Different Presentations of a Mathematical Concept Can Support Learning in Complementary Ways

Andrew K. Lampinen and James L. McClelland
Stanford University

Previous research has found that different presentations of the same concept can result in different patterns of transfer to isomorphic instances of that concept. Much of this work has framed these effects in terms of advantages and disadvantages of concreteness or abstractness. We note that mathematics is a richly structured field, with deeply interconnected concepts and many distinct aspects of understanding of each concept, and we discuss difficulties with the idea that differences among presentations can be ordered on a concrete-abstract dimension. To move beyond this, we explore how different presentations of a concept can affect learning of subsequent concepts and assess several distinct aspects of understanding. Using the domain of elementary group theory, we teach adult participants a group operation using a visuospatial or an arithmetic presentation. We then teach them concepts that build upon this operation. We demonstrate that our presentations differentially support learning complementary aspects of the system presented. We argue that these differences arise from the fact that each presentation supports learning by connecting to different systems of reasoning learners are already familiar with, and that it is these connections to extant knowledge systems, rather than differences in concreteness versus abstractness, that determine whether a presentation will be helpful. Furthermore, we show that presenting both presentations and encouraging participants to recognize the relationship between them improves performance without requiring additional time, at least for some participants.

Educational Impact and Implications Statement
The details of how a concept is taught can have far-reaching effects on students' learning. Using abstract algebra with adult subjects, we show that 2 presentations of a concept that connect to different types of students' prior knowledge can have advantages and disadvantages for later learning that builds on the target concept. We show that one possible solution to the dilemma of choosing which presentation to use is giving students both concepts and explaining how they are related to each other. In summary, when designing pedagogical materials, we should consider not only how they affect learning of the present concept, but also how they support learning of future concepts, and use multiple complementary presentations rather than searching for a single ideal one.

Keywords: presentations, abstraction, generalization, mathematical cognition

Supplemental materials: http://dx.doi.org/10.1037/edu0000235.supp

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? These are central issues for a science of pedagogy and education. We explore them here in the context of learning basic concepts in abstract algebra.

First, we must define our terms. We use the term "presentation" to refer to the details of the pedagogical materials, in contrast to the more common "external representation," to highlight a few distinctions. First, the presentation encompasses more than just an external representation, it also includes the pedagogical explanations used to describe the external representation and the practice problems that students are given. Second, presentations are provided by the curriculum, while external representations could be created in some other way, for example, by a student drawing a diagram. We reserve the term "representation" to refer to participants' mental representation of the concepts, which, while grounded in the presentation used, may differ from it in important ways.

As the name suggests, presentations are generally used to present a 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. These features may highlight or obscure certain aspects of the concept in question. In addition, students may have some prior knowledge 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, Sloutsky, & Heckler, 2008), as we will discuss more fully below.

In this article, we elaborate upon this work in three ways. We explore:

1. The effects of presentations on learning related concepts and other aspects of understanding.

2. How to characterize the ways presentations differ from each other.

3. The potential benefits of combining presentations to achieve the advantages of both.

First, how do presentations of a concept affect learning of 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 earlier concepts affect how later concepts are learned, has been considered for a long time within cognitive psychology (e.g., Bransford & Schwartz, 1999; Fischer, 1980), and more specifically within mathematical cognition, (e.g., Hazzan, 1999; Richland, Stigler, & Holyoak, 2012). Mathematics education research has also highlighted connections between concepts as an essential aspect of what it means to understand (e.g., Hiebert et al., 1997). Thus, it is important to consider the effect of presentations beyond the single concept being presented. For instance, teachers often rely on previously learned concepts to teach a new idea, and students rely on previously learned concepts to reason about it (Hazzan, 1999). Thus, it is possible that the presentations used to teach a concept can also influence students' understanding of later concepts that are related to it. Furthermore, there are often many aspects of understanding of a concept; for example, thinking of it as a process versus an object (Hazzan, 1999), or thinking intuitively versus formally, which may be influenced by features of presentations.

Second, how should we think about the factors that vary between different presentations? The work of Kaminski et al. (2008) and much of the following work has focused on a single axis of concrete (or grounded) versus abstract (or generic). Some previous work has acknowledged that there are concrete features that may be irrelevant to student learning, such as decorative features like toppings on pizza slices when learning fractions (e.g., Belenky & Schalk, 2014). This raises many questions: what types of concrete features are relevant, and which are irrelevant? Is there really only a single concrete versus abstract axis that affects what is learned from a presentation, or is the space of possible presentations more complicated? Are "concrete" and "abstract" broadly useful terms for describing the features that differ between presentations, or are they just a proxy for other factors? Is there an important distinction between abstract and "generic" or "idealized" presentations? Can

presentations of a concept that connect to different types of knowledge result in different types of understanding? Some previous work would suggest that this can occur, at least with large differences in pedagogical strategy (Nokes & Ohlsson, 2005).

Finally, is there any advantage to combining presentations? It has been known for some time that seeing multiple distinct examples can lead to better generalization in some cases (Gick & Holyoak, 1983). Previous work has found that using multiple external representations may be beneficial, although it also imposes higher cognitive demands and can even be detrimental (Ainsworth, 2006; Rau, 2016), and in particular that fading details from concrete to abstract may be beneficial (Fyfe, McNeil, Son, & Goldstone, 2014; Goldstone & Son, 2005). However, when considering the full ramifications of a presentation, including how it affects learning of other concepts, will multiple presentations still be beneficial?

In this project, we explored these issues and found evidence that two different presentations of a concept can have differential advantages, supporting different aspects of students' later understanding. We argue that the two presentations we used cannot be clearly ordered on a single concrete versus abstract axis. Instead, they link the concept to different systems that learners may have prior experience with. Building on this, we explored the possibility that exposure to both presentations might allow students to benefit from the advantages of both. 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.

We examined these issues within the area of elementary group theory, specifically cyclic groups (also used by Kaminski et al. (2008)) and a few concepts about them. An introduction to the relevant concepts is provided in the Appendix, 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 that 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).

The question we attempt to address in this article is how learning of these concepts can be altered by the way the group operation is presented. In the next section, we introduce some of the background needed to explore this question, by examining some of the ways that concepts are related to one another in mathematics.

Relationships Among Mathematical Concepts

How are concepts related to each other in mathematics, and how does this affect mathematical cognition? Some mathematical ed-

ucators have suggested that relationships among concepts are so fundamental to mathematics that the definition of understanding a mathematical concept should be "[seeing] how it is related or connected to other things we know" (Hiebert et al., 1997). Indeed, it has been suggested that students viewing mathematical concepts as isolated rules and procedures, rather than coherent systems of related ideas, is one of the fundamental problems in mathematics education (Richland et al., 2012). There are many kinds of relationships between mathematical concepts, some of which underlie abstract fields 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 the operations.

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., "$5 + ? = 11$" to "$5 + x = 11$, 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), that is, 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. 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 and colleagues 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, and so forth. They might also be asked to generalize this understanding to nonisomorphic 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.

Indeed, it has been noted for some time that the term "understanding" does not have a single meaning within mathematical cognition, but rather can refer to factors such as conformance to a rule when solving problems procedurally, explicit awareness of the rule, ability to transfer the rule to an analogous situation, and so

forth, and that the inferences we make about students level of understanding can depend upon which of these features we use to evaluate it (Bisanz & LeFevre, 1992). For instance, Greeno and Riley (1987) show that students can possess the ability to execute a procedure without having the ability to articulate the rules that the procedure follows. To address this issue, Bisanz and LeFevre (1992) created a framework for analyzing assessments of understanding, based on two features: activity and generality.

Activities vary from applying a procedure on a task to justifying a procedure or evaluating a presented procedure, while generality varies according to how broadly the procedure, justification, or evaluation is applied. This work implies that it is important to not focus solely on evaluating a single feature of what has been learned when evaluating a mathematical presentation. Mathematical education will ideally convey many aspects of understanding, not just a single one.

Concrete Versus Abstract?

What features of presentations might allow them to convey multiple aspects of understanding well? There has been a recent focus on the effects of concrete (or grounded) and abstract (or idealized or generic) materials on learning in mathematics (e.g., Belenky & Schalk, 2014; Kaminski et al., 2008). The broad consensus has been that concrete presentations offer benefits for initial learning and performance, while abstract presentations are beneficial for certain types of transfer (Belenky & Schalk, 2014). Thus, we might hope that more abstract presentations would be beneficial for other aspects of understanding as well.

However, there are reasons to think that the abstract-concrete distinction may be misleading. Schwartz and Goldstone (2015) have argued that rather than thinking of dichotomies in education as an "either-or problem," it's often better to focus on coordinating different learning processes—"rather than choosing one or the other, the best strategy is to choose both." They explicitly reference the abstract-concrete dichotomy, and indeed, some work has shown the benefits of "fading" from concrete presentations to more abstract ones (Fyfe et al., 2014; Goldstone & Son, 2005).

We believe that concrete versus abstract is a false dichotomy more generally, not just because both sides are useful, as the concreteness fading work shows and Schwartz and Goldstone (2015) argue, but because the space of presentations is much more complicated than this description would suggest. One aspect of this is the fact that the concepts of concrete and abstract are themselves difficult to define. For example, some authors have argued that "concreteness is not a property of an object but rather a property of a person's relationship to an object" (Wilensky, 1991). However, this perspective on concreteness seems to elide the physicality that the word commonly conveys. The appropriate definition of concrete may be hard to find. There may be many ideas conflated under this single word.

Similarly, there are many different possible meanings of the word abstract. Hazzan (1999) describes several of these, ranging from abstract as approximately the opposite of concrete as defined by Wilensky (1991) to abstract as an object rather than a process understanding of a concept. Which of these is meant when Kaminski et al. (2008) say that there is an advantage of abstract examples? Is there an important distinction to be drawn between abstract and generic pedagogical materials, and if so, why are these

terms used inconsistently in the literature (Belenky & Schalk, 2014; De Bock, Deprez, Van Dooren, Roelens, & Verschaffel, 2011; Kaminski et al., 2008)?

To avoid these definitional conundrums, instead of thinking of presentations in terms of concrete and abstract, we prefer to think about how presentations may support links to systems of thinking that students already understand. We believe this feature is more relevant to understanding the effects of presentations.

For example, from the perspective of students just learning arithmetic, the set of all integers would be an incredibly abstract concept. Yet, from the perspective of the mathematics students studied by Hazzan (1999), who have experience reasoning with various infinite sets of numbers, this concept is a useful basis for thinking about more formal ideas. The relevant feature of these concepts is not their "concreteness," but instead the students' relationships to the concepts.

When we take the perspective that what matters is how concepts relate to students' extant systems of knowledge, it becomes clear that because students have rich experience, there can be many different presentations of a concept that relate it to the many different reasoning systems that the students are familiar with. These systems can be quite diverse; a presentation may relate a concept to visuospatial thinking (Rau, 2016), to more basic mathematical concepts the students understand (Hazzan, 1999), to physical understanding like measuring cups (Kaminski et al., 2008), or embodiment (Nathan, 2008). Some of these presentations may be concrete in some sense, but it is not clear how to order them on a single concrete-abstract dimension. Indeed, it is not clear that we should, because presentations that relate a concept to different systems of understanding may have quite different support for learning of later concepts. For example, the training and transfer presentations that Kaminski and colleagues used are both concrete in some sense, but they ground the concept in very different ways that we argue support different types of understanding. In the next section, we explore this in more detail.

The Advantage(?) of Abstract Examples

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 1. 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 1, where the objects of the group are grounded as toys in a children's game. 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 results of the game. 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 con-
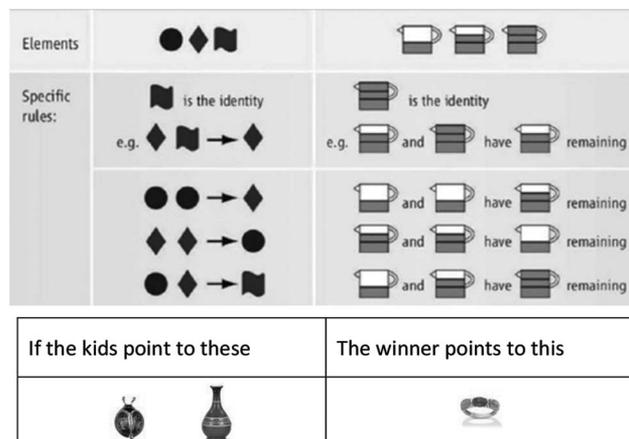

Figure 1. Presentations from Kaminski, Sloutsky, & Heckler (2008). Figures adapted with permission from "The Advantage of Abstract Examples in Learning Math" by J. A. Kaminski, V. M. Sloutsky, & A. F. Heckler, 2008, Science, 320, Supplemental Material.

cluded 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 instantiation with a different contextualization." Thus, he concludes that "The transfer task is more similar to the generic instantiation than to the concrete ones." In a response to this interpretation, Kaminski, Sloutsky, and Heckler (2009) asserted that the generic and transfer domains were not more similar, because after describing the domains to a set of 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 that are represented as one-third of a cup or two-thirds of a cup in the concrete instantiation. Although from a mathematical perspective it is clear that the generic, numeric, and transfer presentations are isomorphic, in the generic and transfer presentations the symmetry between the two nonidentity elements is clear—circle circle = diamond, and diamond diamond = circle—whereas in the numeric presentations the symmetry is broken. 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 one-third cup two times makes two-thirds cups, but $2 + 2 = 1$ because two-thirds cup two times makes 1 and

one-third cups, *and* we throw away the full cup to get back to one-third). We suspect this asymmetry may be to blame for the worse transfer performance, because 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 numeric examples would probably have been biased to choose 1 as a generator, even though 2 is an equally good choice, whereas in the generic or transfer case there would be no such bias. The numeric presentations provide a shared basis (number of identified parts, be they tennis balls, slices of pizza, or one-third 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. Although the transfer example is quite concrete in that it relates to a physical game played with physical objects, it does not naturally support this numerical interpretation. Different ways of presenting a concept may support different types of understanding about it.

This idea is supported by De Bock et al., in their replication of Kaminski's study (De Bock et al., 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 a generic to another generic. Thus, each presentation was better for transferring to presentations that were similar in terms of whether or not they supported a mapping to number. The relevant feature for transfer was not the concreteness of the presentations, but what systems of thinking they connected to and what types of reasoning they supported.

Furthermore, De Bock and colleagues 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. Since Hazzan (1999) showed that students were relating more abstract concepts back to their understanding of simpler ones, the inferences they make about the simpler concepts could provide differential preparation for future learning (Bransford & Schwartz, 1999) of more advanced concepts. Thus, in pedagogically more realistic scenarios where students are asked to learn about a system of related concepts, different presentations of a concept may have effects that propagate to other related concepts.

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 some concepts may have been implicitly communicated 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 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, and students are not assessed on a single outcome, but rather on their grasp of multiple aspects of understanding.

Criteria for Presentations

To more thoroughly assess a presentation of a concept, we propose the following set of questions for evaluating its impact on various aspects of understanding. 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. We think that they will provide a useful basis for beginning to consider the effects of how material is presented more broadly in mathematics education. 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 nonisomorphic 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 nonisomorphic 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? Does it allow them to express (or evaluate the truth of) these generalizations using formal mathematical expressions and language? (Can they explain in words how to find inverses in an arbitrary group? 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?)

These points attempt to span varied types of understanding, from the most procedural and implicit to the most formal and explicit. 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. From the perspective of Bisanz and LeFevre (1992), Kaminski and colleagues assessments of understanding occurred at the "application" activity level. Their participants may have developed other aspects of understanding (e.g., 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 find that two presentations that

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.

Multiple Presentations

If we have multiple presentations of a concept, each with unique advantages, which should we use for instruction? Instead of forcing ourselves to choose one and lose the benefits of the other, we suggest "choosing both" (Schwartz & Goldstone, 2015) by 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. Furthermore, as we noted above, Schwartz and Goldstone (2015) have argued that we should try to "coordinate learning processes so they can do more together than they can alone." The idea that multiple presentations can sometimes be beneficial has also been explored more specifically in educational contexts, for example Ainsworth (2006) has created a theoretical framework for taxonomizing multiple external representations,[1] and Rau (2016) has considered specifically when multiple visual representations may be useful. These authors have suggested that multiple presentations often have both benefits (such as increased generalizability of knowledge) and costs (such as increased cognitive load that may impair learning). Thus, it is interesting to investigate whether presenting multiple presentations that support different aspects of understanding is beneficial or detrimental.

General Experimental Overview

We conducted a series of experiments investigating the effects of two isomorphic presentations of cyclic groups. The two presentations are not easily classified as concrete or abstract. Instead, they relate the cyclic group to different types of reasoning that students have some experience with. Specifically, we either ground the group in a visuospatial presentation, or in a nonvisual presentation that relates it more closely to familiar arithmetic operations. The visuospatial presentation is based on counting around the vertices of a polygon (we call this the "polygon presentation"), and the nonvisual presentation is based on simple arithmetic relationships and is closely related to modular arithmetic (the "modular presentation"). See the Materials and Method section below for more detail.

We show that these different presentations produce equal performance on the base concept (the group operation), but produce differential learning of later concepts. We used the group theoretic concepts of identities, inverses, and generators, as well as generalization from a cyclic group of one order to a cyclic group of another order, and from these 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 a group of a new order. Is there a way to obtain the benefits of both? Indeed, we combine our two presentations into a single hybrid presentation, and demonstrate that the benefits of combining multiple presentations may be more general than ideas like concreteness fading. We show that even over the short time of an experimental session, many subjects develop the ability to exploit the advantageous features of two distinct presentations with complementary advantages.

Experiments

Introduction

In this article, 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, because we did not 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 a hybrid presentation that included both the polygon and modular presentations (while keeping total instruction time approximately the same), and encouraged the participants to integrate them.

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. 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.

Materials and Method

All materials are available on our GitHub,[2] 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)

---

[1] Note that several of the authors in this section use the term "external representations" to refer to a more limited subset of what we term presentations, see above.

[2] https://github.com/lampinen/cyclic_group_presentations.

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 because of 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. (We chose groups of order 6 and order 9 so that each group would have enough elements for demonstrations of concepts like inverses, and sufficiently many generating and nongenerating 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 learning the basic concept/procedure. 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 procedures for finding inverses, identifying generators, and so forth. 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 procedures and concepts learned.

Group presentations.   In all three experiments, participants in one group received a presentation based on modular arithmetic (that is easily explained as a slight variation on regular arithmetic), while participants in another group received a visuospatial presentation based on counting around a polygon (that 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, 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 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 $4 \, ⑥ \, 4 = 2$, 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, which is an important part of learning to reason with multiple presentations (Ainsworth, 2006). 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%.

Because they were presented with both operations, the participants who saw the hybrid presentation received two more sentences of instruction on the operation than subjects in other groups, and saw both types of operation feedback on four of the subsequent operation practice problems. However, they received the same number of practice problems as the other subjects for the operation, and their instruction on all subsequent concepts was identical. We also added one additional page after presentation of the hybrid 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 to control for the effect of the additional prompt. We did not attempt to control for the few extra sentences about the operation that the hybrid group received, but (as shown below) we saw no improvement from the hybrid presentation on the operation, where they had this additional instruction. We only
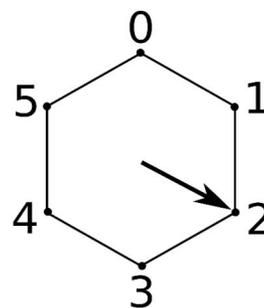

Figure 2.   Order 6 polygon figure.

observed improvement on the concepts where instruction was identical.

Identities and 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 $1 \circledS 5 = 0$ and $5 \circledS 1 = 0$. 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 $1 = 1$, $2 = 1 +_6 1$, and so forth 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.

Order 9 group 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 $\circledS$ with a visual aid analogous to that in 2, except with nine vertices labeled with the numbers 0 through 8. 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:

- A set of seven problems with the group operation, for example, $6 \circledS 4 = ?$, 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 nongenerators 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 $\circledS +_n$ with the visual aid shown in Figure 3. (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 $\circledR +_n$, for 1 and for an arbitrary element x.
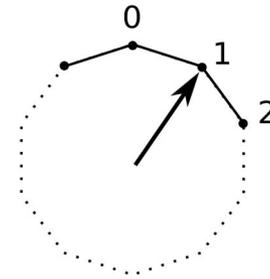

Figure 3. Order n polygon figure.

- 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 nongenerators (If an element x is not a generator under $+_n$, x must be a multiple of a divisor of n.)
- Three always/sometimes/never questions about generators. (e.g., If an element x is a generator under $\circledR +_n$, is its inverse a generator always, sometimes, or, never?)

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

Representation-use questions (Experiment 3). (As a reminder, we will use the term representation to refer to the mental model 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 nonzero elements, identifying generators, and answering T/F questions about generators.

Participants. We recruited participants using Amazon's Mechanical Turk (cf. 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 (gender: 50 female, 49 male, 1 decline to state/other; age range: 20–64; education: 21 high school diploma or less, 38 some college—associates degree or no degree, 32 bachelor's degree, 9 higher than bachelor's degree); Experiment 2 had $n = 50$, $N = 150$ (gender: 75 female, 75 male; age range: 19–67; education: 41 high school diploma or less, 44 some college—associates degree or no degree, 52 bachelor's degree, 14 higher than bachelor's degree); and Experiment 3 had $n = 100$, $N = 300$, (gender: 133 female, 165 male, 1 decline to state/other; age range: 19–69; education: 38 high school diploma or less, 94 some college—associates degree or no degree, 106 bachelor's degree, 34 higher than bachelor's degree).

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 understanding, and a presentation that is beneficial for one concept or aspect may be deleterious for another. This was inspired by the work of (De Bock et al., 2011), discussed above, which showed that there might be some effect on the inferences students draw from presentations. We hypothesized these inferences might affect their ability to learn related concepts. (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 and 3 was to verify our results.)

For Experiments 2 and 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 nonzero 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 all questions, resulting in them being significantly worse overall. For example, Rau (2016) has suggested that multiple presentations must contain relevant and new information, while still overlapping sufficiently with the other presentations. 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.

Measures. We used a variety of measures in our analyses of these data. For the operation questions and inverse questions,

performance was assessed by whether a numerical answer given was correct. For questions identifying whether an element was a generator, and for T/F and A/S/N questions responses were multiple choice, and performance was assessed by whether the response was the correct option. For inverse formula questions, performance was assessed by whether the response was "$n - x$" for[3] the inverse of x and "$n - 1$" for the inverse of 1, no other answers were accepted and no partial credit was given.

We collected typed explanations on some problems, and attempted to build naïve Bayes classifiers (see, e.g., Ng and Jordan (2002)) to classify the representation participants in the modular and polygon groups were using based on the occurrence of particular words in these explanations. The goal was to then use these classifiers to classify the explanations of hybrid group participants, to gain insight into their thought processes. We observed that some words (e.g., subtract, plus) were sometimes used by modular group participants but very rarely used by polygon participants, while other words (e.g., polygon, count, and point) were sometimes used by polygon participants but very rarely by modular participants. However, many participants used none of these words, and classifiers trained to predict group membership based on word use were not very sensitive (see *explanation word use analyses* in Results for further discussion), so we discarded use of these explanations in favor of the representation use questions.

For the representation use questions, we used two 5-item Likert scales (*not at all*, "−", *somewhat*, "−", *very much*) that allowed subjects to rate for each representation to what extent they had used it on the previous problem. Results were coded from 0 to 4 based on which item they had selected. For our analysis evaluating effects of diagram use (see below), as our measure of interaction with the diagram we used whether they had clicked on the diagram at least once.

Analysis. For Experiment 1, we chose to analyze 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. 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. We excluded participants who reported in the background section that they had used modular arithmetic or mathematical groups before.

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 par-

---

[3] This answer is not quite correct—it fails for the case $x = 0$, as noted below. However, we accepted them as finding the correct answer was quite difficult. Only one subject gave a fully correct answer that addressed both cases, and we had to exclude this subject's data anyway as they had prior experience with group theory.

ticipants bootstrapped across 10,000 resamples of the participants, with the predictors being as in Experiments 1 and 2. Bootstrapping provides accurate estimates of uncertainty for logistic regression (Gong, 1986; Wasserman, 2006). 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 preregistered on the Open Science Framework. (We also retrospectively ran this bootstrapped logistic regression on the data from Experiments 1 and 2, 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. Forty-eight participants were excluded for having too much background knowledge.

To present our results in a more coherent and meaningful format, we performed a meta-analysis across our three experiments, using the approach for estimating effect sizes described by Chinn (2000). We present the results of this analysis here. Including within-paper meta-analyses has been suggested as a way to improve the accuracy of conclusions drawn from behavioral science research (McShane & Böckenholt, 2017).

Implementation details. The tasks were developed using the jsPsych framework (de Leeuw, 2015) with a custom plugin to integrate the interactive polygon diagrams where necessary, hosted on Stanford's servers, and embedded in the Mechanical Turk page. To ensure that participants received the correct version of the experiment, we recorded each page of instructions they were given and each question they were asked along with their response, and verified that these matched the intended condition. 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, accuracy 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.

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 nonzero elements. See Figure 4 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 6
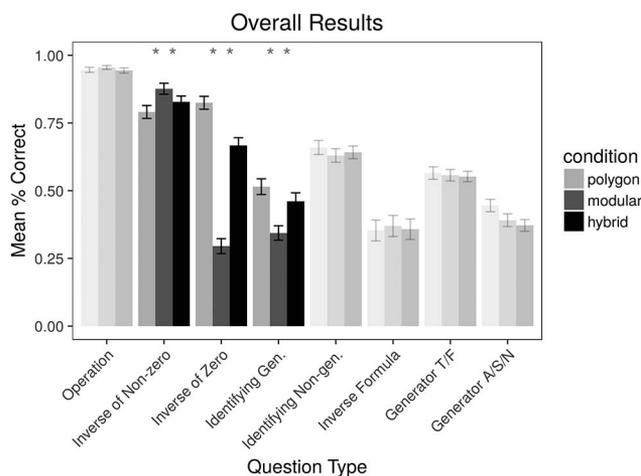

Figure 4. Results aggregated across group orders and Experiments 2 and 3 (Experiment 1 had no hybrid condition, so we omitted it from this graph). Highlighted results are the main findings relevant to our hypothesis, stars mark comparisons where the meta-analysis 95% confidence interval did not overlap zero (statistics include Experiment 1 data for polygon vs. modular comparisons).

for the aggregated results split across group orders showing this pattern of improvement, and the Hierarchical 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.

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 significance based on the inclusion of zero in the bootstrap percentile 95% confidence intervals for the predictors (Gong, 1986; Wasserman, 2006), which provides robust results even when performance is near ceiling or floor (as in the case on some portions of our experiment). We 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 *SD* across the population of participants.

Operation. 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 at first. Specifically, 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). 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).

Inverses. 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.)

Specifically, 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). We estimated the negative effect of the polygon condition on inverse of nonzero questions to be small (order 6: log OR = −0.53, effect size = −0.29; order 9: log OR = −0.81, effect size = −0.45). 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). We estimated the negative effect of the hybrid condition on inverse of nonzero 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).

Generators. 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.

Specifically, 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). We estimated the effect of the polygon condition on identifying nongenerators 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). We estimated the effect of the hybrid condition on identifying nongenerators to be negligible (order 6: log OR = 0.07, effect size = 0.04; order 9: log OR = −0.01, effect size = −0.01). 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).

Questions assessing reasoning about the general case. 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.

Specifically, we estimated the effect of the polygon condition on the inverse formula questions to be negligible (log OR = −0.16, effect size = −0.09), and similarly for the effect of the hybrid condition on the inverse formula questions (log OR = −0.04, effect size = −0.02). 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). 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). 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). 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).

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 because of 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 (Experiment 2: $b = 1.90$, $z = 2.76$, $p = .006$; Experiment 3: $b = 2.24$, $z = 5.12$, $p < 0.001$). Furthermore, this effect was present even when controlling for RT (Experiment 2: $b = 1.62$, $z = 2.26$, $p = .024$; Experiment 3: $b = 1.64$, $z = 3.51$, $p < 0.001$). This might suggest that engagement alone was not the driving factor. This 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 (Experiment 2: $b = 1.26$, $z = 6.93$, $p < 0.001$; Experiment 3: $b = 1.19$, $z = 10.28$, $p < 0.001$), even when controlling for RT (Experiment 2: $b = 1.42$, $z = 7.53$, $p < 0.001$; Experiment 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.

Explanation word-use analyses. In Experiment 2, we attempted to use the explanations of answers from the modular and polygon groups to build a naive Bayes model (see, e.g., Ng & Jordan, 2002) of explanations to classify hybrid group explanations to investigate their thought processes. Unfortunately, although some words were highly specific to condition, the classifiers had low sensitivity even on the modular and polygon group explanations (0.24 for modular and 0.31 for polygon). This is likely because of the design of the experiment—as noted above we attempted to homogenize language as much as possible between groups so that we could ensure that differences we observed were because of the differences in the presentations rather than differences in wording. This meant that participants in all conditions generally used similar language to describe their thought processes. Because of this we discarded this approach in Experiment 3 in favor of explicitly asking about representation use at the end of the experiment.

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* to *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 nonzero questions ($b_{mod} = 1.09$, $z = 2.95$, $p = .003$; $b_{poly} = 1.11$, $z = 2.94$, $p = .003$). This, together with the previous finding, may suggest some integration occurring in the hybrid condition, such that the advantages of each representation are to some extent shared even when the other representation is used (Figure 5).

We found that participants polygon rating, but not modular, was significantly predictive of success on identifying generators ($b_{mod} = 0.19$, $z = 0.95$, $p = .34$; $b_{poly} = 0.75$, $z = 3.86$, $p < 0.001$). This corroborates our other data supporting the superiority of the polygon representation for these questions, 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 predictive of success on the generator True/False questions ($b_{mod} = 0.19$, $z = 1.28$, $p = .20$; $b_{poly} = 0.31$, $z = 1.38$, $p = .17$). This is unsurprising, since we did not observe any significant differences between the polygon and modular groups on these questions in the third experiment (Figure 6).

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 (Gelman, 2006) 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:

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 $1 - \varphi$, 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 $\theta = 0.41$, $\varphi = 0.49$, so the data are best fit under this model by assuming that about 40% the participants are benefitting from both representations, and those that are not are choosing the modular representation and polygon almost equally. We used the Bayesian Information Criterion (BIC; Schwarz, 1978) 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 that is, a fixed $\theta = 0$ and fit $\varphi = 0.56$ ($BIC = 1681.2$), and a model where all participants benefitted from both that is, $\theta = 1.0$ ($BIC = 1689.9$).

The BIC comparisons (all differences >36) provide "very strong" (Kass & Raftery, 1995) evidence that the 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,
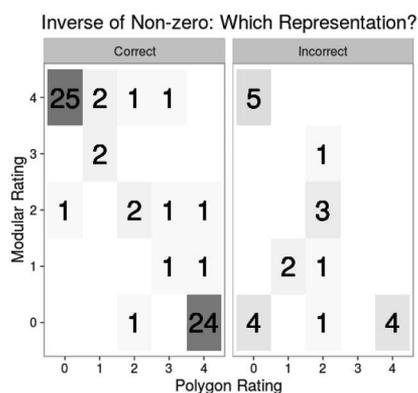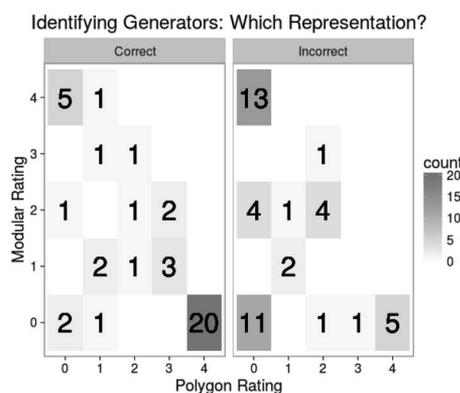
(a) Inverse of non-zero   (b) Identifying generators

Figure 5. Experiment 3—representation-use responses on inverse questions (counts of participants giving each rating, split by whether answer was correct).
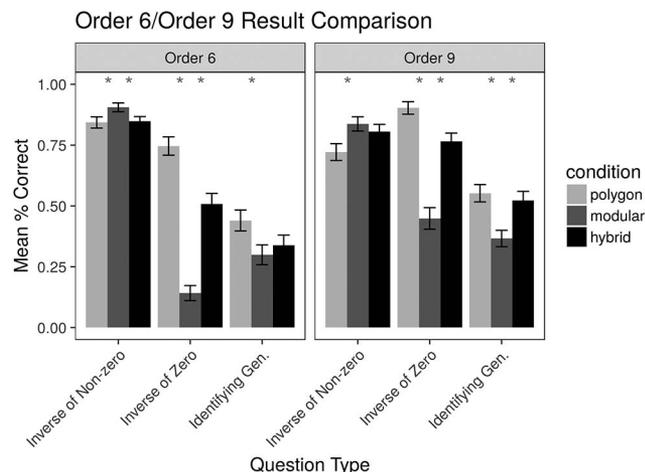
Order 6/Order 9 Result Comparison

Figure 6. Results aggregated across Experiments 2 and 3 (Experiment 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.

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 $\theta = 0.39$, $\varphi = 0.56$, 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 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, that is, a fixed $\theta = 0$ and fit $\varphi = 0.56$ ($BIC = 3584.8$), and a model where all participants integrated, that is, $\theta = 1.0$ ($BIC = 3700.2$). 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 6). This could be because of the fact that it takes some practice and/or feedback for the hybrid group to achieve the benefits of both presentations, and 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 reduce 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 the models described above 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 Figure 6, in Experiment 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 presentations early on, but 50–60% doing so by the time they reached the order 9 material.

Discussion

Our results address the issues with which we began this article. We have shown that different presentations of a concept can have effects on learning of later concepts. In particular, we have shown that relating a concept to different systems of reasoning can differentially support different subsequent types of understanding. Finally, we have shown that combining presentations can be beneficial, in that it can allow at least some learners to exploit the advantages of both.

These findings extend the consideration of the effects of different presentations beyond previous work in the domain of elementary group theory. Our work (consistent with some previous studies in other domains of mathematics learning) supports the following two messages: Rather than focusing on a single learning outcome, it is important to consider multiple aspects of understanding when assessing a presentation. Rather than focusing on concrete versus abstract, it is important to keep in mind that presentations that connect to different underlying knowledge structures may support different types of understanding. Below, we discuss these issues in more detail.

Polygon Versus Modular Presentations

Despite the fact that the participants in the polygon and modular conditions 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 presentation 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 somewhat small, the modular group performed significantly better at finding the inverse of nonzero elements. Thus, the presentations of earlier material had differential effects upon learning of later concepts.

We now consider in more detail how performance with the different presentations measures up to the criteria we proposed in the introduction. Both presentations allowed students to apply the directly instructed concepts correctly in the group of order 6, as well as to transfer these concepts to the group of order 9 with little loss. Furthermore, each presentation allowed students to answer questions about further concepts that built upon the base concept, and to transfer these concepts to the group of order 9. Each presentation had advantages and disadvantages for learning different subsequent concepts. When transferring these concepts to the group of order 9, performance increased on some types of questions and on others it decreased, but the overall advantages of each presentation remained about the same.

However, neither presentation performed particularly well at allowing participants to generalize about cyclic groups or allowing them to express (or evaluate the truth of) generalizations using formal mathematical expressions and language. 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.

These heterogeneous results support our claim that it is important to assess different types of understanding when evaluating a presentation. This continues a long trail of research showing that understanding is complex and nuanced rather than simple and monotonic (Bisanz & LeFevre, 1992; Greeno & Riley, 1987; Nokes & Ohlsson, 2005). Most saliently, our results imply that the results of Kaminski et al. (2008) (and other similar work) should be interpreted with caution—while their transfer measure showed benefits from an abstract presentation, there are many other learning goals in mathematics education that may be better served by different presentations. Our results show that there may not always be a single presentation that is uniformly best.

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, because they are already thinking arithmetically 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 nonzero elements because they were more likely to use the more efficient and accurate subtraction strategy.

Why would the modular group participants 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 (that 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 nonzero questions) may be reinforcing the error.

Generators. The polygon group performed better at identifying elements that are generators. We hypothesize that this is because of 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 $5 \rightarrow 1 \rightarrow 6 \rightarrow 2 \rightarrow \ldots$ 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 all appear to be achieving the full advantages of each presentation, especially 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. Encouragingly, it also suggests that 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 suggest that given sufficient practice with it, the hybrid presentation 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? it has been suggested in previous research on multiple representations that, in addition to benefits, there are costs associated with the additional cognitive load of understanding multiple sources of input (Ainsworth, 2006), especially in the initial learning process when students must both learn *from* and *about* the representations used (Rau, 2016). Another perspective on this 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 (e.g., 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 occurring 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 nonzero 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 nonzero 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 because of 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. We hypothesize that a hybrid presentation must be constructed from distinct presentations with distinct advantages to be beneficial. That is, the benefits of distinct presentations will generally be increased when they support complementary aspects of understanding. This idea is rooted in the literature on multiple representations (Ainsworth, 2006). The different concrete presentations in the work of Kaminski et al., though different in superficial details, can be seen as drawing on the same numerical intuitions that we argue are unhelpful when participants are confronted with one of the nonnumeric presentations. We would expect that by combining one of the numeric and one of the nonnumeric presentations used by Kaminski and colleagues, one might be able to achieve better transfer performance more broadly (e.g., 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).

Beyond Concrete and Abstract

Our results highlight the limitations of organizing presentations into a single concrete-abstract continuum, as they have been in previous work. While Belenky and Schalk (2014) noted that there are concrete details that can be added to a presentation that are irrelevant and do not improve understanding or transfer, and De Bock et al. (2011), Fyfe et al. (2014) and others pointed out that concrete presentations could be advantageous as well, we suggest that the most important features may be how presentations connect to previous types of understanding. Neither of our presentations was clearly more concrete than the other, instead they related the cyclic group to different ways of thinking. The polygon presentation related the group operation to a visuospatial manipulation and counting, while the modular presentation related it more directly to arithmetic operations that the subjects knew. The connections made to prior knowledge in both presentations likely make the group operation easier for students to understand, but they do not do so in the same way. Furthermore, this is not just a superficial difference—our results suggest that these different presentations altered the way participants were able to learn related concepts later, and we have argued that this was because of specific connections that were supported by the different presentations. We suggest that future research should move beyond concrete and abstract to a fuller evaluation of the relationships to other types of understanding given by a presentation.

Formalization and 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; Davidson, Eng, & Barner, 2012; Reber, 1967). This may also explain why the advantages of the presentations did not 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.

Indeed, the process of learning to make the mapping between procedures that generalize 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. Our subjects general failure to formalize is reminiscent of the results of Burger and Shaughnessy (1986), showing that (in their small sample) nobody without college mathematical training was able to reason in a formal way about geometry, and the results of Hazzan (1999), showing that even if students can state theorems, they may not be able to draw upon them when reasoning. This relates to the idea that representations that support certain kinds of reasoning must be learned (Greeno & Riley, 1987), and relates more generally to the idea that students are often failing to understand the broader structures underlying the specific examples they work with (Richland et al., 2012).

Because of the complexity of these issues, it is difficult to say how our presentations impact generalization and formalization performance. It is very likely that people who are more experienced with mathematics would perform better at formalization, and that more explicit practice with formalizing concepts within the experiment would lead to better performance on these types of questions. Once these factors are accounted for, are there other 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 types of understanding, and how presentations may affect each of them.

Time and 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 that 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 it is 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 students are performing is necessary for formalization 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? Does the learner need to be seeking a formal rule to become aware of one? These questions provide a fascinating direction for future research.

Limitations and Future Directions

There are a number of limitations of our current study. First, although we have highlighted the fact that presentations that connect a concept to different types of reasoning may support different types of understanding, we have not delineated fully the different ways that a concept can be related to prior knowledge, and how this will impact future learning. We have suggested that concrete versus abstract may not capture the most interesting features of presentations, but we have not articulated in full a set of dimen-

sions that do suffice to describe the space of presentations. This is a difficult problem, because both the axes and their impacts may depend on the details of the concepts in question and the prior experiences of the students. Exploring these issues is an important direction for future work.

Because of our focus on complete presentations of material rather than single representations, there were aspects of our experimental materials (the operation symbol and the presence of the polygon) that were different for the two experimental groups throughout the experiment. This is a realistic model of certain learning environments in which concepts build on others; in mathematics education notation is often repeated even when it has semantically broadened from how it was originally used or taught, and differences in notation may propagate the influence of different pedagogical choices. However, we would hypothesize that persistent aspects of the presentations are not necessary to observe effects on later concepts, and future research should explore this.

Furthermore, as noted above, the short period of time does not allow for full assimilation of the material or significant practice with the concepts, which limits our conclusions about the long-term effects of different presentations. The fact that the subjects mostly lack training with formal mathematical reasoning and proofs likely contributed to the very low performance on the formalization questions in this experiment, and limits our ability to draw conclusions about the effects of presentations on formalization. Our work shows that presentations can have an effect on later related concepts in the short-term, but future work should explore how these impacts propagate over a longer period of time and in more formal settings.

This also relates to the impact of expertise—because group theory concepts are fairly elementary in advanced mathematics, we had to exclude most subjects who had a background in advanced mathematics. This limits the inferences we can make about the impact of mathematical background. However, the question of how expertise changes the effects of presentation of a concept is very interesting. Are mathematical experts more readily able to reason without resorting to the details of the presentation and how it relates to other knowledge? The work of Hazzan (1999) suggests that students new to a concept may rely on the details more than those who are experts with the concept, but it does not explore the differences in how experts and novices reason about a concept that is new to both groups. This would also be an exciting direction for future work.

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 presentations that ground a concept in different ways can produce differential understanding of related concepts learned later. 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 learning some related concepts or some aspects of reasoning, and less useful for others.

These findings contribute to the ongoing exploration of the effects of presentations in math cognition, illustrating that even presentations that are not clearly more abstract or concrete than one another may have different advantages and disadvantages.

Thus, we suggest that abstract versus concrete may not be the best way to characterize the functionally important differences among presentations. Instead, the relevant features may be the ways a presentation connects the concept to students prior systems of knowledge. Our results show that presentations that connect to different systems of knowledge may support different aspects of understanding. Because of this, trying to find a single best type of presentation may be futile.

Instead, we have highlighted an alternative strategy for improving performance: teaching multiple complementary presentations while encouraging learners to develop an integrated understanding of them. Our results suggest 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 better encourage learners to generalize and formalize their understanding. These questions provide exciting new directions for research in both math cognition and math pedagogy.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16,* 183–198. http://dx.doi.org/10.1016/j.learninstruc.2006.03.001

Anderson, J. R. (1996). A simple theory of complex cognition. *American Psychologist, 51,* 355–365. http://dx.doi.org/10.1037/0003-066X.51.4.355

Belenky, D. M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review, 26,* 27–50. http://dx.doi.org/10.1007/s10648-014-9251-9

Bisanz, J., & LeFevre, J.-A. (1992). Understanding elementary mathematics. In J. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 113–136). Oxford, England: North-Holland.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A Simple proposal with multiple implications. *Review of Research in Education, 24,* 61–100. http://dx.doi.org/10.3102/0091732X024001061

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,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Burger, W., & Shaughnessy, J. (1986). Characterizing the van Hiele levels of development in geometry. *Journal for Research in Mathematics, 17,* 31–48. http://dx.doi.org/10.2307/749317

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19,* 3127–3131. http://dx.doi.org/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 R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making*. Hove, England: Psychology Press.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123,* 162–173. http://dx.doi.org/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, 109. Retrieved from http://www.staff.science.uu.nl/{~}savel101/as-sec410/literature/Bock2011.pdf

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47,* 1–12. http://dx.doi.org/10.3758/s13428-014-0458-y

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87,* 477–531. http://dx.doi.org/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,* 9–25. http://dx.doi.org/10.1007/s10648-014-9249-3

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics, 48,* 432–435. http://dx.doi.org/10.1198/004017005000000661

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

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences, 14,* 69–110. http://dx.doi.org/10.1207/s15327809jls1401_4

Gong, G. (1986). Cross-validation, the jackknife, and the bootstrap: Excess error estimation in forward logistic regression. *Journal of the American Statistical Association, 81,* 108–113. http://dx.doi.org/10.1080/01621459.1986.10478245

Greeno, J., & Riley, M. (1987). Processes and development of understanding. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 289–313). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hazzan, O. (1999). Reducing abstraction level when learning abstract algebra concepts. *Educational Studies in Mathematics, 40,* 71–90. Retrieved from https://link.springer.com/article/10.1023/A:1003780613628

Hiebert, J., Carpenter, T. P., Fennena, E., Fuson, K. C., Wearne, D., Murray, H., . . . Lindquist, M. M. (1997). *Making sense: Teaching and learning mathematics with understanding.* Portsmouth, NH: Heinemann.

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

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science, 320,* 454–455. http://dx.doi.org/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,* 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572

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,* 512–534. http://dx.doi.org/10.1016/j.tics.2016.05.004

Lang, S. (2002). *Algebra* (Rev. 3rd ed.). New York, NY: Springer, Graduate Texts in Mathematics.

McShane, B. B., & Böckenholt, U. (2017). Single paper meta-analysis: Benefits for study summary, theory-testing, and replicability. *Journal of Consumer Research, 43,* 1048–1063. http://dx.doi.org/10.1093/jcr/ucw085

Nathan, M. J. (2008). An embodied cognition perspective on symbols, gesture, and grounding instruction. In M. Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (Vol. 18, pp. 375–396). Oxford, England: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199217274.003.0018

Nathan, M. J. (2012). Rethinking formalisms in formal education. *Educational Psychologist, 47,* 125–148. http://dx.doi.org/10.1080/00461520.2012.667063

Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *Advances in neural information processing systems* (pp. 841–848). Cambridge: MIT Press.

Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: What is learned? *Cognitive Science, 29,* 769–796. http://dx.doi.org/10.1207/s15516709cog0000_32

Rau, M. A. (2016). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review.* Advance online publication. http://dx.doi.org/10.1007/s10648-016-9365-3

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863. http://dx.doi.org/10.1016/S0022-5371(67)80149-X

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist, 47,* 189–203. http://dx.doi.org/10.1080/00461520.2012.667065

Schwartz, D. L., & Goldstone, R. L. (2015). Learning as coordination. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (3rd ed., chap. 5, pp. 61–75). New York, NY: Routledge.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. http://dx.doi.org/10.1214/aos/1176344136

Wasserman, L. (2006). *All of Nonparametric Statistics.* New York, NY: Springer Science + Business Media.

Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education. In I. Harel & S. Papert (Eds.), *Constructionism.* Westport, CT: Ablex Publishing.

(*Appendix follows*)

LAMPINEN AND McCLELLAND

Appendix

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 article). 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 *: $G \times G \rightarrow G$ (*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 does not 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 is 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 is 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

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

then 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 $5 \geq 4$, $3 \; ^* \; 1 = 4 - 4 = 0$, and so forth. 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, for example, Lang (2002).