Adolescent Reasoning in Mathematical and Non-Mathematical Domains: Exploring the Paradox

Eric Knuth

Charles Kalish

Amy Ellis

Caroline Williams

University of Wisconsin

Matthew Felton

University of Arizona

The research is supported in part by the National Science Foundation under grant DRL-0814710. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Abstract

Mathematics education and cognitive science research paint differing portrayals of adolescents' reasoning. A perennial concern in mathematics education is that students fail to understand the nature of evidence and justification in mathematics. In particular, students rely overwhelming on examples-based (inductive) reasoning to justify the truth of mathematical statements, and often fail to successfully navigate the transition from inductive to deductive reasoning. In contrast, cognitive science research has demonstrated that children often rely quite successfully on inductive inference strategies to make sense of the natural world. In fact, by the time children reach middle school, they have had countless experiences successfully employing empirical, inductive reasoning in domains outside of mathematics. In this chapter, we explore this seeming paradox and, in particular, explore the question of whether the skills or knowledge that underlie adolescents' abilities to reason in non-mathematical domains can be leveraged to foster the development of increasingly more sophisticated ways of reasoning in mathematical domains.

A perennial concern in mathematics education is that students fail to understand the nature of evidence and justification in mathematics (Kloosterman & Lester, 2004). Consequently, mathematical reasoning—proof, in particular—has been receiving increased attention in the mathematics education community with researchers and reform initiatives alike advocating that proof should play a central role in the mathematics education of students at all grade levels (e.g., Ball, Hoyles, Jahnke, & Movshovitz-Hadar, 2002; National Council of Teachers of Mathematics, 2000; Knuth, 2002a, 2002b; RAND Mathematics Study Panel, 2002; Sowder & Harel, 1998; Yackel & Hanna, 2003). Proof plays a critical role in promoting deep learning in mathematics (Hanna, 2000); as Stylianides (2007) noted, "proof and proving are fundamental to doing and knowing mathematics; they are the basis of mathematical understanding and essential in developing, establishing, and communicating mathematical knowledge" (p. 289). Yet, despite its importance to learning as well as the growing emphasis being placed on proof in school mathematics, research continues to paint a bleak picture of students' abilities to reason mathematically (e.g., Dreyfus, 1999; Healy & Hoyles, 2000; Knuth, Choppin, & Bieda, 2009; Martin et al., 2005).

In contrast, cognitive science research has revealed surprising strengths in children's abilities to reason inferentially in non-mathematical domains (e.g., Gelman & Kalish, 2006; Gopnik, et al., 2004). Although more traditional (Piagetian) views posit children as limited to understanding obvious relations among observable properties, there is growing evidence that children are capable of developing sophisticated causal theories, and of using powerful strategies of inductive inference when reasoning about the natural world (for review, see Gelman & Kalish, 2006). In the former case, for example, children can integrate statistical patterns to form representations of underlying causal mechanisms (Gopnik & Schulz, 2007). In the latter case, for example, children often organize their knowledge of living things in ways that reflect theoretical principles rather than superficial appearances (Gelman, 2003). Thus, this raises something of a paradox: Why do children appear so capable when reasoning in non-mathematical domains, yet seemingly appear so incapable when reasoning in mathematical domains?

In this chapter, we explore the paradox by considering the research on adolescents' reasoning capabilities within mathematics education as well as within cognitive science. In particular, we briefly consider research that provides a portrayal of adolescents' reasoning in mathematical and non-mathematical domains. Next, we present preliminary results from the first phase of our multi-year research effort to better understand the relationships between adolescents' reasoning in mathematical and non-mathematical domains. We view such relationships as a means for potentially leveraging the strengths adolescents demonstrate when reasoning in non-mathematical domains to foster the development of their mathematical ways of reasoning. Finally, we discuss the implications of our research as well as its future directions.

Situating the Paradox

Mathematics education and cognitive science research paint differing portrayals of adolescents' reasoning, particularly with respect to the nature of their reasoning strategies. In the world outside the mathematics classroom, children typically rely quite successfully on inductive inference strategies—empirical generalizations and causal theories—to make sense of the natural world. For example, preschool-aged children are able to interpret and construct interventions to identify causal mechanisms in simple systems (Gopnik, et al., 2004). Young children also have rich knowledge structures supporting explanation and predictions of physical, biological, and social phenomena (Gelman & Kalish, 2006). In fact, by the time children reach middle school, they have had countless experiences successfully employing empirical, inductive reasoning in domains outside of mathematics.[1] Not surprisingly, many students also employ similar reasoning strategies as they encounter ideas and problems in mathematics (Recio & Godino, 2001); however, they often fail to successfully navigate the transition from inductive to deductive reasoning—the latter being the essence of reasoning in mathematics. As Bretscher (2003) noted, "Proof in everyday life tends to take the form of evidence used to back up a statement. Mathematical proof is something quite distinct: evidence alone might support a conjecture but would not be sufficient to be called a proof" (p. 3).

Adolescent Reasoning in Mathematical Domains

It is generally accepted that students' understandings of mathematical justification are "likely to proceed from inductive toward deductive and toward greater generality" (Simon & Blume, 1996, p. 9). Indeed, various mathematical reasoning hierarchies have been proposed that reflect this expected progression (e.g., Balacheff, 1987; Bell, 1976; van Dormolen, 1977; Waring, 2000); yet, research continues to show that many students fail to successfully make the transition from inductive to deductive reasoning.[2] One of the primary challenges students face in developing an understanding of deductive proof is overcoming their reliance on empirical evidence (Fischbein, 1982). In fact, the wealth of studies investigating students' proving

[1] Unfortunately, children's experiences successfully employing empirical, inductive reasoning in elementary school also tend to engender the belief that such reasoning suffices as proof in mathematical domains.

[2] The hierarchies that have been proposed, although based on empirical data, do not provide accounts regarding the actual transition from inductive to deductive reasoning. Rather, the hierarchies primarily note differences in the nature of students' inductive reasoning (e.g., justifications that rely on several "typical" cases versus those that rely on "extreme" cases) with deductive reasoning being at the "upper end" of the hierarchies, and not how (or if) such inductive reasoning strategies can develop into deductive reasoning strategies.

competencies demonstrates that students overwhelmingly rely on examples to justify the truth of statements (e.g., Healy & Hoyles, 2000; Knuth, Choppin, & Bieda, 2009; Porteous, 1990).[3]

As a means of illustrating adolescent reasoning in mathematics, we briefly present results from our prior work concerning middle school students' proving and justifying competencies. The following longitudinal data are from 78 middle school students who completed a written assessment at the beginning of Grades 6 and 7, and at the end of Grade 8; the assessment focused on students' production of justifications as well as on their comprehension of justifications.[4] In the narrative that follows, we present a representative sample of the assessment items and corresponding student responses. Justifications in which examples were used to support the truth of a statement were categorized as *empirical*, justifications in which there was an attempt to treat the general case (i.e., demonstrate that the statement is true for all members of the set) were categorized as *general*, and justifications that did not fit either of these two aforementioned categories were categorized as *other*.[5]

As an example, students were asked to provide a justification to the following item: *If you add any three odd numbers together, is your answer always odd?* The following two student responses are representative of empirical justifications:

> Yes because 7+7+7=21; 3+3+3=9; 13+13+13=39. Those problems are proof that it is true. (Grade 6 student)

---

[3] For the purposes of this chapter, we define inductive reasoning to be reasoning that is based on the use of empirical evidence, and by empirical evidence we mean the use of examples to justify statements or conjectures. Moreover, inductive reasoning is not to be confused with mathematical induction—a mathematically valid method of proving.

[4] The assessment items presented below were the same for each administration of the assessment, and the same group of students completed the assessment at all three time points.

[5] We have simplified the categorizations described in this chapter as we are primarily interested in highlighting the differences between empirical-based justifications and more general, deductive justifications. See Knuth, Choppin, and Bieda (2009) and Knuth, Bieda, and Choppin (forthcoming) for more detail about the study's results.

> 1+3+3=7. 3+11+1=15. Yes it would be but you will have to do it a 100 times just to make sure. (Grade 7 student)

In contrast, the following two responses are representative of general justifications:

> If you add two odds, the result is even. An even plus one more odd is odd. So three odds added together always results in odd. (Grade 6 student)

> We know that odd and odd equals even. So even (2 odds) added together with odd equals odd. This shows us that no matter what three odd numbers you add together, the sum will always be an odd number. (Grade 8 student)

Responses categorized as *other* (see Footnote 2) were often restatements of the question (without further justification) or nonsensical responses (e.g., "It is not always odd because some problems are even like 1+2+3=6, 3+4+5=12"). Table 1 displays the overall results of students' justifications for this item. As the table illustrates, a significant number of students relied on examples as their means of justification, with very little change occurring across the middle grades. We also see an increase in the number of students attempting to produce more general, deductive justifications from Grade 6 to Grade 8; yet still less than half the students produce such justifications even by the end of their middle school mathematics education.

As a second example, consider students' responses to the following assessment item: *Sarah discovers a cool number trick. She thinks of a number between 1 and 10, she adds 3 to the number, doubles the result, and then she writes this answer down. She goes back to the number she first thought of, she doubles it, she adds 6 to the result, and then she writes this answer* *down.* [A worked-out example, including the computations, followed the preceding text.] *Will Sarah's two answers always be equal to each other for any number between 1 and 10?* In this case it is also worth noting that students could use examples to prove that the statement is always true by testing the entire set of possible numbers (i.e., numbers between 1 and 10). Table 2 presents the results for this item; justifications based on proof-by-exhaustion are also included in the *general* category (only approximately 5% of students in each grade level used this method). Given the significant proportion of students whose responses were categorized as *other*, it is worth briefly discussing potential reasons underlying their responses. The majority of these responses were either a result of (i) students misinterpreting the problem, thinking that the end result must always be twenty (the result that was provided in the worked-out example that accompanied the problem); or (ii) students not being able to articulate a general argument—students could "see" what was going on but were unable to provide an adequate justification. In the former case, the following response is representative: "No, because the number comes out differently if you chose a number like 11. It does not come out as 20" (Grade 8 student). In the latter case, the following response is representative: "The answers will always be equal because you're just doing the same thing" (Grade 7 student). Although the percentage of students who relied on empirical-based justifications decreased relative to the previous example, the grade-level trend regarding the number of students providing general, deductive justifications remained about the same (again, less than 50% of the students at any grade level provided this type of justification).

As a final example, consider the following item in which students were asked to compare an empirical-based justification with a general, deductive justification: *The teacher says the following is a mathematical fact: When you add any two consecutive numbers, the answer is* *always odd. Two students offer their explanations to show that this fact is true* [Note: an empirical-based justification and a general, deductive justification are then presented for the students]. *Whose response proves that if we were to add any two consecutive numbers we would get an answer that is an odd number?* As Table 3 suggests, for many middle school students, an empirical-based justification seems to suffice as proof. We see a slight grade level increase in viewing the general, deductive justification as proving the claim and, interestingly, we see a substantial decrease by Grade 7 of students who think both justifications prove the claim.

In summary, the snapshot of adolescents' mathematical reasoning illustrated above is quite typical of the findings from much of the research: adolescents are limited in their understanding of what constitutes evidence and justification in mathematics and, moreover, they demonstrate a proclivity for empirical-based, inductive reasoning rather than more general, deductive reasoning. Although most studies have focused on adolescents in a particular grade level or across grade levels (i.e., cross-sectional studies), the research discussed above provides a longitudinal view into the development of adolescents' mathematical reasoning. And given this longitudinal view, we see very little development as adolescents progress through their middle school years, and what development we do see falls far short of desired outcomes.

Adolescent Reasoning in Non-mathematical Domains

The difficulties that adolescents show with regard to mathematical reasoning, including the apparent lack of development as they progress through middle school, raise the question of whether there is some developmental constraint that limits adolescents' mathematical reasoning. The most likely candidate would be abilities to do and understand deductive inference. The emergence of deductive inference has been a central focus of research on adolescent cognitive development, spurred in part by Piaget's theory of formal operations. Although there is considerable debate within this literature, a plausible reading suggests there is nothing special about adolescence in terms of acquiring deduction. Younger children, for example, have been shown to appreciate that deductive inference leads to certain conclusions, and is stronger than inductive inference (Pillow, 2002). At the same time, however, even adults struggle to reason formally and deductively.[6] Thus, deductive inference seems neither impossible before adolescence, nor guaranteed after. Rather than review the literature on the development of deductive inference in non-mathematical domains (see Falmange & Gonsalves, 1995), we take a slightly different approach here. Similar to the mathematics education research on proof discussed above, researchers in psychology have often argued that people rely on empirical solutions to deductive problems. An interesting difference with the literature in mathematics education, however, is that these empirical-based solutions are typically evaluated quite positively in non-mathematical domains. That is, the kind of performance that makes people look like poor deductive reasoners is actually consistent with their being quite good inductive reasoners.

Deductive and inductive arguments have very different qualities. On the one hand, in making a deductive argument, one endeavors to show that the hypothesized conjecture must be true as a logical consequence of the premises (i.e., axioms, theorems). On the other hand, in making an inductive argument, one seeks supporting evidence as the means for justifying that the conjecture is likely to be true. We will refer to arguments based on accumulation of evidence as "empirical." Often times the empirical support in inductive arguments is provided by examples. The conclusion "All ravens are black" is supported by encounters with black ravens (and the

[6] Note that in this literature, as in almost all work in psychology, "adult" means college-aged.

absence of no black ones).[7] The critical point is that deductive arguments prove their conclusions though logic, while inductive arguments provide evidence that conclusions are likely. Again, though there is strong debate, one influential view is that people often seek evidence (make inductive arguments based on examples from a class) when asked to evaluate logical validity (make deductive arguments).

One of the clearest examples of inductive approaches to a deductive problem is Oaksford and Chater's (1994) analysis of performance on the Wason selection task. The selection task is a classic test of logical argumentation. A participant is asked to evaluate a conjecture about a conditional relation, such as "If there is a *p* on one side of a card, then there is a *q* on the other side." The participant is presented with four cards: one each showing *p*, not-*p*, *q*, and not-*q*. The task is to select just the cards necessary to validate the conjecture. The logical solution is to ensure that the cards are consistent with the conjecture: that there are no *p* and not-*q* cards. To confirm this involves checking that there is a *q* on the back of the *p* card, and checking that there is a not-*p* on the back of the not-*q* card. In practice, most people do check the *p* card, but very few examine the not-*q* card. Rather, most people opt to explore the *q* card, which is logically irrelevant (both *p* and *q* and not-*p* and not-*q* are consistent with the conjecture). This behavior is often interpreted as akin to the logical fallacy of affirming the consequent (if *p* then *q*, *q*, therefore *p*). Oaksford and Chater argue that selecting the *p* and the *q* cards is actually a reasonable strategy for assessing the evidential support for the conjecture. Though the details are quite complex, they show that given reasonable assumptions about the relative frequencies of *p* and *q*, the cards selected are the most informative tests. That is, people's behavior conforms to a normative standard of hypothesis testing (e.g., optimal experiment design).

---

[7] There are many other sources of inductive support. For example, that one's teacher says, "All ravens are black." provides some reason for adopting the belief.

The distinction turns on two different ways of construing the task. Interpreting the problem as a deductive one (the way experimenters' intend it) can be glossed something like this: Is the statement logically consistent with the features of the four cards on the table? The inductive construal is something like the following: Is the statement likely to be true of cards in general? Although the deductive problem can be solved conclusively, it is not really that interesting or important (who cares about these four cards?) The inductive problem can never be truly solved (absent investigation of every card in the world), however, it is just the kind of problem that people really care about and face in their everyday lives. How can past experience (the four cards) help in the future (expectations about new cards)?

The idea that people often employ inductive, evidential support, strategies to solve deductive problems is part of a general approach to cognition and cognitive development that emphasizes probabilistic reasoning (Chater & Oaksford, 2008). From this perspective, most of the cognitive challenges people face involve estimating probabilities from evidence. This is straightforward for processes of categorization and property projection. Learning that barking things tend to be dogs, and that dogs tend to bark, seems to involve learning some conditional probabilities. Influential accounts of language acquisition suggest that children are not learning formal grammars (deductive re-write rules) but rather patterns of probabilities in word co-occurrences and transitions. Even vision has been analyzed as Bayesian inference about structures likely to have generated a given perceptual experience. The general perspective is that inductive inference is ubiquitous; we are continually engaged in the task of evaluating and seeking evidential support. Given the centrality of inductive inference, it should not be surprising that many psychologists argue that we are surprisingly good at it and, good at it from a surprisingly young age (Xu & Tenenbaum, 2007a; 2007b).

There are several lines of research potentially relevant to understanding adolescents' reasoning strategies in mathematics. Perhaps the most direct connection is with research on evaluations of inductive arguments. In contrast to deductive arguments, which are either valid or invalid, inductive arguments can vary in strength. If people are good at reasoning inductively, then they ought to be able to distinguish better and poorer arguments, stronger and weaker evidence, for conclusions. This work has both a descriptive focus—how do people distinguish stronger and weaker evidence—and a normative focus—do people's strategies conform to normative standards for evidence evaluation?

Most work on adolescent inductive inference has focused on the problem of identifying causal relations in multivariate domains (see Kuhn, 2002, for a review). Questions center on children's abilities to construct and recognize unconfounded experiments, distinguish between hypotheses and evidence, and generally to adopt systematic investigation strategies. Similar to the literature on deductive inference, the conclusions are generally that young children show some important abilities but are quite limited; adults are better, but far from perfect; and adolescents are somewhere in the middle. Other forms or elements of inductive reasoning show a significantly different profile: Even young children are skilled at inductive inference (see Gopnik & Schulz, 2007). Adolescents have not been the direct focus of research, but there seems no reason to believe that inductive inference skills should decline from early childhood to adolescence.

Research on evidential support explores how people respond to or generate evidence. Evidence in this work consists of different kinds of examples or instances.[8] The task is to make or evaluate a conclusion based on that evidence. For example, given that robins are known to

[8] In much of this work, the "examples" are categories of animals. It is unclear whether category-to-category inferences ("robins" to "owls") is different than individual-to-individual inferences ("these 3 robins" to "these 3 owls").

have a certain property (e.g., hollow bones), how likely is it that owls also have the property? The evidence consists of examples known to have the property in question. These examples can be understood as premises in an argument about the conclusion: The strength of the argument is confidence in the conclusion conditional on the evidence. Osherson and colleagues (Osherson, et al., 1990) developed one of the first models and described several criteria for evidential strength. Subsequent research has explored the development and application of these and other criteria (Lopez, Gelman, Gutheil, & Smith, 1992; Heit & Hahn, 2001; Rhodes, Gelman, & Brickman, in press). Table 4 provides a list of proposed criteria; note that some of these criteria are more normatively defensible than others.

Although there remains some debate about preschool-aged children, most researchers would agree that by middle childhood children use the criteria in Table 4 to evaluate examples as evidence. Thus, children judge that many examples are more convincing than are fewer, that a diverse set of examples is better than a set of very similar examples, and that an argument based on a typical example is stronger than an argument based on an atypical example. Research continues on other principles of example-based arguments, such as the role of contrasting cases (e.g., non-birds that do not have hollow-bones; Kalish & Lawson, 2007) and children's appreciation of the importance of sampling. The general conclusion is that children, including adolescents, are similar to adults in their evaluations of evidence. Moreover, children's evaluations accord quite well with normative standards of evidence.

Adolescent Reasoning in Mathematical and Non-mathematical Domains

The preceding discussion highlights some important differences between adolescent reasoning in mathematical and non-mathematical domains. In mathematics education, inductive strategies are typically treated as a stumbling block to overcome rather than as an object of study.

Moreover, mathematics education research has focused primarily on distinctions between empirical/inductive and formal/deductive justifications, and questions such as what makes one empirical justification better than another or what constitutes better/stronger evidence has not been well addressed.[9] In a recent paper Christou and Papageorgiou (2007) argued that the skills involved in induction, such as "comparing" or "distinguishing," were similar in mathematical and non-mathematical domains. Christou and Papageorgiou's work showed that students can identify similarities among numbers, distinguish non-conforming examples, and extend a pattern to include new instances. Thus, identifying how adolescents use such abilities to evaluate both mathematical and non-mathematical conjectures and how they think about the nature of evidence used to support conjectures may suggest a means for leveraging their inductive reasoning skills to foster the development of more sophisticated (deductive) ways of reasoning in mathematical domains.

Exploring the Paradox

What kinds of skills or knowledge underlie adolescents' abilities to reason in non-mathematical domains, and might such skills or knowledge have any relevance to reasoning in mathematical domains? There are many different accounts of inductive inference, but one fairly consistent component is a representation of relevant similarity in the domain. To make or evaluate empirical-based, inductive inferences one must have a sense of the significant relations among the examples or objects. For example, if the task is to decide whether birds have hemoglobin in their blood or not, the most informative examples will be objects similar to birds. The argument that since spiders lack hemoglobin, birds must lack it as well is not particularly convincing because spiders and birds seem very different. In contrast, knowing that reptiles have

[9] Although mathematics education researchers have noted differences in the nature of empirical justifications—checking a few "random" cases, systematically checking a few cases (e.g., even and odd numbers), and checking extreme cases (e.g., Balacheff, 1987)—they have not engaged in any deeper study of empirical justifications.

hemoglobin seems quite relevant if we believe birds and reptiles are relevantly similar. The critical question, then, is “what makes two things relevantly similar?” Other principles of inductive inference described earlier also depend on similarity relations (e.g., typical examples are better because they are similar to many other examples). The relevant similarity relations are, in part, knowledge dependent. Airplanes are similar to birds and may be useful examples to use when making inferences about aerodynamics, however, the question about hemoglobin calls for a biological sense of similarity. Getting the right similarity relations is a critical part of expertise. For example, experts tend to see “deep” similarities (e.g., evolutionary history), while novices often rely on shallow, domain-general similarities (e.g., appearances; Bedard & Chi, 1992). Put another way, reasoning from similar cases will only be successful if one’s representation of “similar” really does capture important relations in the domain.

A considerable amount of research and debate in the cognitive developmental literature involves just what kinds of similarity relations children recognize and how such relations are acquired. Some argue, for example, that evolution has equipped us to be sensitive to significant similarities (Spelke, 2000; Quine, 1969). Others argue that domain general learning principles allow children to hone in on the important relations (Rogers & McClelland, 2004). Regardless, the general finding is that quite young children seem to display useful and productive intuitions about similarity in the empirical domains studied. Even preschoolers recognize that reptiles are more like birds than are airplanes when biological questions are involved, but that airplanes may be more informative about birds when the questions involve aerodynamics (for example, Kalish & Gelman, 1992). Unfortunately, the mechanisms that have been hypothesized to underlie the development of similarity in empirical domains may fail to support a sense of mathematical similarity useful for evaluating mathematical conjectures.

The nativist view of similarity suggests that evolutionary pressures have shaped the human cognitive system to focus on important relations. For example, snakes often look like sticks, but an organism that focused on these similarities would find itself in significant peril. Clearly, quantitative relations have adaptive significance, and a long enough history that our species could have evolved specific cognitive dispositions to represent such relations. Indeed, there are important claims of just such a "number sense" involving representations of approximate magnitude (see Dahaene, 1999). Beyond the early grades, however, such relations are generally not important parts of mathematical thinking or reasoning. A sense of numerical similarity based on approximate magnitude is a limited basis for evaluating or making inferences about mathematical relations. The principles of mathematical relations depend on a formal system, which is too recent an invention to have had any significant selective pressure on the human cognitive system (see Geary, 1995). In geometry, basic mechanisms for representing shape provide a natural organization to the domain. It seems possible that this sense of similarity may be more productive, more related to mathematically significant properties, than representations of number.

Empiricist views of the development of similarity also suggest pessimism about a mathematical sense of similarity. The empiricist idea is that children form representations in a domain by tracking statistical patterns. In the natural world, objects tend to form clusters: There are natural discontinuities (Rosch, et al., 1976). The features that are important for representing animals tend to come in groups, with high intra-group correlations among features and low inter-group correlations. For example, birds tend to fly, have wings, and have feathers. These features co-occur and tend to be distinctive from the features of mammals that walk, have legs, and have fur. These patterns in the distributions of observed features allow people to pick out informative features and represent kinds or categories that reflect those distributions. A sense of mathematical similarity will also be dependent on the kinds of relations and distributions of properties observed in experience. Again, it seems likely that many of the most significant relations among mathematical objects in children's regular as well as school experiences may not be particularly well correlated with mathematically significant relations. A tendency to notice similarity in appearance does lead one toward a fairly useful notion of similarity among animals, because biological properties tend to be correlated with appearance. In contrast, a tendency to notice frequency or magnitude among numbers does not typically lead to a mathematically useful notion of similarity of numbers. Moreover, mathematical objects, at least numbers, have a network organization: There are many cross-cutting dimensions of similarity. In contrast, there is one, taxonomic, way of representing similarity relations among living things that seems primary (though see Ross, Medin, & Cox, 2007 on significance of ecological relations). Again, geometric objects, with a strong hierarchical organization, and a closer tie to psychological mechanisms of shape perception, may be somewhat different than numbers in this regard.

Thus, an important first step toward developing a deeper understanding of students' inductive reasoning in the domain of mathematics is to explore their representations of similarity relations among mathematical objects.[10] Successful inductive reasoning depends on seeing objects as similar to the degree they really do share important features or characteristics. How do students make judgments about whether two numbers or two geometric shapes are similar? What features or characteristics do students attend to when considering the similarity of numbers or geometric shapes? How do students' similarity judgments compare with experts' similarity judgments? Answers to such questions may provide insight into students' choices for the

---

[10] Note that our use of similarity refers to conceptual similarity unless we explicitly write mathematical similarity.

empirical evidence they use to justify mathematical conjectures, which, in turn, may provide insight into means to foster their transition to more deductive ways of reasoning.[11]

Assessing Similarity in Mathematical Domains

The first phase of our current research was to determine which features and characteristics individuals attend to when considering whole numbers and common geometric shapes; in particular, on what features might individuals base their decisions when determining whether a particular number or shape is typical? We conducted semi-structured interviews with 14 middle school students, 14 undergraduates, and 14 doctoral students in mathematics and engineering fields (hereinafter, STEM experts). Participants examined various numbers and shapes on individual cards and then sorted and re-sorted them into groups according to whatever principles they chose (Medin et al., 1997). The numbers and shapes presented to participants for inclusion are shown in Figures 2 and 3.

Participants engaged in three types of sorts: an open sort, a prompted sort, and a constrained sort. For the open sort, participants grouped and re-grouped numbers or shapes into categories of their own choosing until they had exhausted the types of categories they deemed relevant. For the prompted sort, the interviewer grouped some numbers (or shapes) according to a characteristic and asked participants to place additional numbers (or shapes) into the group. For instance, the interviewer might place the numbers 4, 25, and 81 (all perfect squares) into a group and ask the participant to include other numbers in the group. For the constrained sort, the interviewer provided a group of numbers (such as 4, 25, and 81) and then included an additional set of numbers (such as 23, 36, 51, and 100) and asked the participants which, if any, of the

[11] Mathematics education research has revealed very little insight into students' thinking regarding their choices of empirical evidence, yet such insight is critical in helping students develop more sophisticated ways of reasoning. For example, selecting examples that provide insight into the structure underlying why a conjecture is true can offer a potential means for generating a general, deductive justification (e.g., Yopp, 2009).

additional numbers should be included in the group. The prompted and constrained sorts allowed us to determine whether participants would sort by particular features deemed mathematically interesting, such as a number being a perfect square, versus other noticeable, yet mathematically uninteresting features, such as the number of digits of a number or the value of one of its digits.

The participants' responses to the sorting and tree building interview yielded 13 number categories and 13 shape categories denoting features deemed relevant in each domain. Table 5 presents the number categories and their meanings. One of the more salient results from the number-sorting task was the number of similarities between the middle school students and the STEM experts in terms of which features they noticed. For instance, consider the parity category. Figure 4 shows the percentage of participants from each group who sorted according to parity in the three sorts. The "Parity 1st" part of the graph shows the percentage of participants who sorted by parity as their first-choice sort in the open sort. The "Parity Open" part shows the percentage of participants who sorted by parity in the open sort, but not as their first sort, and the "Parity Prompt" section shows the percentage of participants who were able to sort according to parity only in the prompted or constrained sorts. In this case it is clear that parity was a particularly salient feature for all three groups.

The similarities between the middle school students and the STEM experts led us to wonder, were there any features that one group attended to but the other did not? The *factors* category was the only category that appeared to be salient to the middle school students but not to the undergraduates or the STEM experts. Just over 20% of the middle school students sorted according to factors in the open sort, whereas none of the undergraduates or STEM experts sorted according to factors. There was also just one category that the STEM experts could sort by more readily than the middle school students, and this was the *squared* category, referring to numbers that are perfect squares. Figure 5 shows the percentage of participants in each group who were able to sort by perfect squares.

The “prime” category was the other category of number we expected to be more salient to STEM experts, but this turned out not to be the case. Almost 90% of the STEM experts could sort according to primes, but 70% of the middle school students and the undergraduates could sort according to primes as well. Additionally, there were some mathematically uninteresting features that we expected the middle school students to attend to more than the experts, such as *intervals*, *value of the digits*, *number of digits*, *contains a digit*, and *arithmetic*. But of those five categories, differences only emerged for *contains a digit* and *number of digits*, and the differences were not large: 36% of the middle school students versus 21% of the experts sorted according to *contains a digit*, and 43% and 29% respectively sorted according to *number of digits*.

Table 6 presents the categories for the number sort organized by the features to which each group of participants attended, in order from the most salient to the least. The middle school students and undergraduates were somewhat more attentive to common digits and number of digits than the STEM experts, whereas the STEM experts were more attentive to perfect squares and arithmetic relationships.

Table 7 presents the categories of shape that the participants identified in the sorting task. The most salient category across all three groups was the number of sides, which all of the participants in each group used for grouping in either the first sort or the open sort. The other two categories that were also salient across all three groups were *shape* and *size*.

The only category that was more salient for the STEM experts than for the other participants was the *regular* category. Forty-three percent of the STEM experts grouped shapes according to whether they were regular, but only 14% of the middle school students sorted according to this principle, and those students did so only in a prompted sort. None of the undergraduates sorted according to regularity. There were three more categories that we anticipated would be more salient for the STEM experts: *similar*, *symmetry*, and *tessellate*. This turned out to be the case only for symmetry (see Figure 6): STEM experts sorted according to symmetry more often than middle school students (47% versus 21% for both the middle school students and the undergraduate students). Contrary to our expectations, as shown in Figure 6, middle school students attended to similarity slightly more than STEM experts did (50% versus 40%). Only one participant across the three groups attended to tessellations, and this participant was a middle school student.

We anticipated that categories such as *size*, *familiar*, and *orientation* would be ones that would be more salient for middle school students, particularly because we consider these categories to denote principles of shape that are not mathematically important. However, orientation and size were actually slightly more salient for the STEM experts, and the familiar category was equally salient across all three groups (36% of each group sorted according familiarity of shape). Twenty-seven percent of the STEM experts grouped according to orientation, while only 14% of the middle school students and 7% of the undergraduates sorted by orientation. All of the STEM experts sorted according size, versus 78% of middle school students and 85% of the undergraduates.

Table 8 presents the categories of shape sorted by which features each group of participants attended to in order from the most salient to the least. We found that STEM experts noticed symmetry and regularity more than did middle school students, whereas middle school students attended to similarity and equal sides more readily.

Findings from the sorting and tree building study showed that in general, there were not many differences between the middle school students, the undergraduates, and the STEM experts in terms of the characteristics of number and shape that they attended to. Furthermore, we found that some of the most salient features of number included multiples, parity, primes, and intervals (i.e., the relative size of numbers). Several of the more salient features of shape included the number of sides, the shape’s size, recognizable features of a shape, and the size of its angles. These findings are important in that they reveal particular characteristics that participants find noticeable, such as a number’s relative size or a shape’s size, that matter to students but are not mathematically important from our perspective.

Discussion and Concluding Remarks

The results from our initial study suggest that adolescents’ and experts have very similar representations of similarity among (some) mathematical objects. In particular, adolescents did notice mathematically significant relations among the objects; part of what makes two numbers or shapes similar is that they share properties relevant to mathematical theorems and conjectures. Of course, participants did notice less significant properties as well, for example, shared digits of numbers, and shared orientation of shapes. There was some evidence that these less significant properties played a larger role in adolescents’ representations of number and shape, however, such properties also showed up in the STEM experts’ sorts. One possible explanation for this result is due in part to the extremely open-ended, unconstrained, nature of our similarity measures. Participants, for example, were not instructed to focus on “mathematical” similarity. As noted above, part of expertise consists of being able to select an appropriate similarity metric for the task at hand. We suspect that experts would tend to ignore irrelevant features (e.g., orientation) in the context of evaluating mathematical conjectures. It is less clear whether adolescents would show the same selectivity—exploring the significance of contextual variations (e.g., mathematics class) is one aspect included in the next steps for this research program. The importance of the current findings, though, is that adolescents do represent mathematically significant similarity relations. The pressing question for future research, however, is how they use such relations to evaluate conjectures.

In this chapter we have taken a relatively new perspective (in mathematics education research) on adolescents' use of empirical strategies for evaluating mathematical conjectures. Rather than seeing such strategies as limited or as failures to adopt deductive strategies, we suggest that there may be value in such inductive strategies. The argument so far has been that inductive inference is a powerful and useful form of reasoning, and one that people (especially adolescents) seem both disposed to use and use relatively successfully. Our proposal is to consider inductive inference about mathematical conjectures as an object of study in and of itself. To that end, we seek to better understand the adolescents' inductive reasoning in the domain of mathematics. The empirical work presented in this chapter is a first step in a larger project of exploring just how adolescents use empirical examples and inductive methods to reason about mathematical objects. In short, we believe inductive inference strategies should play an important role in mathematics, and understanding adolescents' inductive reasoning may provide important insight into helping adolescents transition to more sophisticated, deductive ways of reasoning in mathematics.

In closing we want to make a more extended argument in favor of our perspective that inductive inference can and should play a productive role in school mathematics? Can this kind of reasoning support the transition to more general, deductive ways of reasoning? We began the chapter by noting that inductive arguments are commonplace in mathematics classrooms among middle school adolescents, and that more general, deductive reasoning is relatively rare. In contrast, we also noted that inductive arguments are important outside of mathematics and that adolescents often employ quite sophisticated inductive strategies on many tasks that seem to call for deductive inference. The perspective from cognitive science is that people are not so much poor deductive reasoners as they are reluctant deductive reasoners. One reason for this reluctance may be that (outside of mathematics) we rarely care about the deductive implications of some set of facts or propositions. Invariably, it is the empirical significance that people seem to care about in most aspects of their lives. Yet, mathematics is different, precisely because of the demand to attend to deductive relations. To the question of the proper place of inductive inference in mathematics education we offer three responses.

Although induction is not the accepted form of mathematical inference, it is a form of inference. Inductive reasoning can help students develop a feel for a mathematical situation and can aid in the formation of conjectures (Polya, 1954). It also provides a means of testing the validity of a general proof, especially where students are uncertain about the scope and logic of their argument (Jahnke, 2005). A major challenge in mathematics education, however, lies in moving students from reasoning based on empirical cases to making inferences and deductions from a basis of mathematical structures. By using more accessible inductive inference strategies, at least as an intermediate step, students may begin to appreciate that mathematics is a body of knowledge that can be reasoned about, explained, and justified. A concern with justification and explanation, even if inductively based, may support, rather than undermine, acquisition of more formal proof strategies.

Inductive inferences are important mathematical strategies in their own right. Mathematical problems do not always demand formal solution approaches. This point is very much akin to the value of estimation in relation to exact computation. For example, it is very useful to be able to guess whether a novel problem will be like some familiar problem; perhaps the same solution strategies will work in both cases. One does not need a formal proof that the two problems are isomorphic. Of course, the induction may be false and the apparent similarities misleading. However, developing better inductive strategies, such as recognizing which dimensions of similarity are important, is an important mathematical skill. Even in the context of theorem proving, inductive strategies are invaluable as they can often be used to provide evidence that suggests a conjecture may be true (or false). Second, the process of producing a proof depends on intuitions about the likely value of different steps or transformations. Intuitions that certain problems are related, or that some problems are more difficult, are expectations derived from experience. The critical point is that some intuitions and perceptions of similarities will be more useful than others. If students employ inappropriate inductive strategies they will not develop adequate mathematical reasoning skills.

Although empirical induction is not an accepted form of proof within mathematics, it is *a* form of justification (and as previously discussed, a very common form among students). If students are encouraged to reason in more familiar ways, inductively, they may come to recognize the limitations of such reasoning with regard to proof as well as the power (in terms of proving) of deductive methods. Moreover, reflecting on the strengths and limitations of inductive argumentation may be an excellent bridge to introduce deductive methods. For example, empirical methods cannot conclusively prove conjectures, but they can conclusively disprove them (by exposing counter-examples). The idea of proof by contradiction could flow naturally from discussion of this feature of inductive reasoning. A similar trajectory might work for introducing mathematical induction as a kind of systemization or grounding of empirical induction. For example, students could be prompted to consider the limits of empirical induction and challenged to identify how (or if) mathematical induction overcomes those limits.

We argue that inductive strategies are an important and valuable part of mathematical reasoning. Yet, even from the perspective that inductive strategies are shortcomings in the long run, there is overwhelming evidence that students do rely on them. Understanding inductive strategies is critical to understanding what students are taking from mathematics instruction. For example, teachers illustrate mathematical concepts with specific instances, but what do students infer from these particular instances? In such cases, are students led to believe that examples suffice as proof?

The overwhelming message from mathematics education and cognitive science is that students do use empirical, inductive, strategies to reason about their world (including mathematics). Mathematics education can either ignore such strategies by treating them as "errors" to be overcome, or it can ask whether there is some value, as instructional tools, or as important mathematical content, to supporting inductive approaches. The perspective from cognitive science emphasizes the value of induction; to be a good reasoner is largely to be a good inductive reasoner. Mathematics may be different, but that difference does not obviate the need for or value of inductive reasoning. The study of adolescents' inductive reasoning in the domain of mathematics is at a very early stage. If the literature on non-mathematical domains is any guide, we should expect to see powerful and sophisticated strategies of inference in the domain of mathematics. Inductive inference is likely a real source of strength upon which mathematics education can build.

References

Balacheff, N. (1987). Treatment of refutations: Aspects of the complexity of a constructivist approach to mathematics learning. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 89–110). The Netherlands: Kluwer Academic Publishers.

Ball, D., Hoyles, C., Jahnke, H., & Movshovitz-Hadar, N. (August, 2002). *The teaching of proof*. Paper presented at the International Congress of Mathematicians, Beijing, China.

Bedard, J., & Chi, M. T. (1992). Expertise. *Current Directions in Psychological Science, 1*, 135-139.

Bell, A. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics, 7,* 23–40.

Bretscher, N. (2003). The spirit of the subject. *Mathematics Teaching*, *184*, 1–7.

Chater, N., & Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford: Oxford University Press.

Christou, C., & Papageorgiou, E. (2007). A framework of mathematics inductive reasoning. *Learning and Instruction*, *17*, 55–66.

Dehaene, S. (1999). *The Number Sense*. New York: Oxford University Press.

Dreyfus, T. (1999). Why Johnny can't prove. *Educational Studies in Mathematics*, *38*(1/3), (pp. 85–109).

Falmagne, R. J., & Gonsalves, J. (1995). Deductive inference. *Annual Review of Psychology, 46*, 525-559.

Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics, 3*(2), 9–24.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist, 50*, 24-37.

Gelman, S. A., & Kalish, C. W. (2006). Conceptual development. In D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol 2, cognition, perception, and language (6th ed.).* (pp. 687–733). Hoboken, NJ: John Wiley & Sons Inc.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological review*, *111*, 3-32.

Gopnik, A., & Schulz, L. (2007). *Causal learning: Psychology, philosophy, and computation*. New York: Oxford University Press.

Hanna G. (2000). A critical examination of three factors in the decline of proof. *Interchange*, *31*(1), 21–33.

Healy, L. & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, *31*(4), 396–428.

Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology, 43*, 243-273.

Jahnke, H. N. (2005). A genetic approach to proof. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education (CERME IV)*, (pp. 428–437). Sant Feliu de Guíxols.

Kalish, C. W., & Lawson, C. A. (2007). Negative evidence and inductive generalization. *Thinking and Reasoning, 13*, 394-425.

Kloosterman, P., & Lester, F. (2004). *Results and interpretations of the 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

Knuth, E. (2002a). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education*, *33*(5), 379–405.

Knuth, E. (2002b). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education*, *5*(1), 61–88.

Knuth, E., Bieda, K., & Choppin, J. (Forthcoming). *The longitudinal development of middle school students' justifying and proving competencies*.

Knuth, E., Choppin, J., & Bieda, K. (2009). Middle school students' production of mathematical justifications. In D. Stylianou, M. Blanton, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 153-170). New York, NY: Routledge.

Kuhn, D. (2002). What is scientific thinking, and how does it develop? In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development.* (pp. 371): Blackwell Publishing.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63*, 171-190.

Martin, T. S., McCrone, S. M. S., Bower, M. L., & Dindyal, J. (2005). The interplay of teacher and student actions in the teaching and learning of geometric proof. *Educational Studies in Mathematics*, *60*(1), 95–124.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, *32*, 49.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608-631.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.

Pillow, B. H. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development, 73*, 779-792.

Polya, G. (1954). *Induction and analogy in mathematics*. Princeton, NJ: Princeton University Press.

Porteous, K. (1990). What do children really believe? *Educational Studies in Mathematics, 21*, 589–598.

Quine, W. (1969). Natural kinds. In N. Rescher et al (Eds.), *Essays in Honor of Carl G. Hempel*, pp. 5-23, Dordrecht: Reidel

RAND Mathematics Study Panel Report. (2002). *Mathematical proficiency for all students: A strategic research and development program in mathematics education*. Washington, DC: U.S. Department of Education.

Recio, A. M., & Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, *48*(1), 83–99.

Rhodes, M., Gelman, S. A., & Brickman, D. (In press). Children's attention to sample composition in learning, teaching, and discovery. *Developmental Science*.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Ross, N., Medin, D., & Cox, D. (2007). Epistemological models and culture conflict: Menominee and Euro-American hunters in Wisconsin. *Ethos, 35*, 478-515.

Simon, M. & Blume, G. (1996). Justification in the mathematics classroom: A study of prospective elementary teachers. *Journal of Mathematical Behavior, 15,* 3–31.

Sowder, L. & Harel, G. (1998). Types of students' justifications. *Mathematics Teacher, 91*(8), 670–675.

Spelke, E. (2000). Core knowledge. *American Psychologist, 55*, 1233-1243.

Stylianides, A. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, *38*(3), 289–321.

van Dormolen, J. (1977). Learning to understand what giving a proof really means. *Educational Studies in Mathematics, 8*, 27–34.

Waring, S. (2000). *Can you prove it? Developing concepts of proof in primary and secondary schools*. Leicester, UK: The Mathematical Association.

Xu, F., & Tenenbaum, J. B. (2007a). Word learning as Bayesian inference. *Psychological review, 114*, 245-272.

Xu, F., & Tenenbaum, J. B. (2007b). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*, 288.

Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to the Principles and Standards for School Mathematics* (pp. 227–236). Reston, VA: NCTM.

Yopp, D. (2009). From inductive reasoning to proof. *Mathematics Teaching in the Middle School*, *15*(5), 286-291.

Grade	**Empirical**	**General**	**Other**
6	37%	21%	42%
7	42%	36%	22%
8	40%	46%	14%

Table 1. Categories of Student Justifications to the Three Odd Numbers Sum item.

Grade	Empirical	General	Other
6	30%	28%	42%
7	28%	32%	40%
8	22%	42%	36%

Table 2. Categories of Student Justifications to the Number Trick item.

Grade	**Empirical**	**General**	**Both**	**Other**
6	37%	32%	20%	11%
7	40%	39%	3%	18%
8	36%	49%	7%	8%

Table 3. Categories of Student Responses to the Consecutive Numbers Sum item.

Principle	**Description: Arguments with…**	**Example for conclusions concerning “all birds” hve X**
Amount	More examples are stronger than fewer	(robins, sparrows, & cardinals have X) > (robins have X)
Diversity	Dissimilar examples are stronger than similar	(robins, hawks, & penguins have X) > (robins, sparrows, & cardinals have X)
Typicality	Typical examples are stronger than atypical	(robins have X) > (penguins have X)
Contrast	Negative examples are stronger than those without	(robins have X, cats lack X) > (robins have X)

Table 4. Some Examples of Criteria for Evidential Strength.

Category	**Meaning**
Parity	Even vs. odd
Multiples	3, 9, 36, 81, 90 are all multiples of 3
Factors	3, 9, 15, 30, and 90 all go into 90 evenly
Prime	2, 5, 11, 17 go together because they're prime
Composite	25, 30, and 60 are all composite numbers
Squared	4, 25, and 81 are perfect squares
Sequence	3, 9, and 15 go together because they go up by 6
Value of digit	1, 11, 21, and 81 all have a "1" at the same spot
Intervals	11, 14, 15, 17 are all between 10 and 20
Number of digits	1, 2, 3, 5, and 9 are all one-digit numbers
Contains a digit	5, 15, and 51 all contain a 5 so they belong together
Arithmetic	2, 3, and 5 are a group because 2 + 3 = 5
Relational	100, 25, 36, & 9 because 100/25 = 4 and 36/9 = 4

Table 5: Number categories and their meanings.

Principle	**Middle School**	**Undergrad**	**STEM**
Multiples	86% (1)	93% (2)	93% (1)
Parity	79% (2)	93% (1)	87% (2)
Prime	50% (3)	71% (3)	80% (3)
Vale of Digit	50% (3)	29% (9)	47% (6)
# of Digits	43% (5)	36% (6)	20% (8)
Intervals	36% (6)	50% (4)	53% (4)
Contains Digit	29% (7)	43% (5)	13%
Squared	21% (8)	36% (6)	53% (4)
Sequence	21% (8)	14% (11)	20% (8)
Factors	21% (8)	0% (12)	0% (13)
Arithmetic	14% (11)	36% (6)	40% (7)
Composite	14% (11)	21% (10)	20% (8)
Relational	7% (13)	0% (12)	13% (8)

Table 6: Number categories sorted according to salience.

Category	What it Means
# of Sides	Grouping shapes with the same number of sides
Angles	Grouping shapes on angle size (all have an obtuse)
Equal sides	Two or more equal sides
Regular	Grouping regular shapes together
Shape	Resemblance to a shape (e.g., “arrows”, “sharp”)
Familiar	Common shapes you see in school
Size	Grouping large or small shapes together
Orientation	Grouping according to orientation on paper
Compose	A group of shapes that can be made from others
Tessellate	Grouping shapes that would tessellate
Similar	Grouping similar shapes together
Symmetry	Grouping symmetric shapes together
Convex/Concave	Grouping according to concavity

Table 7: Shape categories and their meanings.

Principle	**Middle School**	**Undergrad**	**STEM**
Size	78% (2)	85% (3)	100% (3)
Shape	64% (2)	78% (4)	93% (2)
Angles	64% (2)	85% (2)	80% (4)
Similar	50% (5)	43% (8)	40% (11)
Concavity	43% (5)	14% (8)	40% (7)
Equal Sides	36% (5)	21% (6)	26% (11)
Composition	36% (10)	43% (9)	27% (8)
Familiar	35% (5)	43% (5)	34% (8)
Symmetry	21% (8)	21% (6)	47% (5)
Orientation	14% (8)	7% (9)	27% (8)
Regular	14% (11)	0% (11)	53% (6)
Tessellate	7% (11)	0% (11)	0% (12)
# of Sides	100% (1)	100% (1)	100% (1)

Table 8: Shape categories sorted according to salience.