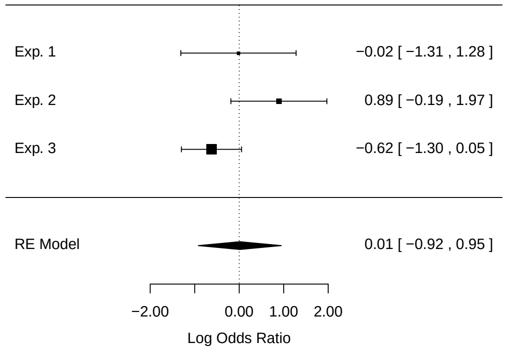
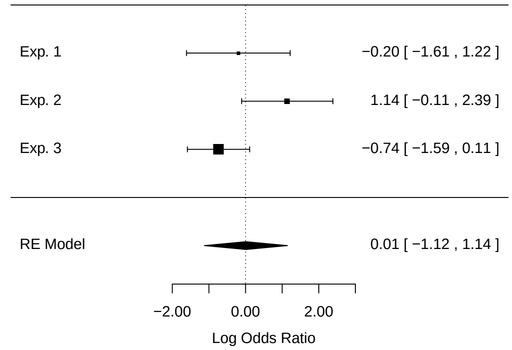


Figure 5. Results aggregated across group orders and experiments 2 & 3 (Exp. 1 had no hybrid condition, so we omitted it from this graph). Highlighted results are the main findings relevant to our hypothesis, stars mark comparisons where the meta-analysis 95% confidence interval did not overlap zero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

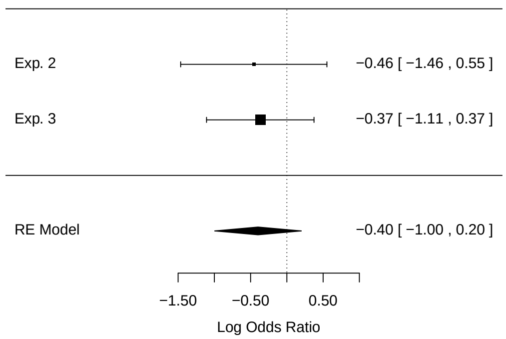
The participants in the polygon and modular groups differed significantly on a number of question types, with the polygon group consistently performing better at identifying elements that were generators and finding the inverse of zero, while the modular group performed significantly better at finding the inverse of non-zero elements. See Figure 5 for a summary of the results aggregated across experiments and group orders. Note that after performing some post hoc analyses we noticed that these aggregated results may understate the hybrid group's final level of understanding relative to the other two groups, because the hybrid group showed an improvement on a number of aspects of understanding between the order 6 and order 9 questions. (See Figure 16 for the aggregated results split across group orders showing this pattern of improvement, and the Hiearchical modeling section for further discussion of this difference.) For the sake of brevity, we present below only the results from our meta-analysis of the logistic regressions performed on each experiment, (but note that the forest plots for each predictor includes the estimates and confidence intervals for that predictor from each experiment). For the full results and significance tests from the analyses for each experiment, see Appendix B.



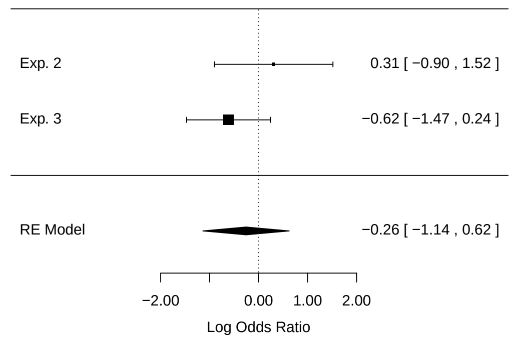
(a) Polygon vs. modular order 6



(b) Polygon vs. modular order 9



(c) Hybrid vs. modular order 6



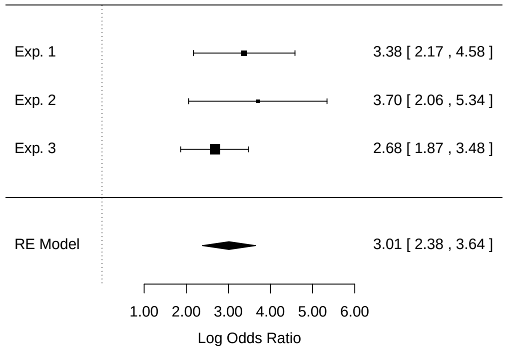
(d) Hybrid vs. modular order 9

Figure 6. Meta analysis – operation

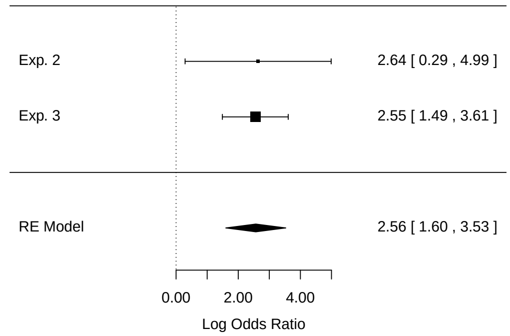
Main Results

In this section we present the result of our meta-analysis of the logistic regressions performed on the data from all the experiments. We report estimated effect sizes using the approach described by Chinn (2000). In this approach, performance is thought to depend on a normally distributed random variable, and the effect of a manipulation is viewed as shifting the mean upward (for a positive effect size) or downward (for a negative effect size) by the indicated units of the distribution's standard deviation across the population of participants.

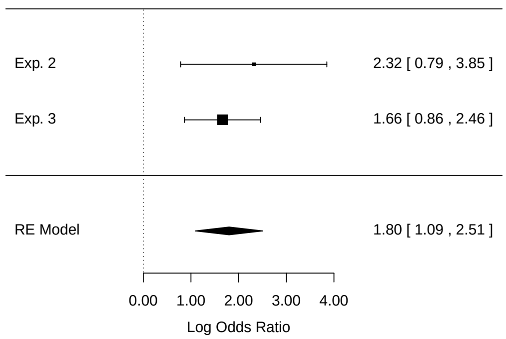
Operation. We estimate any effect of the polygon presentation on the ability to perform the group operation to be negligible (order 6: log OR (Odds Ratio) = 0.01, effect size = 0.01; order 9: log OR = 0.01, effect size = 0.01; see Fig. 6 top). We estimate any effect of the hybrid presentation on the ability to perform the group operation to be small or negligible (order 6: log OR = -0.40, effect size = -0.22; order 9: log OR = -0.26, effect size = -0.14; see Fig. 6 bottom). Despite learning different methods for performing the group operation, the experimental groups do not differ substantially in their ability to perform it, although the hybrid group appears to lag a little.



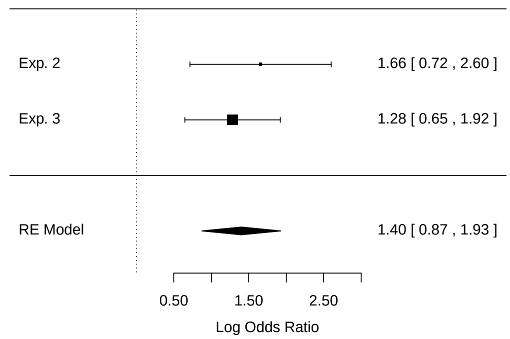
(a) Polygon vs. modular, order 6



(b) Polygon vs. modular order 9



(c) Hybrid vs. modular, order 6

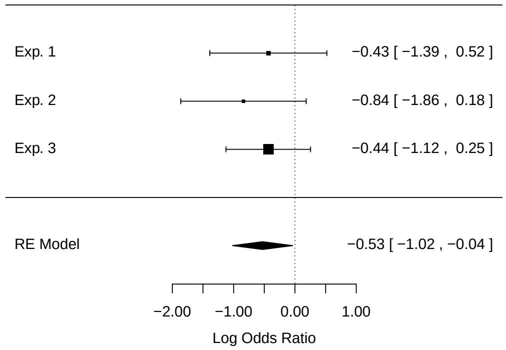


(d) Hybrid vs. modular, order 9

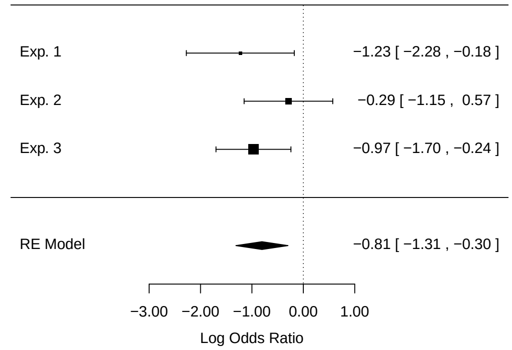
Figure 7. Meta analysis – inverse of zero

Inverses. 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, effect size = 1.66; order 9: log OR = 2.56, effect size = 1.41; see Fig. 7 top). We estimated the negative effect of the polygon condition on inverse of non-zero questions to be small (order 6: log OR = -0.53, effect size = -0.29; order 9: log OR = -0.81, effect size = -0.45; see Fig. 8 top). 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, effect size = -0.99; order 9: log OR = 1.40, effect size = -0.77; see Fig. 7 bottom). We estimated the negative effect of the hybrid condition on inverse of non-zero questions to be small, and negligible after further practice in the order 9 group (order 6: log OR = -0.58, effect size= -0.32; order 9: log OR = -0.16, effect size = -0.09; see Fig. 8 bottom).

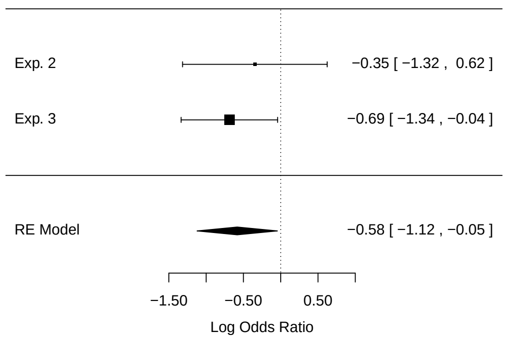
Overall, it seems that the modular presentation is generally beneficial for finding inverses, except in the case of zero, where the polygon presentation participants perform much better. (See discussion for a possible explanation of this result.)



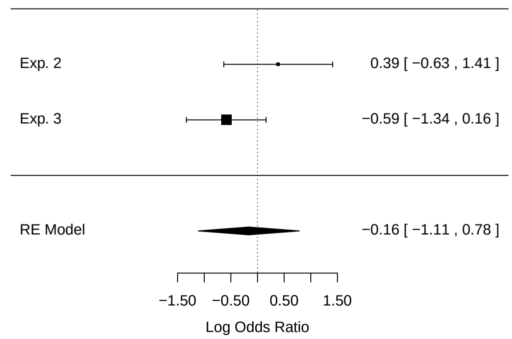
(a) Polygon vs. modular, order 6



(b) Polygon vs. modular, order 9



(c) Hybrid vs. modular, order 6

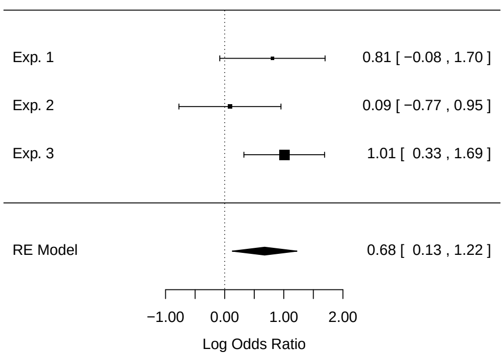


(d) Hybrid vs. modular, order 9

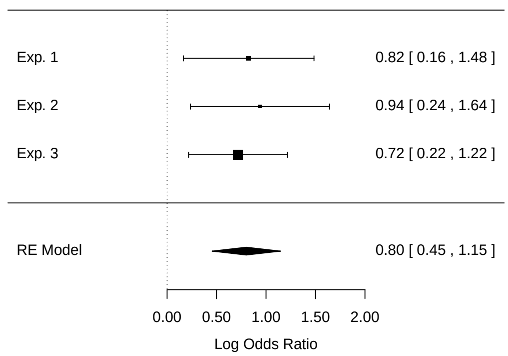
Figure 8. Meta analysis – inverse of non-zero

Generators. We estimated the positive effect of the polygon condition on identifying generators to be small (order 6: log OR = 0.68, effect size = 0.38; order 9: log OR = 0.80, effect size = 0.44; see Fig. 9 top). We estimated the effect of the polygon condition on identifying non-generators to be negligible in group of order 6, but trending toward a small positive effect in the group of order 9 (order 6: log OR = 0.07, effect size = 0.04; order 9: log OR = 0.36, effect size = 0.20; see Fig. 10 top). We estimated the effect of the hybrid condition on identifying non-generators to be negligible (order 6: log OR = 0.07, effect size = 0.04; order 9: log OR = -0.01, effect size = -0.01; see Fig. 10 bottom). We estimated the positive effect of the hybrid condition on identifying generators to be negligible in the order 6 group, but increasing in the order 9 group, (order 6: log OR = 0.19, effect size = 0.10; order 9: log OR = 0.65, effect size = 0.36; see Fig. 9 bottom).

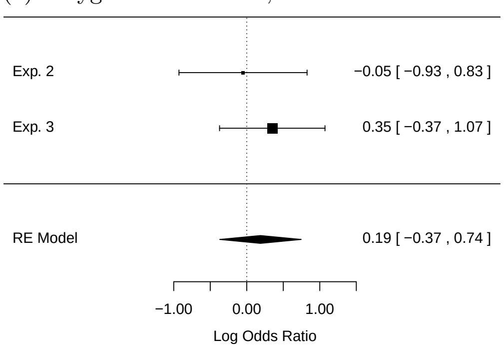
Overall, it seems that the polygon presentation is beneficial for identifying generators, and the hybrid presentation seems to be similarly beneficial for identifying generators in the order 9 group, once the participants have had some practice.



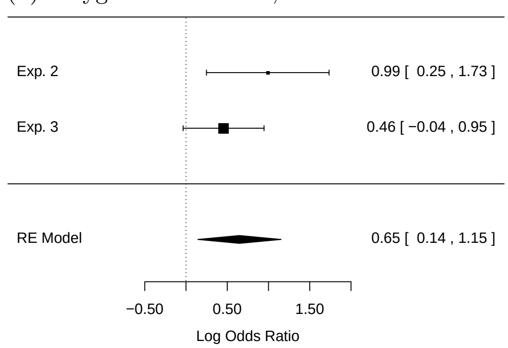
(a) Polygon vs. modular, order 6



(b) Polygon vs. modular, Order 9

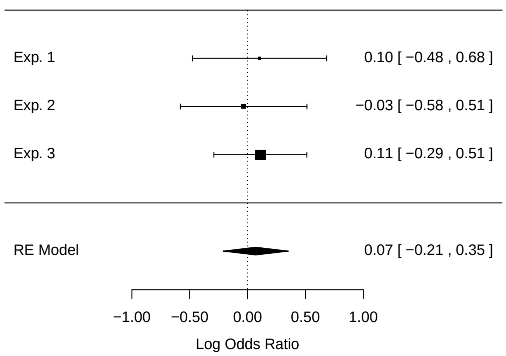


(c) Hybrid vs. modular, order 6

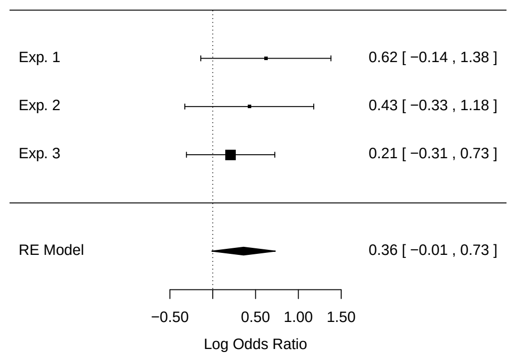


(d) Hybrid vs. modular, order 9

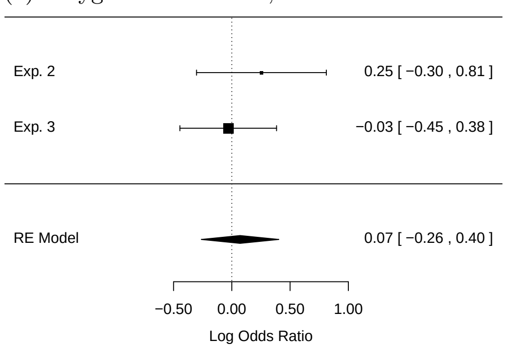
Figure 9. Meta analysis – identifying generators



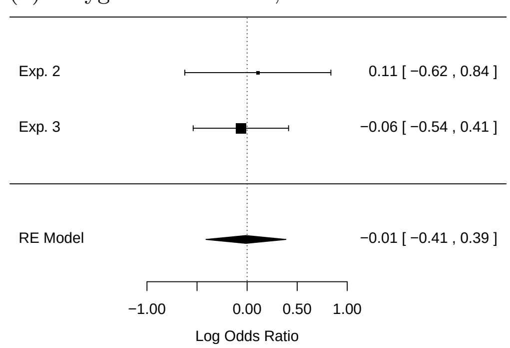
(a) Polygon vs. modular, order 6



(b) Polygon vs. modular, order 9



(c) Hybrid vs. modular, order 6

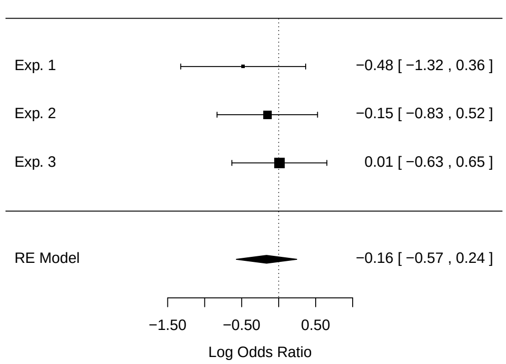


(d) Order 9

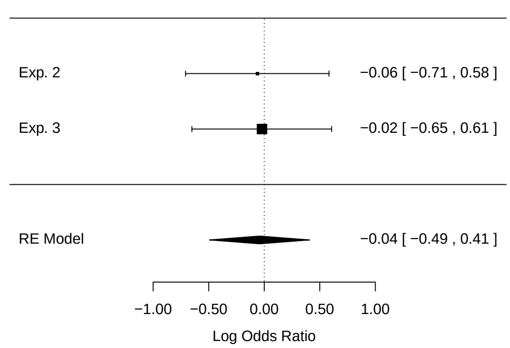
Figure 10. Meta analysis – identifying non-generators

Questions assessing reasoning about the general case. We estimated the effect of the polygon condition on the inverse formula questions to be neglible (log OR = -0.16, effect size = -0.09, see Fig. 11), and similarly for the effect of the hybrid condition on the inverse formula questions (log OR = -0.04, effect size = -0.02, see Fig. 11). We estimated that any effect of the polygon condition on answering True/False questions about generators is negligible (log OR = -0.14, effect size = 0.08; see Fig. 12). We estimated that any effect of the polygon condition on answering Always/Sometimes/Never questions about generators is negligible (log OR = 0.17, effect size = 0.09; see Fig. 12 top). We estimated the effect of the hybrid condition on answering True/False questions about generators to be negligible (log OR = -0.05, effect size = -0.03; see Fig. 12 bottom). We estimated the effect of the hybrid condition on answering Always/Sometimes/Never questions about generators to be negligible (log OR = -0.11, effect size = -0.06; see Fig. 12 bottom).

None of the presentations seem particularly beneficial for discovering formulas for the inverse, or for answering T/F or A/S/N questions about generators; performance was quite low on these questions, especially the T/F and A/S/N.

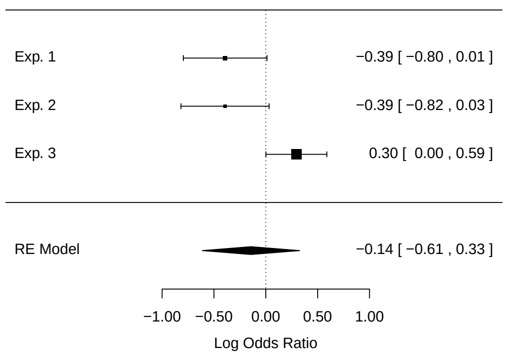


(a) Polygon vs. modular

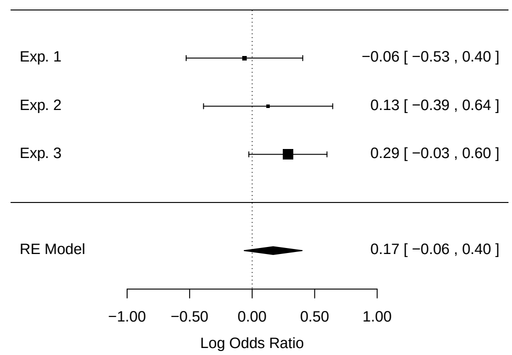


(b) Hybrid vs. modular

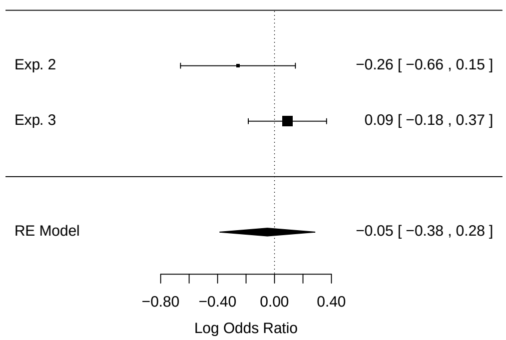
Figure 11. Meta analysis – inverse formula



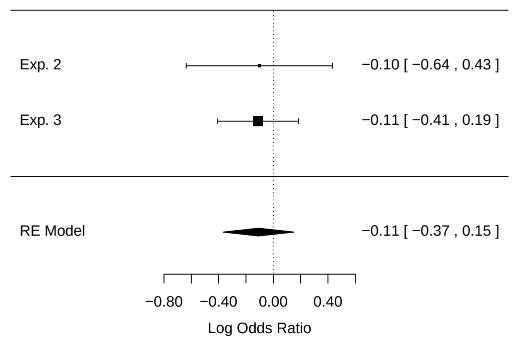
(a) Polygon vs. modular T/F



(b) Polygon vs. modular A/S/N



(c) Hybrid vs. modular T/F



(d) Hybrid vs. modular A/S/N

Figure 12. Meta analysis – T/F & A/S/N

Other Analyses

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

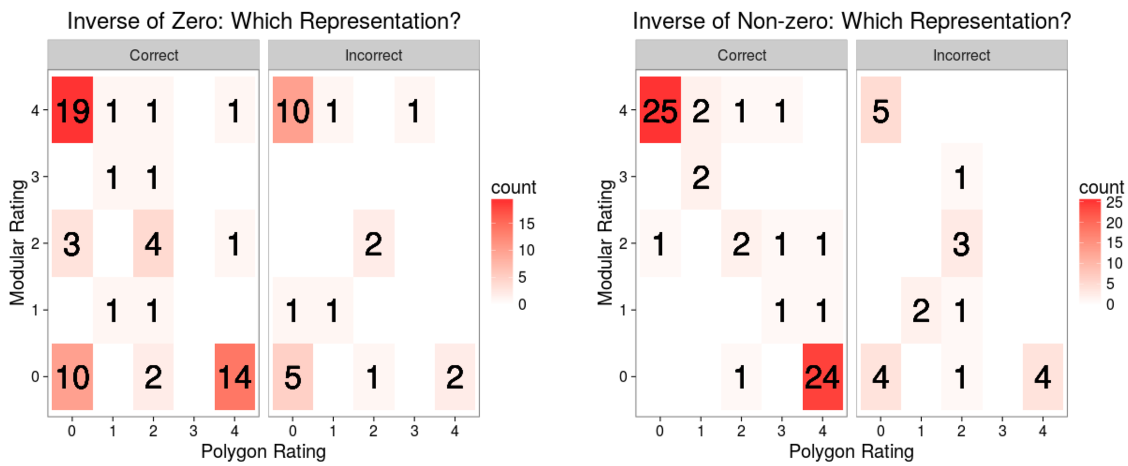
Diagram use. We hypothesized that the polygon group's superior performance on identifying generators might be due to the ability to use the spatial structure of the polygon to more easily visualize the elements generated by an element (see discussion). One possible prediction of this hypothesis would 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 participants 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 participants 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: $b = 1.90$, $z = 2.76$, $p = 0.006$; Exp. 3: $b = 2.24$, $z = 5.12$, $p < 0.001$). Furthermore, this effect was present even when controlling for reaction time (Exp. 2: $b = 1.62$, $z = 2.26$, $p = 0.024$; Exp. 3: $b = 1.64$, $z = 3.51$, $p < 0.001$). This 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: $b = 1.26$, $z = 6.93$, $p < 0.001$; Exp. 3: $b = 1.19$, $z = 10.28$, $p < 0.001$), even when controlling for reaction time (Exp. 2: $b = 1.42$, $z = 7.53$, $p < 0.001$; Exp. 3: $b = 1.25$, $z = 10.71$, $p < 0.001$). However, the estimated effect sizes were smaller than for the generator questions. This suggests that the diagram may have been especially helpful on these generator questions, as we hypothesized.

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 ($b_{mod} = -0.02$, $z = -0.15$, $p = 0.88$; $b_{poly} = 0.25$, $z = 1.37$, $p = 0.17$). 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 participants may have simply recalled the answer. Performance was very high in the hybrid group overall on this question, the majority (71%) of the participants got the question right in the representation-use section, so it had the only positive intercept of any of the representation-use regressions.

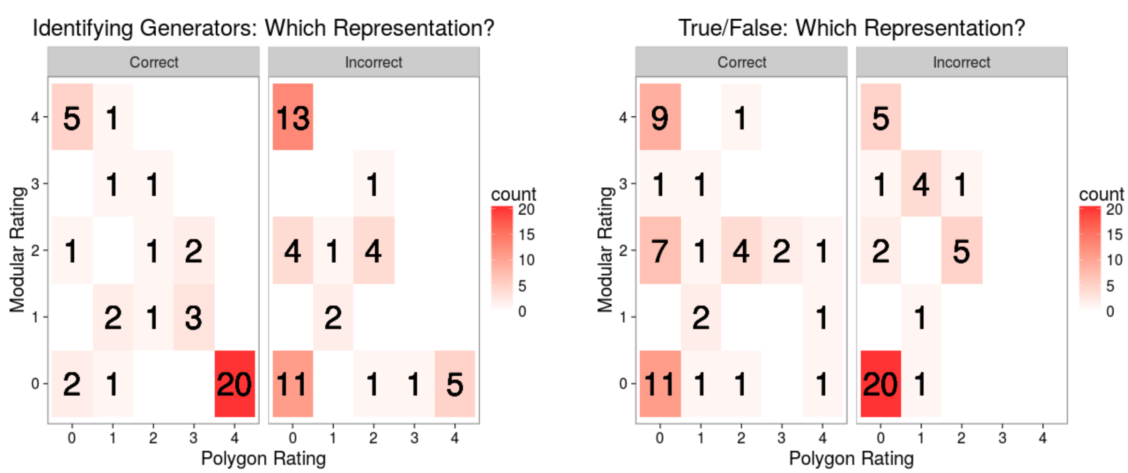
Intriguingly, we found that both modular and polygon rating were significantly predictive of success on inverse of non-zero questions ($b_{mod} = 1.09$, $z = 2.95$, $p = 0.003$; $b_{poly} = 1.11$, $z = 2.94$, $p = 0.003$). This, together with the previous finding, may suggest some integration occuring 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 participants polygon rating, but not modular, was significantly predictive of success on identifying generators ($b_{mod} = 0.19$, $z = 0.95$, $p = 0.34$; $b_{poly} = 0.75$, $z = 3.86$, $p < 0.001$). 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 ($b_{mod} = 0.19$, $z = 1.28$, $p = 0.20$; $b_{poly} = 0.31$, $z = 1.38$, $p = 0.17$). This is unsurprising, since we did not observe any significant differences between the polygon and modular groups on these questions in the third experiment.



(a) Inverse of zero

(b) Inverse of non-zero

Figure 13. Experiment 3 – representation-use responses on inverse questions (counts of participants giving each rating, split by whether answer was correct)



(a) Identifying generators

(b) Generator T/F

Figure 14. Experiment 3 – representation-use responses on generator questions (counts of participants giving each rating, split by score)

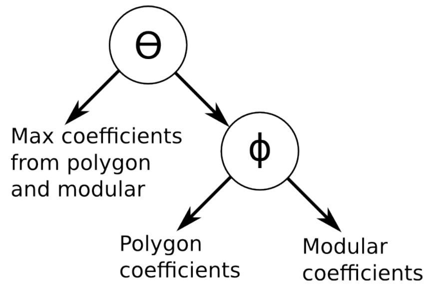


Figure 15. Hierarchical model

Hierarchical modeling of hybrid participants

Although we found that 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. In this section, we explore alternative ways of accounting for this finding, using a post-hoc, hierarchical modeling approach. One explanation for the pattern of results might be that some participants 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 that assumed that the data were generated by the following process (depicted schematically in Fig. 15):

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

2. If the participants did not benefit from both presentations (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 $\theta = 0.41, \phi = 0.49$, so the data are best fit under this model by assuming that about 40% the participants are benefitting from both representations, and those that aren't are choosing the modular representation and polygon almost equally. We used the Bayesian Information Criterion (BIC) to compare this model ($BIC = 1653.1$) to models where all participants chose modular ($BIC = 1829.0$), all chose polygon ($BIC = 1720.5$), where no participants benefitted from both i.e. a fixed $\theta = 0$ and fit $\phi = 0.56$ ($BIC = 1681.2$), and a model where all participants benefitted from both i.e. $\theta = 1.0$ ($BIC = 1689.9$). The BIC comparisons (all differences > 36) provide "very strong" (Kass & Raftery, 1995) evidence that the

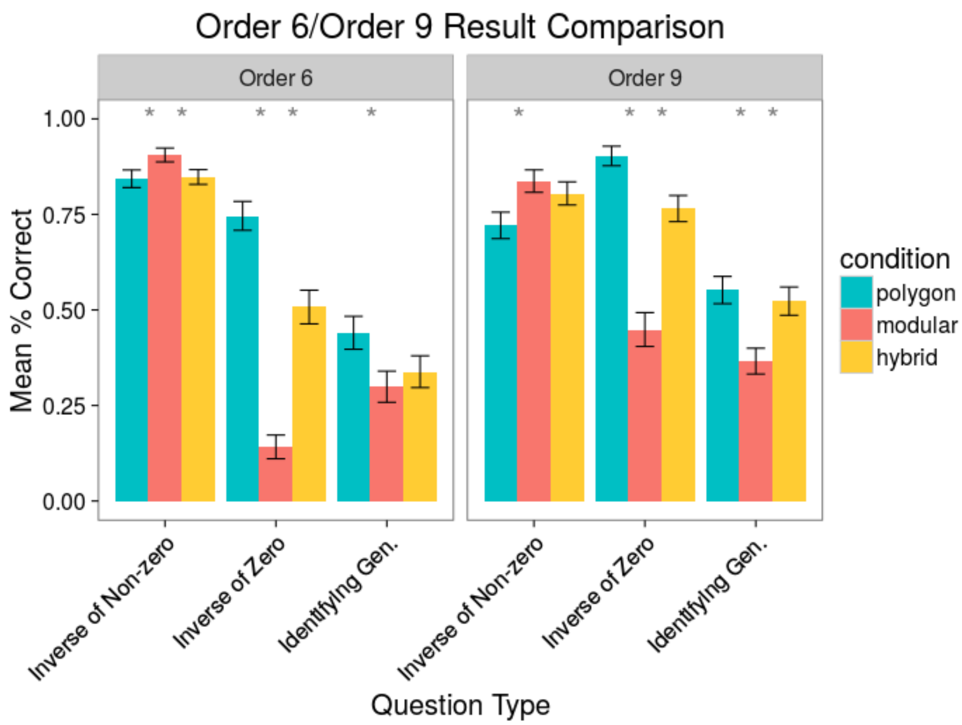


Figure 16. Results aggregated across experiments 2 & 3 (Exp. 1 had no hybrid condition, so we omitted it from this graph). This plot shows how the hybrid group participants, while not initially achieving best-of-both-worlds performance, appear to be much closer to achieving the benefits of both presentations later in the experiment.

full model is significantly better than any of these comparison models. To get an intuition for how strong the evidence is, the difference in log-likelihood is 15.97 between the full model and the next best model, meaning that the data are $e^{15.97} \approx 9$ million times as likely to have occurred under the full model (while this estimate does not include the compensation for the extra parameter that is taken into account in the BIC, the effect of the extra parameter is relatively small, and is swamped by the difference in log-likelihood). However, there are many other possible ways people could use the two representations beyond what we have modeled here (such as picking arbitrarily on each question), so further investigation is needed.

Similarly, with the experiment 3 data we estimated that $\theta = 0.39, \phi = 0.56$, so the data are best fit under this model by assuming that a little less than 40% of the participants are integrating, and those that aren't are choosing the polygon representation slightly more frequently than the modular. We used the Bayesian Information Criterion (BIC) to compare this model ($BIC = 3525.9$) to models where all participants chose modular ($BIC = 3826.5$), all chose polygon ($BIC = 3734.6$), where no participants integrated, i.e. a fixed $\theta = 0$ and fit $\phi = 0.56$ ($BIC = 3584.8$), and a model where all participants integrated, i.e. $\theta = 1.0$ ($BIC = 3700.2$). The comparisons again provide very strong evidence that the full model is significantly better than any of these comparison models. (Again, for intuition, the data are $e^{31.74} \approx 6.1 \cdot 10^{13}$ times as likely under the full model as the next best model, again swamping the penalty for extra parameters included in the BIC.) However, as above there are other possible ways that the participants 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 participants.

After running these analyses, we noticed that the hybrid group seems to be achieving much closer to best of both worlds performance in the group of order 9 (see Figure 16). This could be due to the fact that it takes some practice and/or feedback for the hybrid group to achieve the benefits of both presentations, or it could be because they are transferring more successfully than the participants in the other group. Either way, this effect might be crucial for evaluating the effectiveness of the hybrid presentation.

Because the change in performance between group orders for the hybrid participants was observed post-hoc, it is important to assess it carefully – and in particular to make sure that the effect is sufficiently strong to defray the concern that it is simply one of many possible patterns that might have arisen by chance. Accordingly, we ran an additional post-hoc analysis fitting these models on the subsets of the data from order 6 and order 9 separately to assess the strength of the evidence for greater integration later on. In accordance with Fig. 16, 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 with the best model using a single set of parameters for both group orders, we find the new model improves substantially ($BIC = 1438.8$, very strong evidence that the model fitting order 6 and order 9 separately is better; for intuition the data are $e^{118.6} \approx 3.4 \cdot 10^{51}$ times as likely under this model, which entirely dominates the penalty for the extra parameters in the BIC). In experiment 3, we estimated $\theta_6 = 0.22$, whereas $\theta_9 = 0.50$. As above, this substantially improves on the earlier model ($BIC = 3317.6$, very strong evidence that this model is better; for intuition the data are $e^{118.0} \approx 1.8 \cdot 10^{51}$ times as likely under this model as under the full model, which entirely dominates the penalty for the extra parameters in the BIC). This corroborates the idea that integration may have increased as the experiment went on, with only 20-30% of participants in the hybrid group appearing to acquire the benefits of both presentaions early on, but 50-60% doing so by the time they reached the order 9 material.

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; though the effect was rather small, the modular group performed significantly better at finding the inverse of non-zero elements. Thus our initial hypothesis that there would be differences in performance between the groups was confirmed.

We now consider in more detail how performance with the different presentations measures up to the criteria we proposed in the introduction:

- Does each presentation allow students to apply the directly-instructed concepts correctly? Does it allow them to transfer these concepts to a non-isomorphic group?

  Yes. Both the polygon and modular presentations produced excellent performance on the operation in the order 6 group, and were able to transfer with similar performance to the order 9 group.

- Does each presentation allow students to answer questions about further concepts that build upon the base concept, within the original instance? Does it allow them to transfer these concepts to a non-isomorphic group?

  Generally yes, with each presentation having advantages and disadvantages. Subjects were able to transfer these concepts to the group of order 9 reasonably well, performance on some types of problems increased and on others it decreased, but overall the advantages of each presentation remained the same.

- Does each presentation allow students to generalize about a class of instances? Does it allow them to express (or evaluate the truth of) these generalizations using formal mathematical expressions and language?

  Not very well. Both groups had fairly low success on these portions of the experiment, and there did not appear to be many differences between the groups on these questions.

Process differences underlying the performance differences. We have some hypotheses about the process differences that may underlie the pattern of results we observed, based on responses to problems where we asked the participants to explain their answers, and our post-hoc analyses of things like explicit use of the diagram:

Inverses: The modular group performed better at finding the inverses of elements other than zero, while the polygon group performed better at finding the inverse of zero. One possible explanation is that the modular presentation cued the participants to recognize an algorithm for finding most of the inverses: simply subtract the element from the group order. For example, under $+_6$, the inverse of 2 is 4, and $6 - 2 = 4$. We expect that the modular group participants would be more likely to recognize this relationship, since they perform subtraction much of the time when computing the group operation. By contrast, the polygon participants may have been less likely to infer this algorithm, using instead the less reliable strategy of counting around to 0. We hypothesize that the modular group outperformed the polygon group on computing the inverses of non-zero elements because they recognized the more efficient and accurate subtraction strategy.

Why would the modular subjects then do worse at the inverse of zero questions? Because this is the only case where the subtraction algorithm fails. The inverse of 0 under the operations we have defined is 0, but the subtraction algorithm gives the group order (which is not even an element of the group). A large majority ($> 75\%$ in all experiments and group orders) of the incorrect responses to the inverse of zero questions were the group order. (Note that this means that if we consider these answers to be correct, this particular advantage for the polygon group would go away, but this would reflect a different interpretation of what the elements of a cyclic group are.) It is interesting that this result is robust even after participants receive feedback on this question in the order 6 group explaining the correct answer; about half the participants who got the inverse of zero question wrong in the order 6 group persisted in their error in the order 9 group. This may suggest that the effect was sufficiently strong that one piece of feedback was insufficient to overcome it, or that some aspect of the intervening experience (such as using the subtraction algorithm on inverse of non-zero questions) may be reinforcing the error.

Generators: The polygon group performed better at identifying elements that are generators. We hypothesize that this is due to a spatial structure to the generator questions in the polygon case which may assist in solving them. For example, consider evaluating whether 5 is a generator on the nonagon. Adding 5 to itself repeatedly, we get the sequence $5 \to 1 \to 6 \to 2 \to \cdots$. 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; after a few steps the participant might observe that the pattern covers successive items on every other step (5, 6, ... on odd steps, 1, 2, ... on even 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).

Hybrid Group

The results of our meta-analysis suggest that by the time they reached the order 9 group, the hybrid group as a whole performed at a level that approaches the hoped-for “best of both worlds” performance. However, they did not appear to be achieving the full advantages of each group, at least initially. Our hierarchical modeling results suggest that this imperfect performance may be explained by some individual variation, with some participants picking just one representation, while others achieved the benefits of both. It also suggests that, encouragingly, the number of participants 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 overall, the results seem to suggest that given sufficient practice with it, the hybrid representation might be beneficial for most participants.

Why might the hybrid presentation be beneficial, and why might these benefits emerge more slowly than in the single presentation experimental groups? One possibility is that the benefits of the hybrid group come from a slower process of integration or coordination (Schwartz & Goldstone, 2015) of the different ways of thinking about the problem. There are a number of possible forms of integration that might occur:

- Learning which of the representations is best used for which types of problems.
- Transferring concepts learned using one representation to the other.
- Creating a unified single representation that incorporates aspects and benefits of the individual presentations, as well as the relationships between them.

These possibilities are neither exhaustive nor mutually exclusive. We find it likely that the unified representation could be beneficial in at least some circumstances, and could possibly give even better than best-of-both-worlds performance. (For example, consider for the identifying generator questions combining the spatial intuitions of the polygon presentation with the computational reliability of the modular presentation.) Thus we attempted to encourage unification of the representations through the question where we asked participants to reflect on the relationship between the different ways of thinking about the operation. However, our data do not provide the ability to fully dissociate which types of integration were occuring for our participants.

Furthermore, in practice, there may be some heterogeneity in the type of integration that occurs even within one experiment, depending on the type of representations being integrated; consider the pattern of effects of representation use on performance we observed in the representation-use questions. On some question types, such as finding the inverse of non-zero elements, it appears that most hybrid group participants have transferred or unified their understanding between presentations sufficiently so that using either representation is equally beneficial. On other questions, such as identifying generators, one representation is still much more beneficial than the other. 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 easily transferable between representations. 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.

Along these lines, we note that between the polygon and modular presentation, the polygon presentation seems overall more advantageous. It is possible that this is due to the hybrid elements inherent in the polygon presentation. By including both the visuospatial presentation and numbers as symbols, it may cue participants to recognize some of the arithmetical patterns that are more explicitly explained in the modular presentation. This might explain why the polygon presentation group did not perform too much worse than the modular presentation group even when thinking in an arithmetic way seemed more useful.

It remains a question for future research how the choice of presentations affects what type of integration occurs. Hopefully exploring this would shed light on how the hybrid presentation could be altered to encourage more uniform improvement across all question types and participants. Nevertheless, our results suggest that teaching multiple presentations may be beneficial to students' overall understanding.

"Hybrid" presentations in previous work. The reader might notice that some of the experimental groups in the work of Kaminski and colleagues could be viewed as having hybrid-like elements. Specifically, some groups of participants saw multiple distinct presentations (e.g. both fractional cups of liquid and fractional slices of pizza), and had their attention implicitly or explicitly drawn to the connections between them. It is important to clarify that to be beneficial, a hybrid presentation must be constructed from distinct presentations with distinct advantages. That is, the benefits of distinct presentations will generally be increased when they support complementary aspects of understanding. The different concrete presentations in the work of Kaminski et al., though diferent in surface details, can be seen as drawing on the same numerical intuitions that are unhelpful when participants are confronted with one of their generic presentations. We would expect that by combining one of the concrete and one of the generic presentations used by Kaminski and colleagues, one might be able to achieve better transfer performance more broadly (for example, seeing the generic presentation might allow participants to transfer to Kaminski and colleagues' original transfer task, while seeing one of the concrete presentations might give better ability to generalize to a new group order or a new concrete instantiation that shares the numerical structure of the taught concrete presentations).

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. For example, despite the fact that performance on the inverse questions was around 75% on average, only about one third of the participants were able to articulate the general formula for computing an inverse in a cyclic group of unspecified order n. This may reflect the fact that this study was too short for participants understanding of specific groups to develop into a more general understanding, or it may be because the processes of reasoning explicitly or formally and the processes of reasoning implicitly and/or procedurally are not perfectly linked, and some effort is needed to go from one to the other (e.g. Anderson, 1996; Reber, 1967; Davidson, Eng, & Barner, 2012). This may also explain why the advantages of the presentations didn't transfer to the order n questions – it may be difficult for many participants without formal mathematical training to give an explicit formula using the group order as a variable even if they can apply the subtraction procedure to find the inverse of an element in a cyclic group of arbitrary order n. Similarly, it may be difficult to use spatial intuitions that support detecting that a particular element is a generator on a polygon of specific size when thinking about all possible polygons. The process of learning to make this mapping between procedures that might be applicable to any specific instance of a general class of problems and a formal understanding of the generalization itself may be a skill that is generally acquired through experience working with and evaluating formal mathematical reasoning.

Because of the complexity of these issues, it's difficult to say how we could most easily improve generalization and formalization performance. 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 participants 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 participants 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 that examines how formalization interacts 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 participants to formalize their understanding of inverses in the cyclic group of order 6, and then later ask participants to give an 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 over days or months. It has been suggested that across many domains, the brain relies on complementary learning systems which learn at different rates, and in particular that cortex must assimilate knowledge slowly to avoid catastrophic interference (Kumaran, Hassabis, & Mcclelland, 2016). Because the progression from concept introduction to final assessment of understanding occurs in about an hour in our experiment, we may be short-changing the presentations by not allowing the participants enough 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, which might accelerate learning from the beginning.)

It is interesting to ask whether simple practice with a concept can lead to formalization, and if so, under what circumstances. What sort of introspection about the processes they are performing is necessary for this insight to arise? Do students acquire these insights suddenly after practicing for a while, or does formal understanding emerge by a more gradual learning process, moving from an inarticuable intuition to an understanding that can be explained in words and finally to a formal expression of an idea? It has been suggested previously that this sort of graded transition from implicit to conscious knowledge can occur (Cleeremans & Jiménez, 2002). How does this transition 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 presented, and of concepts related to it, using elementary group theory as our test domain. We found that even if two presentations produce equal ability to apply the directly instructed 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. These findings contribute to the ongoing exploration of the role 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 multiple presentations while encouraging participants to develop an integrated understanding which both presentations contribute to. Our data suggests that this may have positive effects even if the total instruction time is the same. By the end of the experiment, participants 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 participants 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 (masked)

References

Anderson, J. R. (1996). A Simple Theory of Complex Cognition. *American Psychologist*, *51*(4), 355–365. doi: 10.1037//0003-066X.51.4.355

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. Retrieved from `http://pps.sagepub.com/lookup/doi/10.1177/1745691610393980` doi: 10.1177/1745691610393980

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*(22), 3127–3131. Retrieved from `http://doi.wiley.com/10.1002/1097-0258(20001130)19:22{\%}3C3127::AID-SIM784{\%}3E3.0.CO;2-M` doi: 10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making (frontiers of cognitive science).*

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162–173. Retrieved from `http://dx.doi.org/10.1016/j.cognition.2011.12.013` doi: 10.1016/j.cognition.2011.12.013

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 Kaminski, 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

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness Fading in Mathematics and Science Instruction: A Systematic Review. *Educational Psychology Review*, *26*(1), 9–25. doi: 10.1007/s10648-014-9249-3

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. doi: 10.1016/0010-0285(83)90002-6

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

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kumaran, D., Hassabis, D., & Mcclelland, J. L. (2016). What Learning Systems do Intelligent Agents Need ? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, *20*(7), 512–534. Retrieved from http://dx.doi.org/10.1016/j.tics.2016.05.004 doi: 10.1016/j.tics.2016.05.004

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

Reber, A. s. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863.

Schwartz, D. L., & Goldstone, R. L. (2015). Learning as Coordination. In *Handbook of educational psychology* (3rd ed., chap. 5).

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 Results

Operation

Experiment 1: There was no significant difference in the performance on the basic operation questions between the experimental groups (see Figure 17) in either the order 6 group ($b = 0.006$, $t(235) = 0.18$, $p = 0.86$) or the order 9 group ($b = -0.009$, $t(4178) = -0.26$, $p = 0.79$). These are the questions where the wording varied in accordance with the different presentations 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 on the basic operation questions between the modular and polygon experimental groups (see Figure 18) in either group order (order 6: $b = 0.02$, $t(480) = 0.735$, $p = 0.46$; order 9: $b = 0.02$, $t(6030) = 0.63$, $p = 0.53$). Furthermore, there was no significant difference between the hybrid and modular groups (order 6: $b = -0.02$, $t(449) = -0.72$, $p = 0.47$; order 9: $b = 0.04$, $t(6042) = 1.10$, $p = 0.27$).

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 19) in either group order (order 6: b 95%-CI $= [-2.55, 0.15]$, $p > 0.05$; order 9: b 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: b 95%-CI $= [-2.13, 0.13]$, $p > 0.05$; order 9: b 95%-CI $= [-1.90, 0.17]$, $p > 0.05$).

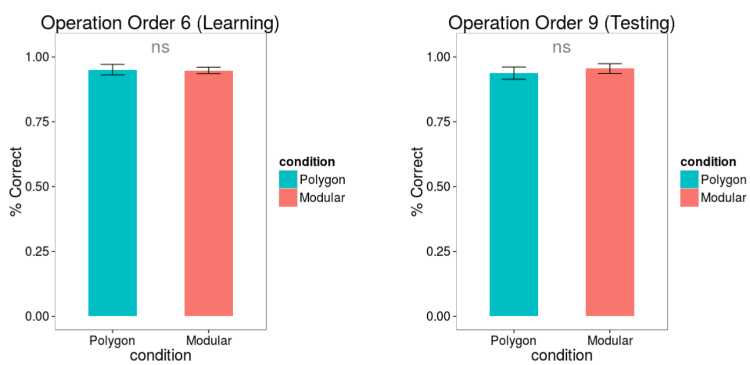


Figure 17. Experiment 1 – operation results

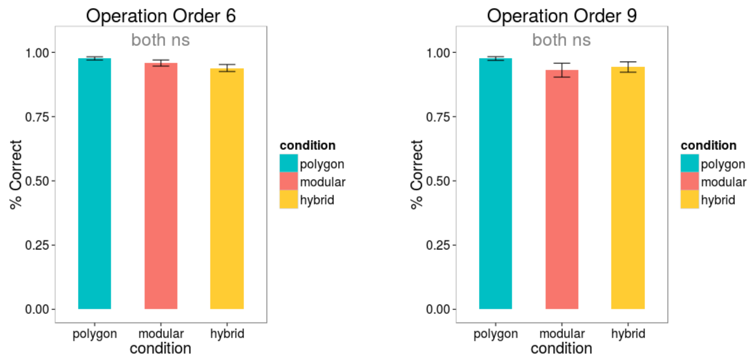


Figure 18. Experiment 2 – operation results

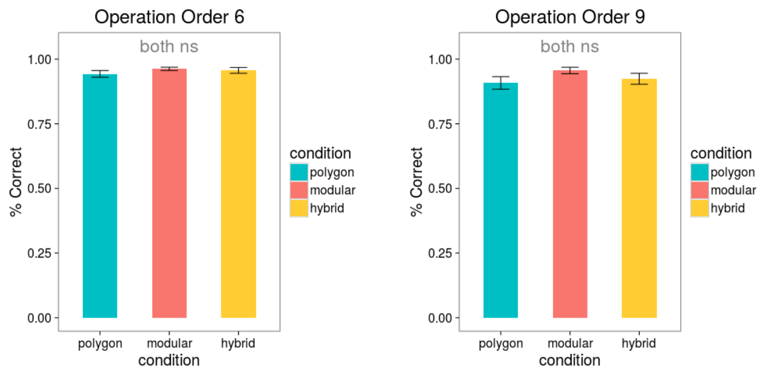


Figure 19. Experiment 3 – operation results

Inverses

Experiment 1: For inverses the polygon group was significantly worse at finding the inverse of non-zero elements in the order 9 group ($b = -0.169$, $t(4178) = -3.73$, $p < 0.001$), but not significantly worse in the order 6 group ($b = -0.04$, $t(4178) = -1.01$, $p = 0.31$), and was significantly better at finding the inverse of zero (see Figure 20) in the order 6 group ($b = 0.66$, $t(4178) = 9.13$, $p < 0.001$). (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 ($b = -0.11$, $t(4178) = -1.99$, $p = 0.047$).

Experiment 2: We replicated our results that the modular group performed better at inverses of non-zero elements (order 6: $b = -0.11$, $t(6026) = -2.59$, $p = 0.009$; order 9: $b = -0.07$, $t(6026) = -1.48$, $p = 0.13$), and the polygon group performed better at inverse of zero questions (order 6: $b = 0.68$, $t(6024) = 8.842$, $p < 0.001$; order 9: $b = 0.44$, $t(6024) = 5.80$, $p < 0.001$). (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 ($b = -0.05$, $t(6025) = -0.94$, $p = 0.35$). The hybrid group did not perform significantly worse than the modular group at inverses of non-zero elements (order 6: $b = -0.007$, $t(6033) = -0.173$, $p = 0.86$; order 9: $b = 0.08$, $t(6031) = 1.56$, $p = 0.12$), and did perform significantly better at inverses of zero (order 6: $b = 0.46$, $t(6027) = 5.95$, $p < 0.001$; order 9: $b = 0.38$, $t(6027) = 4.91$, $p < 0.001$). 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 ($b = 0.01$, $t(6029) = 0.125$, $p = 0.90$). (See Figure 21.)

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: b 95%-CI $= [-1.14, 0.24]$, $p > 0.05$; order 9: b 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: b 95%-CI $= [1.93, 3.55]$, $p < 0.05$; order 9: b 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 (b 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: b 95%-CI $= [-1.92, 0.01]$, $p > 0.05$; order 9: b 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: b 95%-CI $= [0.21, 2.65]$, $p < 0.05$; order 9: b 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 (b 95%-CI $= [-1.32, 0.78]$, $p > 0.05$) (See Figure 22.)

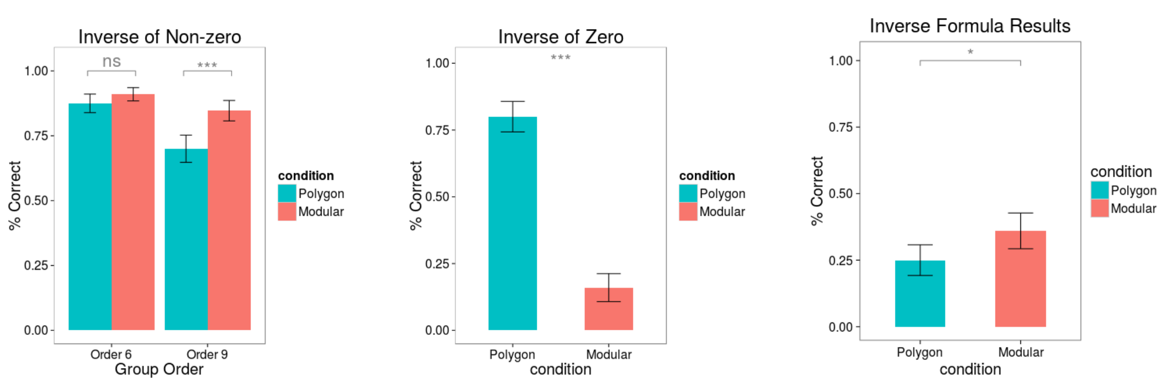


Figure 20. Experiment 1 – inverse results

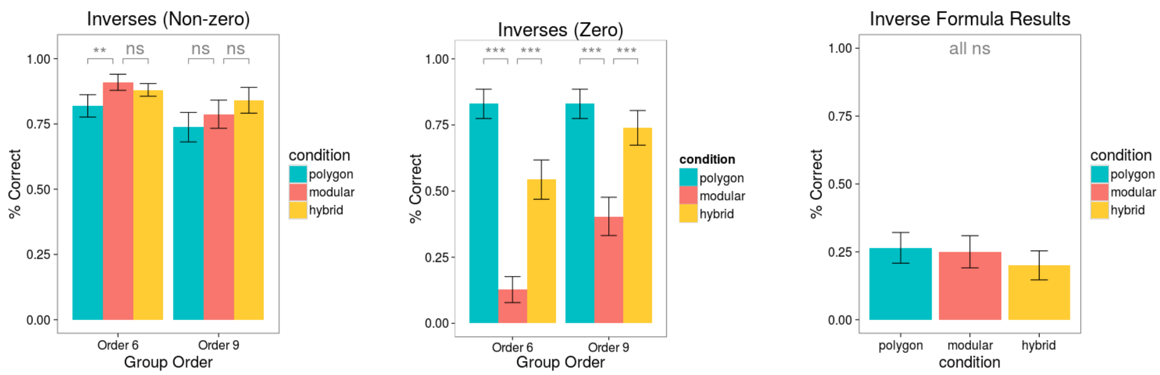


Figure 21. Experiment 2 – inverse results

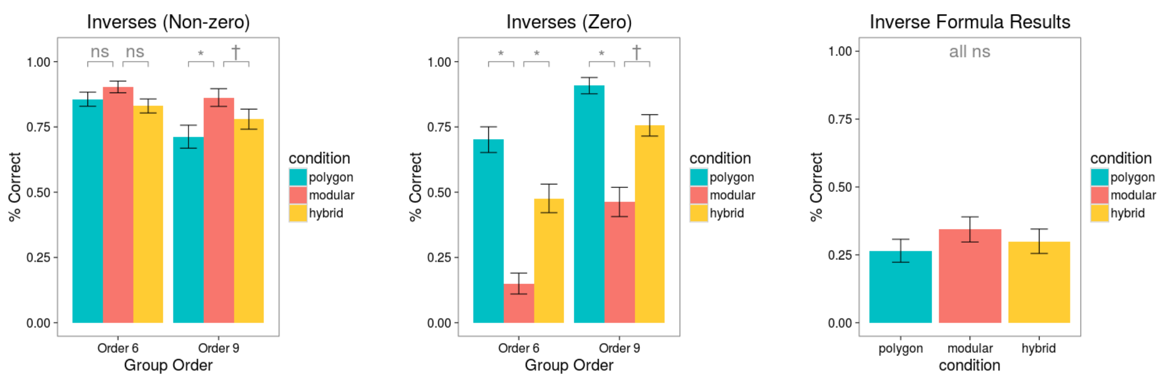


Figure 22. Experiment 3 – inverse results

Generators

Experiment 1: The polygon group performed significantly better at identifying elements that were generators (order 6: $b = 0.164$, $t(4178) = 2.27$, $p = 0.02$; order 9: $b = 0.184$, $t(4178) = 3.45$, $p < 0.001$). 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: $b = 0.01$, $t(4178) = 0.235$, $p = 0.81$; order 9: $b = 0.12$, $t(4178) = 2.32$, $p = 0.02$). The modular group performed better at the T/F questions about generators in the order n group ($b = -0.11$, $t(4178) = -2.60$, $p = 0.009$). However, performance on these questions was not too far above chance. The groups did not differ significantly on the A/S/N questions ($b = -0.03$, $t(4178) = -0.5$, $p = 0.62$). (See Figure 23.)

Experiment 2: We replicated our result that the polygon group is better at identifying generators only in the order 9 group (order 6: $b = 0.00$, $t(6024) = -0.05$, $p = 1.00$; order 9: $b = 0.19$, $t(6025) = 3.37$, $p < 0.001$), and the polygon group did not perform better at identifying non-generators in either group (order 6: $b = -0.03$, $t(6026) = -0.60$, $p = 0.55$; order 9: $b = 0.07$, $t(6025) = 1.31$, $p = 0.19$). We replicated our result that the modular group performed better at T/F questions about generators ($b = -0.12$, $t(6026) = -2.71$, $p = 0.007$). We replicated our result that the polygon and modular groups did not significantly differ at the A/S/N questions ($b = 0.01$, $t(6026) = 0.139$, $p = 0.89$). The hybrid group perfomed significantly better than the modular group at identifying generators in the group order where our previous finding replicated ($b = 0.25$, $t(6029) = 4.31$, $p < 0.001$), did not significantly differ at identify non-generators (order 6: $b = 0.07$, $t(6031) = 1.51$, $p = 0.13$; order 9: $b = 0.05$, $t(6029) = 0.80$, $p = 0.42$), and did not perform significantly worse at the T/F questions ($b = -0.042$, $t(6033) = -0.98$, $p = 0.33$). 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 ($b = -0.00$, $t(6031) = -0.01$, $p = 0.99$). (See Figure 25.)

Experiment 3: We replicated our result that the polygon group is better at identifying generators (order 6: b 95%-CI $= [0.35, 1.72]$, $p < 0.05$; order 9: b 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: b 95%-CI $= [-0.29, 0.50]$, $p > 0.05$; order 9: b 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 (b 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 (b 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: b 95%-CI $= [-0.37, 1.07]$, $p > 0.05$; order 9: b 95%-CI $= [-0.04, 0.95]$, $p > 0.05$). We replicated our result that the hybrid group did not significantly differ at identifying non-generators (order 6: b 95%-CI $= [-1.17, 0.60]$, $p > 0.05$; order 9: b 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 (b 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 (b 95%-CI $= [-1.26, 0.53]$, $p > 0.05$). (See Figure 27.)

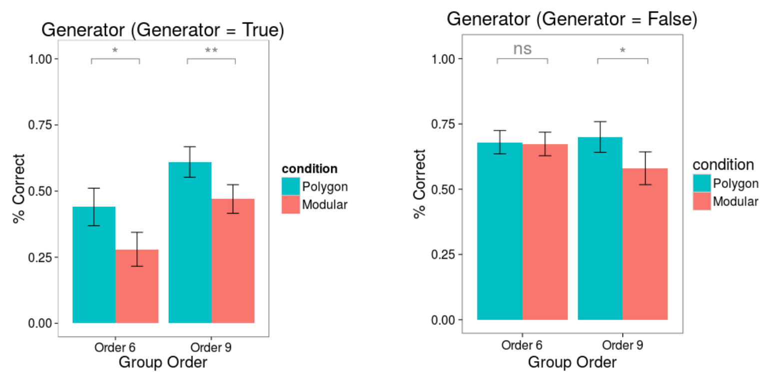


Figure 23. Experiment 1 – generator results

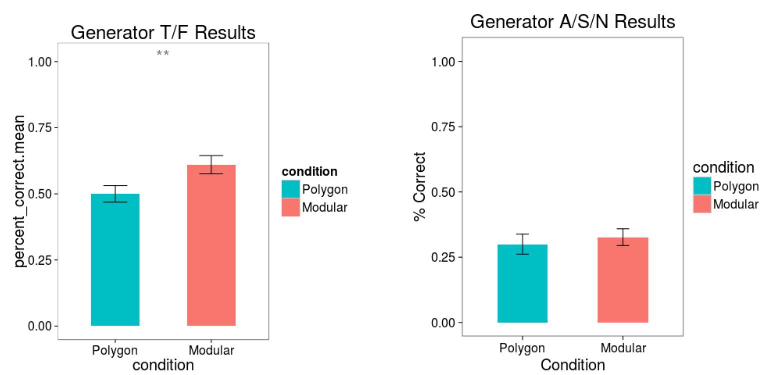


Figure 24. Experiment 1 – generator T/F & A/S/N results

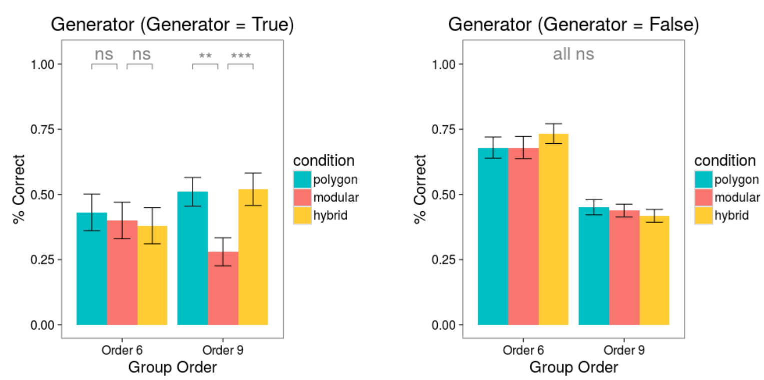


Figure 25. Experiment 2 – generator results

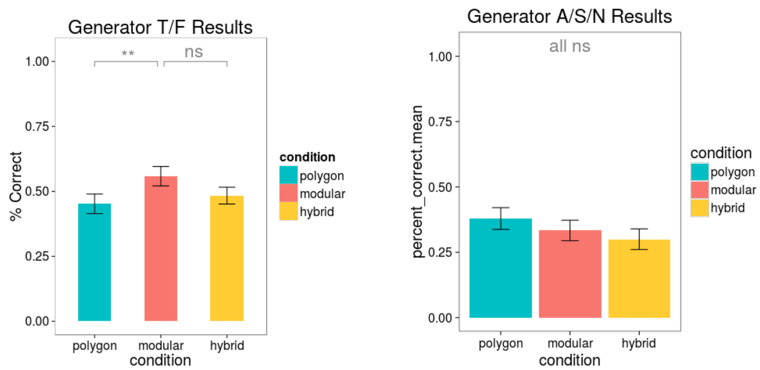


Figure 26. Experiment 2 – generator T/F & A/S/N results

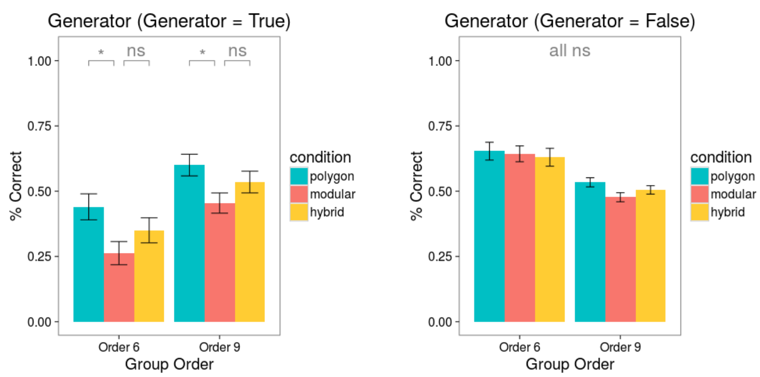


Figure 27. Experiment 3 – generator results

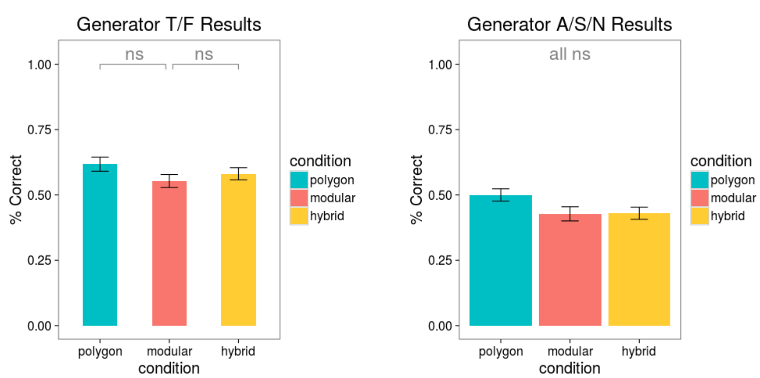


Figure 28. Experiment 3 – generator T/F & A/S/N results

T/F & A/S/N Result Detail

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?"

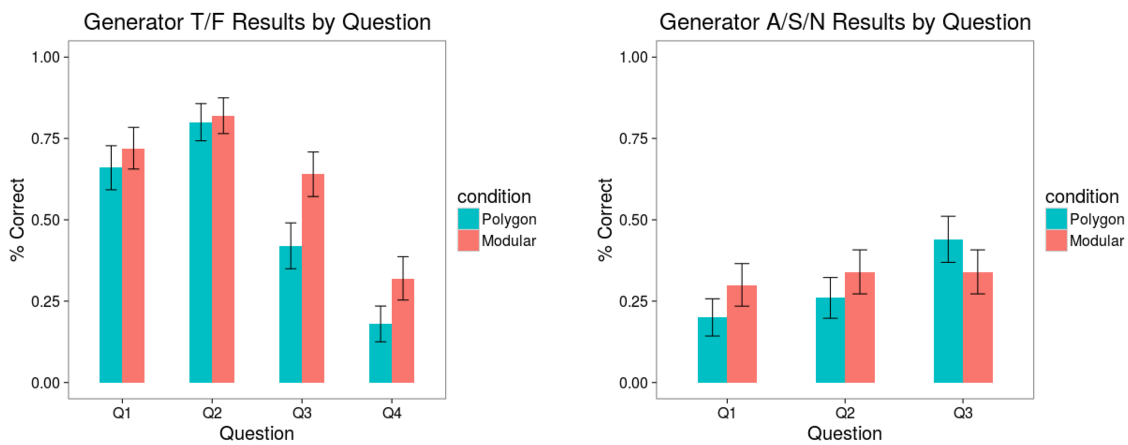


Figure 29. Experiment 1 – T/F & A/S/N results by question

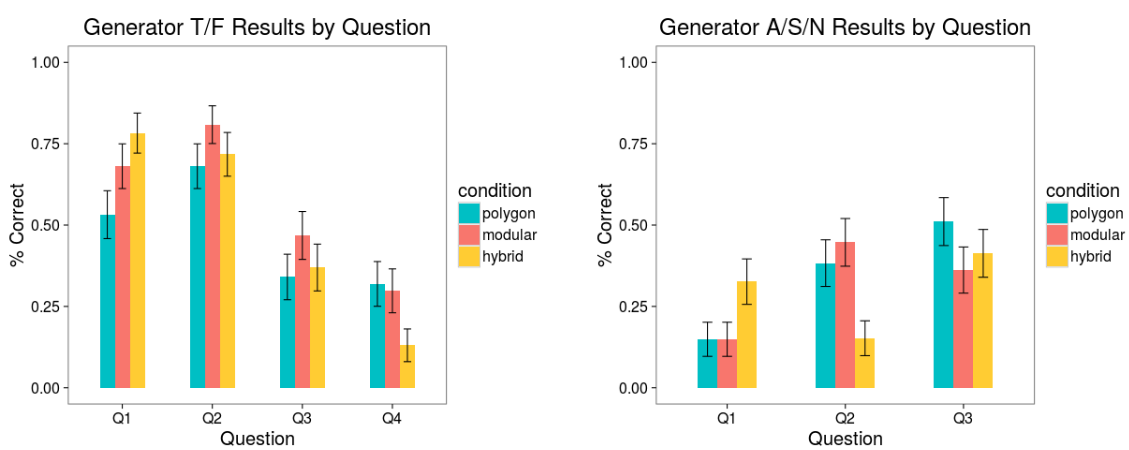


Figure 30. Experiment 2 – T/F & A/S/N results by question

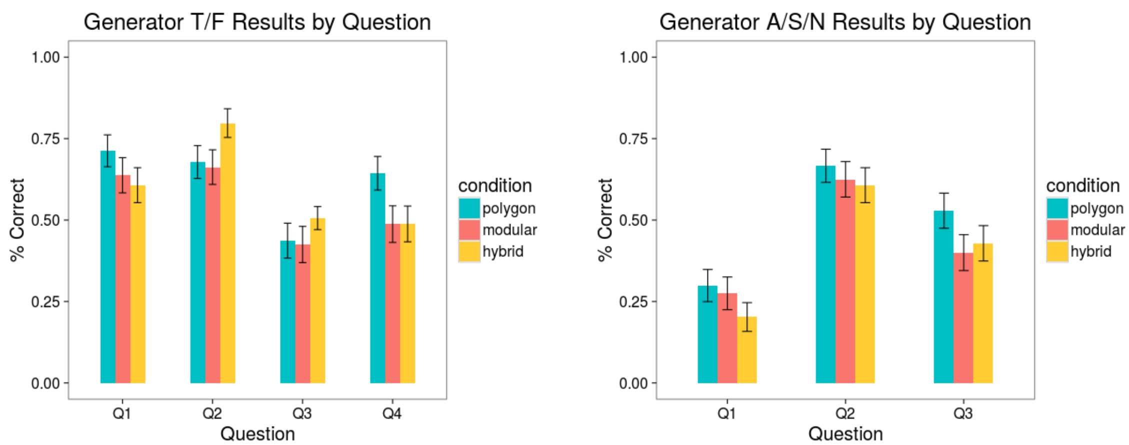


Figure 31. Experiment 3 – T/F & A/S/N results by question