notes

WHAT COULD THEY POSSIBLY BE THINKING!?!

UNDERSTANDING YOUR COLLEGE MATH STUDENTS

David Kung
and
Natasha Speer

MAA PRESS

What Could They Possibly Be Thinking!?!

Understanding your college math students

"After mastering mathematical concepts, even after great effort, it becomes very hard to put oneself back into the frame of mind of someone to whom they are mysterious."

— Fields Medal winner William Thurston

Library of Congress Control Number: 2020934681

Print ISBN 978-0-88385-209-5
Electronic ISBN 978-1-61444-326-1

Printed in the United States of America

Current Printing (last digit):
10 9 8 7 6 5 4 3 2 1

What Could They Possibly Be Thinking!?!

Understanding your college math students

Dave Kung
St. Mary's College of Maryland

and

Natasha Speer
The University of Maine

MAA

Published and Distributed by
The Mathematical Association of America

The MAA Notes Series, started in 1982, addresses a broad range of topics and themes of interest to all who are involved with undergraduate mathematics. The volumes in this series are readable, informative, and useful, and help the mathematical community keep up with developments of importance to mathematics.

MAA Notes

14. Mathematical Writing, by *Donald E. Knuth, Tracy Larrabee, and Paul M. Roberts.*
16. Using Writing to Teach Mathematics, *Andrew Sterrett,* Editor.
17. Priming the Calculus Pump: Innovations and Resources, Committee on Calculus Reform and the First Two Years, a subcomittee of the Committee on the Undergraduate Program in Mathematics, *Thomas W. Tucker,* Editor.
18. Models for Undergraduate Research in Mathematics, *Lester Senechal,* Editor.
19. Visualization in Teaching and Learning Mathematics, Committee on Computers in Mathematics Education, *Steve Cunningham and Walter S. Zimmermann,* Editors.
20. The Laboratory Approach to Teaching Calculus, *L. Carl Leinbach et al.,* Editors.
21. Perspectives on Contemporary Statistics, *David C. Hoaglin and David S. Moore,* Editors.
22. Heeding the Call for Change: Suggestions for Curricular Action, *Lynn A. Steen,* Editor.
24. Symbolic Computation in Undergraduate Mathematics Education, *Zaven A. Karian,* Editor.
25. The Concept of Function: Aspects of Epistemology and Pedagogy, *Guershon Harel and Ed Dubinsky,* Editors.
26. Statistics for the Twenty-First Century, *Florence and Sheldon Gordon,* Editors.
27. Resources for Calculus Collection, Volume 1: Learning by Discovery: A Lab Manual for Calculus, *Anita E. Solow,* Editor.
28. Resources for Calculus Collection, Volume 2: Calculus Problems for a New Century, *Robert Fraga,* Editor.
29. Resources for Calculus Collection, Volume 3: Applications of Calculus, *Philip Straffin,* Editor.
30. Resources for Calculus Collection, Volume 4: Problems for Student Investigation, *Michael B. Jackson and John R. Ramsay,* Editors.
31. Resources for Calculus Collection, Volume 5: Readings for Calculus, *Underwood Dudley,* Editor.
32. Essays in Humanistic Mathematics, *Alvin White,* Editor.
33. Research Issues in Undergraduate Mathematics Learning: Preliminary Analyses and Results, *James J. Kaput and Ed Dubinsky,* Editors.
34. In Eves' Circles, *Joby Milo Anthony,* Editor.
35. You're the Professor, What Next? Ideas and Resources for Preparing College Teachers, The Committee on Preparation for College Teaching, *Bettye Anne Case,* Editor.
36. Preparing for a New Calculus: Conference Proceedings, *Anita E. Solow,* Editor.
37. A Practical Guide to Cooperative Learning in Collegiate Mathematics, *Nancy L. Hagelgans, Barbara E. Reynolds, SDS, Keith Schwingendorf, Draga Vidakovic, Ed Dubinsky, Mazen Shahin, G. Joseph Wimbish, Jr.*
38. Models That Work: Case Studies in Effective Undergraduate Mathematics Programs, *Alan C. Tucker,* Editor.

39. Calculus: The Dynamics of Change, CUPM Subcommittee on Calculus Reform and the First Two Years, *A. Wayne Roberts,* Editor.

40. Vita Mathematica: Historical Research and Integration with Teaching, *Ronald Calinger,* Editor.

41. Geometry Turned On: Dynamic Software in Learning, Teaching, and Research, *James R. King and Doris Schattschneider,* Editors.

42. Resources for Teaching Linear Algebra, *David Carlson, Charles R. Johnson, David C. Lay, A. Duane Porter, Ann E. Watkins, William Watkins,* Editors.

43. Student Assessment in Calculus: A Report of the NSF Working Group on Assessment in Calculus, *Alan Schoenfeld,* Editor.

44. Readings in Cooperative Learning for Undergraduate Mathematics, *Ed Dubinsky, David Mathews, and Barbara E. Reynolds,* Editors.

45. Confronting the Core Curriculum: Considering Change in the Undergraduate Mathematics Major, *John A. Dossey,* Editor.

46. Women in Mathematics: Scaling the Heights, *Deborah Nolan,* Editor.

47. Exemplary Programs in Introductory College Mathematics: Innovative Programs Using Technology, *Susan Lenker,* Editor.

48. Writing in the Teaching and Learning of Mathematics, *John Meier and Thomas Rishel.*

49. Assessment Practices in Undergraduate Mathematics, *Bonnie Gold,* Editor.

50. Revolutions in Differential Equations: Exploring ODEs with Modern Technology, *Michael J. Kallaher,* Editor.

51. Using History to Teach Mathematics: An International Perspective, *Victor J. Katz,* Editor.

52. Teaching Statistics: Resources for Undergraduate Instructors, *Thomas L. Moore,* Editor.

53. Geometry at Work: Papers in Applied Geometry, *Catherine A. Gorini,* Editor.

54. Teaching First: A Guide for New Mathematicians, *Thomas W. Rishel*.

55. Cooperative Learning in Undergraduate Mathematics: Issues That Matter and Strategies That Work, *Elizabeth C. Rogers, Barbara E. Reynolds, Neil A. Davidson, and Anthony D. Thomas,* Editors.

56. Changing Calculus: A Report on Evaluation Efforts and National Impact from 1988 to 1998, *Susan L. Ganter*.

57. Learning to Teach and Teaching to Learn Mathematics: Resources for Professional Development, *Matthew Delong and Dale Winter.*

58. Fractals, Graphics, and Mathematics Education, Benoit Mandelbrot and Michael Frame, Editors.

59. Linear Algebra Gems: Assets for Undergraduate Mathematics, *David Carlson, Charles R. Johnson, David C. Lay, and A. Duane Porter,* Editors.

60. Innovations in Teaching Abstract Algebra, *Allen C. Hibbard and Ellen J. Maycock,* Editors.

61. Changing Core Mathematics, *Chris Arney and Donald Small,* Editors.

62. Achieving Quantitative Literacy: An Urgent Challenge for Higher Education, *Lynn Arthur Steen.*

64. Leading the Mathematical Sciences Department: A Resource for Chairs, *Tina H. Straley, Marcia P. Sward, and Jon W. Scott,* Editors.

65. Innovations in Teaching Statistics, *Joan B. Garfield,* Editor.

66. Mathematics in Service to the Community: Concepts and models for service-learning in the mathematical sciences, *Charles R. Hadlock,* Editor.

67. Innovative Approaches to Undergraduate Mathematics Courses Beyond Calculus, *Richard J. Maher,* Editor.

68. From Calculus to Computers: Using the last 200 years of mathematics history in the classroom, Amy Shell-Gellasch and Dick Jardine, Editors.

69. A Fresh Start for Collegiate Mathematics: Rethinking the Courses below Calculus, *Nancy Baxter Hastings,* Editor.

70. Current Practices in Quantitative Literacy, *Rick Gillman*, Editor.

71. War Stories from Applied Math: Undergraduate Consultancy Projects, *Robert Fraga*, Editor.

72. Hands On History: A Resource for Teaching Mathematics, *Amy Shell-Gellasch*, Editor.

73. Making the Connection: Research and Teaching in Undergraduate Mathematics Education, *Marilyn P. Carlson and Chris Rasmussen*, Editors.

74. Resources for Teaching Discrete Mathematics: Classroom Projects, History Modules, and Articles, *Brian Hopkins*, Editor.

75. The Moore Method: A Pathway to Learner-Centered Instruction, *Charles A. Coppin, W. Ted Mahavier, E. Lee May, and G. Edgar Parker*.

76. The Beauty of Fractals: Six Different Views, *Denny Gulick and Jon Scott*, Editors.
77. Mathematical Time Capsules: Historical Modules for the Mathematics Classroom, *Dick Jardine and Amy Shell-Gellasch*, Editors.
78. Recent Developments on Introducing a Historical Dimension in Mathematics Education, *Victor J. Katz and Costas Tzanakis*, Editors.
79. Teaching Mathematics with Classroom Voting: With and Without Clickers, *Kelly Cline and Holly Zullo*, Editors.
80. Resources for Preparing Middle School Mathematics Teachers, *Cheryl Beaver, Laurie Burton, Maria Fung, and Klay Kruczek*, Editors.
81. Undergraduate Mathematics for the Life Sciences: Models, Processes, and Directions, *Glenn Ledder, Jenna P. Carpenter, and Timothy D. Comar*, Editors.
82. Applications of Mathematics in Economics, *Warren Page*, Editor.
83. Doing the Scholarship of Teaching and Learning in Mathematics, *Jacqueline M. Dewar and Curtis D. Bennett*, Editors.
84. Insights and Recommendations from the MAA National Study of College Calculus, *David Bressoud, Vilma Mesa, and Chris Rasmussen,* Editors.
85. Beyond Lecture: Resources and Pedagogical Techniques for Enhancing the Teaching of Proof-Writing Across the Curriculum, *Rachel Schwell, Aliza Steurer and Jennifer F. Vasquez,* Editors.
86. Using the Philosophy of Mathematics in Teaching Undergraduate Mathematics, *Bonnie Gold, Carl E. Behrens, and Roger A. Simons,* Editors.
87. The Courses of History: Ideas for Developing a History of Mathematics Course, *Amy Shell-Gellasch and Dick Jardine,* Editors.
88. Shifting Contexts, Stable Core: Advancing Quantitative Literacy in Higher Education, *Luke Tunstall, Gizem Karaali, and Victor Piercey,* Editors.
89. MAA Instructional Practices Guide, *Martha L. Abell, Linda Braddy, Doug Ensley, Lewis Ludwig, Hortensia Soto,* Project Leadership Team.
90. What Could They Possibly Be Thinking!?! Understanding your college math students, *Dave Kung and Natasha Speer*.

MAA Service Center
P.O. Box 91112
Washington, DC 20090-1112
1-800-331-1MAA FAX: 1-301-206-9789

Preface

This book was conceived over a couple of tasty pies at Pizzaria Bianco in Phoenix during the 2004 Joint Mathematics Meetings. “Mathematics education researchers know so much about student thinking,” we opined to our friend Gavin LaRose. “But mathematicians—the people doing the teaching—don’t know about it” Gavin didn’t miss a beat, “The two of you should write that book!” Although many college mathematics instructors may know a great deal about how students think about particular ideas, it is common for that knowledge to have developed from years of experience doing things such as looking at students’ writing on exams and talking with students during office hours. This book is a resource to help accelerate this learning process so that more instructors of college mathematics can use knowledge of student thinking from early in their careers as they work to design and offer learning opportunities to their students.

And so it began.

This book’s 15 year gestation period spanned one job change, three sabbaticals, three promotions, and endless other endeavors. However, the need to bridge the chasm between the mathematics education world (where Natasha grew up) and the land of theoretical mathematics (Dave’s native land) has never been bigger. Still 15 years after gobbling down that pizza, most college mathematics teachers remain unaware of reams of important work on how their students understand (or don’t understand) the topics in the undergraduate mathematics curriculum.

We hope this book helps pull our two worlds slightly closer together, providing college mathematics instructors with insights that will help them better serve their students.

Why does our community need this book?

While more than a decade has passed since we began this project, its completion couldn’t have been more timely. Leaders point to science, technology, engineering, and mathematics (STEM) as providing hope for overcoming economic and other challenges. In 2012, the President’s advisors insisted that to have an adequate number of people for STEM jobs, we (universities, colleges) need to produce 1 million **additional** graduates with STEM degrees within a decade [8]. As if that challenge were not daunting enough, fewer than half of the students who enter college intending to major in a STEM field actually complete such a major [8]. Students drop STEM degrees for many reasons, but many who switch to non-STEM majors earned high grades in their introductory STEM courses. Contrary to views sometimes expressed in our community, introductory STEM courses are actually “weeding out” some very capable students who are quite successful in our courses. And those “high-performing students frequently cite uninspiring introductory courses as a factor in their choice to switch majors” [8, p. i]. Many of those leaving STEM majors (in the neighborhood of 90%) state that poor teaching contributed to their decision to leave [11].

As a gateway to other STEM majors, mathematics plays a key role in attracting and retaining students. The learning opportunities we provide determine whether students make it through the mathematics gate and into their STEM careers. “The first two years of college are the most critical to the retention and recruitment of STEM majors” [8, p. ii]. Instructors of undergraduate mathematics courses, whether faculty or graduate students, can make substantial contributions to the enrollment and retention rates for STEM majors by improving the learning opportunities they provide to students.

What can we do? Teaching methods that interactively engage students can play a vital role in improving student success, recruitment and retention in mathematics courses and in STEM disciplines.

In case we're not just preaching to the choir on this point, we find the research on the effectiveness of interactive teaching methods extremely persuasive. Students in courses dominated by passive lectures have higher drop/fail/withdrawal (DFW) rates—over 50% higher—than other course formats [6]. The differences between passive and active teaching methods are so stark that the Freeman report, a meta-analysis of over 200 separate studies of college STEM teaching, suggests that a medical study with such drastic results would be stopped prematurely for ethical reasons. The control group (passive lectures) would be given the treatment (anything more interactive) [6]. If those reasons aren't enough to prompt pedagogical change, active approaches have also proven to be particularly beneficial to traditionally under-represented students, including women and minorities [6, 9]. In addition, results from a national study of undergraduate calculus showed that programs found to be effective (as measured by outcomes, retention rates, etc.) include the use of student-centered and active-learning approaches to instruction [2]. On the basis of these and other findings, The Conference Board of the Mathematical Sciences endorsed such approaches to instruction in their statement on *Active Learning in Post-Secondary Mathematics Education* [3].

To leverage these research findings, professional development (PD) for college STEM teachers has turned its focus toward preparing instructors to actively engage their students in the learning process. In the math world, many PD programs for graduate student teaching assistants (TAs) focus on using group work in discussion sections. In our work with the NSF-funded College Mathematics Instructor Development Source project (CoMInDS), we help those running such programs retool their curricula to better prepare TAs for interactive teaching. Most of the PD activities in the College Math Video Cases, developed by a team we were honored to be included in, focus on active learning strategies for new college math instructors. When those graduate students get faculty positions, they may have opportunities to participate in additional teaching-focused professional development either on their home campuses or through national programs such as MAA Project NExT (a program which Dave now directs), American Mathematical Association of Two-Year College (AMATYC) Project ACCCESS or The Academy of Inquiry Based Learning, all of which include sessions focused on research-based, interactive methods.

At all of the stages of a math faculty member's career, the Mathematical Association of America's *Guide to Evidence-Based Instructional Practices in Undergraduate Mathematics* [10] provides an amazing resource that surveys interactive, research-based pedagogies. With colorful vignettes, it describes a wide range of ways that instructors can implement the engaged approaches to instruction that research shows will enhance student learning opportunities in mathematics classes.

Our community's transition from passive lectures to interactive methods has been neither easy nor quick, in large part because it comes with a change in what knowledge is required for teachers. While subject matter knowledge is vital for all teaching styles, knowledge of student thinking is especially crucial when instructors implement active learning approaches (e.g., [1, 5]). As our community asks (begs?) instructors to adopt more interactive methods, they'll need more knowledge of student thinking to be successful. They need to anticipate and be able to understand what students do and say while working on problems or engaging in some inquiry-oriented assignment. They need to correctly diagnose student difficulties and, instead of just telling them "the right way" to think, respond in ways that will help students see their errors and then refine their reasoning.

As we mentioned above, typically, instructors gain such knowledge, bit by bit, over years of experience interacting with students in class, talking with them during office hours, and reading and trying to make sense of students' written work on homework and tests. Our goal in writing this book is to expedite that learning process by providing instructors with descriptions of how students think about key ideas in undergraduate mathematics. We aim to arm them with knowledge of the typical ways that students think (both correctly and incorrectly) about key ideas. We believe that instructors who know how students think are more likely to:

- anticipate common ways of thinking as well as difficulties and plan instruction accordingly
- recognize a variety of ways students approach particular problems and correctly diagnose the source of student difficulties
- write engaging activities that leverage likely ways students will think and then help students learn challenging ideas

If the process of learning about student thinking is expedited, instructors will be able to provide higher quality instruction earlier in their careers, student learning will improve, fewer STEM-interested students will leave the pipeline, more will graduate, we'll reach the one million additional STEM graduates, and the economy (and our jobs) will be saved!

Hyperbole aside (if it annoys you, blame Dave), our perspective boils down to a simple idea, supported by research:

> The more you know about how students think, the more your students will learn [4, 7].

Who might benefit from this book?

We created this book with three primary audiences in mind.

Novice college mathematics instructors, including graduate students and faculty. Novice instructors can use this book as a reference, reading the chapters relevant to their teaching responsibilities as they plan their courses, prepare daily lessons, and design student assignments and assessments. The chapters provide a guided tour of the ideas and topics that are apt to be especially challenging for students as well as insights into how students think about those ideas as they are in the process of learning. The goal is for this information to benefit novice instructors by providing them with wisdom that might otherwise only be acquired over years of experience teaching.

Those who provide teaching-related professional development to novice college mathematics instructors (e.g., instructors of graduate student teaching seminars). Novice instructors can also be exposed to the contents of the book via a professional development seminar. For example, those who design and provide such seminars can plan readings from the book such that seminar participants are reading chapters a week or two prior to when that content will be addressed. Even if participants in such a seminar are not currently teaching, assignments might involve creating a lesson or assessment informed by information in a particular chapter. Becoming familiar with how students think about that content will be useful later when instructors are facing the challenges of teaching it.

Experienced instructors teaching a course for the first time. Even the most experienced teachers are sometimes asked to teach a new course or topic. This book can serve as a reference as instructors teach a new course—or one they haven't taught in years. Instead of relying only on their own recollections from learning the content (oh so many years ago ...), the chapters provide distillations of information to help ground their class plans in what they are apt to see in typical students' ways of thinking. In addition, even for experienced instructors, occasionally students do or say perplexing things. Referring to specific examples in the relevant chapters of the book can provide insight into a way of thinking that was unfamiliar. These previously-unseen ways of thinking are especially likely to surface when experienced instructors incorporate new instructional practices into their classes such as collaborative groupwork, clickers, or other active learning approaches.

Anatomy of a Chapter

Chapter sections are organized around examples of student work, to showcase student thinking. Following a brief introduction, we present sets of questions and student answers, each of which has the same format:

- a mathematical task
- sample student responses to that task
- analysis/discussion of those responses.

The content of the chapters is based on findings from research. Many of the tasks that appear in the "question and student answers" sections are taken directly from research papers about student thinking and learning. In these cases, the sample student work may be reproduced verbatim. For some tasks discussed in the literature, we generated sample student work based on research findings, informed by our own experiences working with students. Other tasks and student responses were created to illustrate ideas from research but were not modeled directly on any specific research study. Citations indicate whether we took a problem or student work directly from the research or adapted it for our purposes.

As you read, you are strongly encouraged to try out the problem yourself and to anticipate how students might answer prior to looking at the sample student solutions. This will help you get the most out of the discussion of student thinking and is the best we can do to create an active, engaged learning experience for you via text.

Many researchers suggested ways to address known student difficulties and to help students develop strong understanding of the ideas. Those are discussed in the latter part of each chapter in **What You Can Do**. While most of the suggestions are supported by citations to research, occasionally we have drawn from our own experiences and from those of our reviewers.

Each chapter concludes with **A Good Place to Start**. This is a single suggestion, informed by the research, that we find especially helpful when teaching that chapter's material. These suggestions, grounded in our reading of the available research, represent our own opinions. We hope instructors will implement these ideas as first steps towards incorporating many of the ideas contained in the chapters.

We deviate somewhat from this format in Chapter Zero, which contains information about how students think about select, key ideas from K–12 mathematics. These ideas are ones known to be foundational to college mathematics and an underdeveloped understanding of them can interfere with students learning new ideas. For each of these we provide a mini-chapter including sample student work, analysis and teaching recommendations.

Acknowledgements

Each chapter was reviewed by a mathematics education researcher with expertise in student thinking about that content, as well as a mathematics professor with extensive teaching experience in that area.

Lara Alcock	Jess Ellis	Alex Meadows
Spencer Bagley	Shawn Firouzian	David Meel
Lake Bookman	Sandy Ganzell	Teri Jo Murphy
Josh Case	Kent Johnson	Michael Oehrtman
Allison Dorko	Matt Jones	Chris Rasmussen
Casey Douglas	Sean Larsen	Vicki Sealey
Paula Drewniany	Sara Malec	Jen Tyne
Ed Dubinsky		

We thank them for their constructive feedback and for catching various errors and typos. We also extend our thanks to Bruce Birkett and Ian Thackray for their comments and for helping us get our physics right. Any remaining errors are probably Dave's fault, but let's just say that we take joint responsibility for them. We are also grateful to Kelcey Landgraf who provided assistance with the preparation of materials for some of the chapters.

We are indebted to our institutions for providing us with the time and support to engage in this long-term project, and to our families for doing the same.

Finally, thanks to Gavin for suggesting that we take on this project. The next pizza's on us.

Dave Kung
Mathematics & Computer Science
St. Mary's College of Maryland
Director, MAA Project NExT
Director of Strategy and Implementation,
Transforming Post-Secondary Education - Math (TPSE-Math)

Natasha Speer
Mathematics & Statistics
Maine Center for Research in STEM Education
The University of Maine

Bibliography

[1] D. Ball, M. Hoover Thames, and G. Phelps. Content knowledge for teaching: What makes It special? *Journal of Teacher Education*, 59(5):389–407, 2008.

[2] David M. Bressoud, Vilma Mesa, and Chris Rasmussen. *Insights and Recommendations from the MAA National Study of College Calculus*. MAA Notes.

[3] Conference Board of the Mathematical Sciences. Active Learning in Post-Secondary Mathematics Education. www.cbmsweb.org/2016/07/active-learning-in-post-secondary-mathematics-education/, 2016. Accessed: 2018-06-06.

[4] E. Fennema, T. Carpenter, M. Franke, Linda Levi, Victoria R Jacobs, and Susan B Empson. A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, 27(4):403–434, 1996.

[5] E. Fennema, M. Franke, T. Carpenter, and D. Carey. Using children's mathematical knowledge in instruction. *American Educational Research Journal*, 30(3):555–583, 1993.

[6] M. P. Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

[7] H. Hill, B. Rowan, and Deborah L. Ball. Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2):371–406, 2005.

[8] J. P. Holdren and E. S. Lander. Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Technical report, President's Council of Advisors on Science and Technology, Washington, D.C., 2012.

[9] Sandra L. Laursen, Marja-Liisa Hassi, Marina Kogan, and Timothy J. Weston. Benefits for Women and Men of Inquiry-Based Learning in College Mathematics: A Multi-Institution Study. *Journal for Research in Mathematics Education*, 45(4):406–418, 2014.

[10] MAA. *MAA Instructional Practices Guide*. maa.org/ip-Guide. Accessed: 2018-06-06.

[11] Elaine Seymour and Nancy Hewitt. Talking About Leaving: Factors Contributing to High Attrition Rates Among Science, Mathematics, and Engineering Undergraduate Majors, 1994.

Contents

0

Foundational Ideas

Introduction

Chances are that you have had the experience of trying to learn something new and discovering that there were relevant, related ideas that you needed to understand better. Teachers and researchers have noticed this over and over again in their observations of students. For students who are learning undergraduate mathematics, years and years of more basic mathematics are vital as they learn ideas in calculus and beyond. Researchers have found that students' understanding of elementary, middle, and high school mathematics ideas can have a profound impact on their abilities to learn undergraduate mathematics—a strong understanding can help facilitate that learning and a weak understanding can make it challenging or even prevent learning from occurring.

There is much that undergraduate mathematics instructors can do to help strengthen students' understanding of these fundamental ideas. But to help, first you need to be able to recognize the symptoms of these issues as they appear as students are learning undergraduate mathematics. That's where this chapter comes into play—it is sort of a field guide to the things you may see or hear as students work on undergraduate mathematics tasks that are symptoms of under-developed understandings of K–12 mathematics concepts. We include the kinds of student work and thinking you might observe, explanations for the sources of that thinking that are connected to their thinking about earlier ideas, and suggestions for ways to help your students enhance their understanding of some fundamental ideas that can help them be better equipped to learn the undergraduate mathematics ideas you are teaching to them. We also discuss some issues related to language, notation and representations that may materialize in your students' work across multiple topics in the undergraduate curriculum.

For each topic, we provide:

- a mathematical task
- sample student responses to that task
- analysis/discussion of those responses
- discussion of what you can do

The structure of each of these mini-chapters mimics the structure used in later chapters that are devoted to a single topic. Many of these topics show up in multiple parts of the undergraduate curriculum and, as a result, the ordering used in this chapter is not chronological (as is the case in the overall book and in most other chapters).

Being able to do this kind of detective and diagnostic work is valuable no matter what instructional approach is being used. However, when active learning or inquiry-based approaches are used instructors are more likely to see and hear how students are thinking and therefore are more likely to have opportunities to diagnose and address these less-than-fully-formed ways of thinking.

Active learning approaches where students spend time working on problems during class are also apt to give instructors a window into students' skills and understandings of prior-to-calculus ideas and this chapter is designed to help raise instructors' awareness of the symptoms of student difficulties with these ideas and to provide insights into that thinking so the difficulties can be addressed.

Remember: To get the most out of the discussion about student answers and thinking, try each problem yourself and generate a few different ways students might respond before looking at the sample student work.

The Distributive Property (and Factoring)

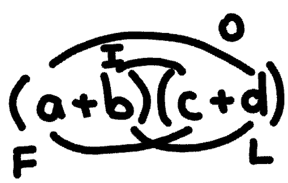
1. Anna's work:
Given that $y = 3(x^2 + 7x)$, Find y'.

$$y' = 3(2x+7) = 6x+7$$

The first stop on our tour of thinking for this chapter is something veteran college mathematics instructors are very familiar with—troubles with distributing. Anna correctly takes the derivative but then encounters some trouble. She uses the distributive property appropriately for part of the calculation, but not so with the second term in the parentheses, and writes 7 instead of 21. Instructors may see similar work when students are asked to multiply two binomials. For example, $(x + 2y)(x^2 + 4y)$ may generate answers such as $x^3 + 8y^2$ and $(a + b)^2$ is apt to occasionally be simplified to $a^2 + b^2$.

What are these students thinking (or failing to think) when carrying out computations that they've been working on since Algebra I?

For some, a quick reminder about FOIL (first, outer, inner, last) might be all that is needed to refresh their understanding of the fundamental properties of multiplication. But for others (many, perhaps), these calculations are a symptom of an under-developed understanding of multiplication. In those cases, at best, when FOIL is mentioned it cues some procedure or some collection of arcs linking various terms in the problem:



College students, even quite successful ones, may be unable to explain why these procedures (when used correctly) work and may have little to fall back on to figure out these answers other than the (mis)remembered procedure.

Similar issues are seen when students are asked to reverse the process and factor an expression. Other than guess and check, the quadratic formula, and perhaps a few procedures (factor out the leading coefficient, group like terms, etc.), students may lack tools or problem solving strategies to help them when faced with a factoring problem.

What You Can Do

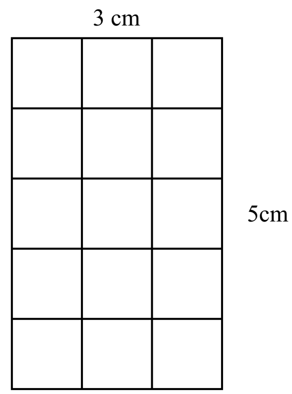
The concept underlying distribution and factoring is multiplication. Representing the result of a multiplication as the area of a rectangle has been shown to be a very powerful tool for understanding why FOIL and other things work and for figuring out answers without a memorized procedure [2].

Most students were first introduced to the idea of multiplication as a description of the process of adding something a certain number of times. This *repeated addition* portrayal of multiplication can be illustrated with story problems such as "Every classroom has two blackboards. There are four classrooms in the building. How many blackboards are in the building?" And it can be modeled with $2 + 2 + 2 + 2$. The result of this computation gives the answer of "8 blackboards." Note that the unit of the answer is the same as the unit of one of the things in the problem statement ("two blackboards"). This repeated addition way of describing multiplication is very commonly used to teach multiplication and is typically well understood by students. This is referred to as *scalar multiplication* because we are "scaling" (up or down) the amount of something (or number of somethings) you have. You probably heard about this in your linear algebra class when you did things like this:

$$3\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 3a & 3b \\ 3c & 3d \end{pmatrix}.$$

In these kinds of problems you are scaling each entry by a factor of 3.

In many cases, a more powerful model for multiplication is area. The computation 3×5 could also give you the area of a rectangle (see Figure 0.1) that is 3 units by 5 units (or, say 3 cm by 5 cm). The answer (15 units^2 or 15 cm^2) is not measured in the same units we used to measure the quantities we started with.



Area = 3 cm × 5 cm = 15 cm^2

Figure 0.1.

Those with a sophisticated understanding of multiplication can likely explain FOIL or other steps they use to get the solution. For example, to explain $3(x + 7) = 3x + 21$, someone might draw what is shown in Figure 0.2.

	x	7
3	$3x$	21

Figure 0.2.

This representation can then be used to illustrate the simplified version of expressions such as $(x+3)(x+5)$, shown in Figure 0.3.

	x	3
x	x^2	$3x$
5	$5x$	15

Figure 0.3.

Your college math students are unlikely to have seen this area model approach for multiplication of binomials. Using it a few times is an effective way to help students build their understanding and improve their algebra computation skills. This area model for multiplication can also be used to support work on factoring tasks of the sort encountered in min-max and other problems. For example, if you have $8x^2 + 22x + 15$, you can begin with a blank 2×2 grid and place the first and last terms on one diagonal, as seen in Figure 0.4.

Then the $22x$ needs to be spread between the two remaining boxes in a way so that the lengths of the sides generate the appropriate products for each term (see Figure 0.5). This draws on guess and check strategies (and number facts)

$8x^2$	
	15

Figure 0.4.

	$2x$	3
$4x$	$8x^2$	$12x$
5	$10x$	15

Figure 0.5.

but does so in a more systematic way than what is typically seen when people approach these tasks in a linear (non-area) fashion. And as with the binomial multiplication example illustrated in Figure 0.3, showcasing this technique can help your calculus students solidify their factoring skills so their difficulties with this topic are not the obstacle to their success on min-max and other problems they encounter that require these factoring skills.

About that Incomplete Square

2. Use an appropriate technique of integration to compute $\int \frac{1}{x^2+4x+5}\,dx$.

Bill:

$x^2+4x+5=(x+?)^2+?$

something with b/2 ??

Can't remember!

As part of the techniques of integration section of a calculus course students may be expected to tackle integrals such as the one in Problem 2 by transforming the denominator into the sum of two squares and then utilizing some sort of trigonometric substitution. For some students, this produces little more than vague memories of some formula involving $\frac{b}{2}$. At the heart of this procedure, however, is the idea of "completing" the values in an area representation of a square - something that is very similar to the process described above for rectangles.

The goal is to transform x^2+4x+5 so it is the sum of two squares. To accomplish this, we create a square (in the geometric sense), place the x^2 in one quadrant and then *equally* divide the 4 xs between two quadrants (that's why it's $\frac{b}{2}$ n the formula Bill has partially recalled) like this:

	x	2
x	x^2	$2x$
2	$2x$	

Figure 0.6.

Now "complete the square" by placing a 4 in the lower right corner box. Your created square has area x^2+4x+4. All that remains is to see what the difference is between the area of that square and the value you started with. In this case, we get $x^2+4x+5=x^2+4x+1+4=(x+2)^2+1$.

The Equal Sign

3. Solve for y in the equation $5y - 7 = 15x + 3$.

Claudette:

$$5y - 7 = 15x + 3 + 7$$
$$= 15x + 10/5$$
$$= 3x + 2$$
$$\text{So } y = 3x + 2$$

Such blatant disregard for the sanctity of the equal sign! If you've spent any time looking at college students' work, you've likely seen many examples of this sort. Although it may be tempting to chalk this up to sloppy work habits or poorly learned algorithms for solving equations, the root cause may actually be the misapplication of an algorithm that served them very well in the early phrases of their mathematical careers [3, 6]. Becoming a well-informed detective who can recognize the symptoms of this difficulty can be very useful as you try to make sense of lines and lines of the algebra your students generate as they work on various problems from calculus and beyond.

Typically, students are first introduced to the equal sign in the context of simple arithmetic problems. These are often formatted using a box for where the answer belongs:

$$7 + 11 = \boxed{}$$

Students encounter many examples like this, with numbers on the left and a place for the answer on the right, separated by the equal sign. Students learn that this arrangement of symbols means they are to "do something" to the things on the left and put the result of that work on the right. Some students may also treat the equal sign as a form of punctuation used to connect a sequence of steps they are carrying out. Viewing the equal sign as a "do something" symbol serves them very well as they proceed through elementary school where most of their encounters with the equal sign are completely compatible with this view of it.

More sophisticated students might have advanced to "do the same thing" to *both* sides of the equal sign. Such strategies, largely successful, can still mask an impoverished understanding of equality. They might not understand which transformations preserve the truth of the equation (e.g., differentiation) and which do not (e.g., anti-differentiation—because of that pesky constant of integration).

Unfortunately, when things get more complex, such as in a first algebra class, students may not have adequate opportunities to learn that the equal sign is no longer to be interpreted as a signal to manipulate the left side of the equation. This can result in strings of seemingly illogical terms such as the ones Claudette produces above.

Here Claudette sees the $5y - 7$ on the left side of the equal sign, springs into action and, doing her best to apply the "do something" algorithm, adds 7 to the right-hand side of the equation. This is of course a fine way to proceed, assuming that you do not maintain the equality link between your answer and the original left-hand term of the equation. Moving along, Claudette carries out all the correct steps to isolate y but by operating repeatedly only on the right-hand side of the equation, she generates a string of non-equal terms—all leading up to the correct conclusion.

What You Can Do

Be on the lookout for students who generate strings of expressions inappropriately chained together with equal signs. Recognizing this as a symptom of possible difficulties with equality instead of a neatness issue can give you opportunities to tackle the issue head on, at moments when it will be most relevant and meaningful to students. Telling them directly that "=" is used in different ways when introducing arithmetic than it is in algebra, followed with a few examples that illustrate the differences such as the ones above, can help them update their understanding in ways that will pay dividends when working on all sorts of undergraduate math problems that require various forays into the land of algebraic manipulation.

Variables

4. Given that $3 + a + a = a + 10$, how many values could the letter a assume?

Danny:

3

Erika:

infinitely many

Fred:

none

The task in Problem 4 is the sort that could be found in many pre-algebra textbooks and involves the kind of thinking and skills that we may feel very comfortable assuming our college math students have command over. However, tied up with this seemingly simple task are some very complex ideas related to variable that can impact their work on undergraduate math, some of which your students may not have had ample opportunities to master but that you can learn to spot.

The first time students see variables they are being used as placeholders for some particular unchanging value (making "variable" an odd choice of name). For example, students are shown problems such as $2 + \square = 9$, where the goal is to place a value in the box that makes the statement true. Then the problem is restated as $2 + x = 9$. Many examples typically follow where the goal is to determine the specific unknown value for the variable that makes the statement true. And students also learn that just because $x = 7$ in one problem, that doesn't mean it will equal 7 in all problems. Although their struggles rarely bubble to the surface, many students struggle to reconcile this unchanging-yet-changing concept with the word "variable." As we saw with the use of the equal sign earlier, the way that variable is used in elementary school can follow students into their college years. This use of "variable," and a letter to represent it, may be what Danny is thinking about. Seeing that a appears three times in the problem, Danny concludes that there are three values that are possible for a.

Erika may be drawing on a use of variable that students typically first encounter at the very beginning of their study of algebra. Early examples students see may include equations such as $y = x + 5$. Here, x is no longer a *specific* unknown value but is instead a placeholder for an infinite number of values—a concept better in line with the word "variable." This transition from a specific value view of variable to one of a general number is very challenging for some students and perhaps Erika is still struggling to recognize the difference. For Fred, the issues related to variable interpretation may have been compounded along the way when he encountered problems such as $x + 2 = x + 6$ and learned that in some cases there are no values that make the statement true. He may be misapplying this idea to the given problem.

Sorting out these differences is nontrivial for students. Researchers have found elementary and middle school students struggle with these ideas [7, 8] and in a study of first year university students only 41% produced the correct answer for Problem 4—that only $a = 7$ is possible [12]. Similar results were found across a large collection of items designed to assess student interpretation of variables.

These different uses of variables are of course not the end of the story. In addition to using variables to represent the input values for functions, we also use variables to represent the functions themselves. The ability to appropriately interpret this function-related use of variable was also part of the same study [12] and student performance was poor. For example, only 46% of students were able to correctly answer the question, "If $x + 3 = y$, which values can y have?" Common wrong answers were "$x + 3$" and all natural numbers. Other researchers have also found that calculus students often struggle to use variables appropriately to define quantities in various kinds of applied calculus problems [13].

What You Can Do

Many students may be unaware of these different uses for variables; even those who move fluidly from one use to another may be unfamiliar with the characteristics of the different uses. Raising students' awareness of three types (specific value, general number, function) can be valuable and noting the challenges of moving from one view to the next can go a long way towards acknowledging the (often hidden) complexities we expect them to grasp. One way to do this is to point out and discuss what kind of variable is being used in particular problems you work on in class.

It can also be very valuable to let students in on the conventions mathematicians have for the use of particular letters. Students may be unaware that letters at the start of the alphabet (i.e., a, b, c, d) are commonly used to represent specific unknowns of the sort they solved for way back in elementary school. Other letters found in the middle of the alphabet are often used for functions of one sort of another (i.e., f, g, h, p, q, r, s, etc.) and letters at the end (i.e., x, y, z) are commonly used as general numbers. Of course these are not absolutes and these conventions may not hold in all situations (or may vary in other disciplines such as physics or engineering), but letting students know that the particular letter being used may be a clue to the kind of variable they are dealing with may help them refine their understanding of these ideas.

A good place to do this is when students are working with functions and functions evaluated at a specific value such as $f'(x)$ and $f'(a)$. You can check your students' understanding of variable by asking them: Let $f(x) = x^3 + 5$. State whether each of the following is a function or a number: $f(3)$, $f'(x)$, $f'(x)$ when $x = 4$, $f(a)$ when a is a constant.

Functions

5: Given that $f(x) = 2x + 3$, compute $f(x + a)$.

Grace:

$$f(x+a) = 2x + 3 + a$$

Henri:

$$f(x+a) = 2x + a + 3$$

Ida:

$$\begin{aligned} f(x+a) &= f(x) + f(a) \\ &= 2x+3 + 2a+3 \\ &= 2x+2a+6 \end{aligned}$$

Joaquin:

$$\begin{aligned} f(x+a) &= 2(x+a)+3 \\ &= 2x+2a+3 \end{aligned}$$

Students are typically first introduced to the idea of a function as represented by something in the form of $y = mx + b$. This form has useful aspects, but eventually they see it with new notation, namely $f(x) = mx + b$. This form conveys information about independent variables in a nicely compact way. Unfortunately, for many students adapting to and competently using this function notation presents many challenges.

We see evidence of some of these in the student work in Problem 5. Grace's work illustrates a difficulty that some students have understanding the role that the input to the function plays. Here Grace is interpreting "$x + a$" as meaning that a should be added to the output of the function instead of to the input. In other words, her train of thought, if made more explicit, might be $f(x+a) = f(x) + a$ so therefore $f(x+a) = f(x) + a = (2x+3) + a = 2x + 3 + a$.

Although technically equivalent, something different can be seen in the work of Henri. We imagine many teachers (including us!) have said something like, "Just replace the x with $x + a$." Henri did exactly that—writing in $x + a$ (instead of $(x+a)$). Although this looks exactly like a failure to distribute, Henri might actually be faithfully carrying out precisely the procedure he was told to use! The real problem, however, may be that he has an underdeveloped understanding of functional notation and does not understand that $f(x+a)$ represents applying the function to the (entire) quantity $x + a$.

In Ida's work is evidence of a different kind of difficulty with functions. You might catch yourself saying, "But wait, not all functions are linear!" Ida is doing something that is completely appropriate in many mathematical situations but not in ones involving this type of function or function notation: she distributes. When dealing with expressions like $5(x+a)$, it is OK and often useful to transform it into $5x + 5a$. There is remarkable similarity between $f(x+a)$ and $5(x+a)$ and Ida is likely viewing the use of f as no different from any other variable or value that can be distributed to the values inside parentheses. To us, the differences are obvious, but for many students the unfortunate dual use of parentheses as indicators for multiplication and for function input causes confusion.

Joaquin produces the correct answer. This work may indicate that he has a strong understanding of function and what is going on when a has been added to the input value. However, it is also possible (perhaps even likely) that Joaquin has mastered the procedures for generating these answers without developing a strong understanding of the difference between $h(x)$ and $h(x+a)$ or what it means to apply a function to a set of values.

Using this and other tasks, researchers investigated undergraduate students and found weaknesses in students' understanding of key ideas related to function [1]. They found that students generally have a good understanding of what it means to apply a function to a single value or to individual values in succession. What is more challenging, however, is conceptualizing a function as something that transforms *all* of the input values simultaneously. This way of thinking (and visualizing) is key to understanding how a dependent variable varies as an independent variable varies and this "covariational reasoning" is essential for the deep understanding of function needed in undergraduate mathematics (and is discussed in more detail below).

What You Can Do

Developing strong conceptual understanding and the necessary procedural skills to handle situations involving functions is a challenge for many students when they are first exposed to the topic and for some, those challenges persist in significant ways into their undergraduate studies. There are several things you can do to strengthen your students' understanding and skills. First, you can address the notation issue head-on by acknowledging the multiple meanings ascribed to parentheses. For example, you can mention this when you use "$y =$" or "$f(x) =$" notation—explaining how they could be used interchangeably but that the "$f(x) =$" version uses parentheses to indicate what input the function can accept. Then you can make the point that those parentheses are *not* conveying multiplication as would be the case with something like $5(x)$. Parentheses are just one of many mathematical symbols and words that have multiple, context-dependent meanings (several of which are discussed in detail in the next section of this chapter). In addition to discussing differences between the meaning of parentheses in $f(x)$ and $5(x)$, this would also be a fine time to mention other ambiguous uses of parentheses: intervals and points!

In addition to raising students' awareness about the notation itself, it can be useful to spend a bit of time on what it means for something to be the input for a function. You can make this point by using a variation on the notation where the input goes into a box. For example, $f(x) = 3x^2 + x$ can be written as $f \boxed{} = 3\boxed{}^2 + \boxed{}$. This representation can help students recognize that the function is operating on the *entire* input (i.e., contents of the box). This can help when we ask students to work with things like $f(x+h)$ or to compose functions (in which case the thing that goes into the box is itself a function). By using the box representation you can focus student attention on the idea of function input without the potential confusion from the multiplication meaning of parentheses.

Another way you can help students develop a richer conception of function is to focus on families of functions and how new functions can be seen as the result of a transformation on the entirety of another function. (This approach also has the side benefit of strengthening students' graphing skills.) For example, starting with the base function $y = x$ students can be asked to consider what happens to its graph when 5 is added to the output for each input value.

Illustrating this with a vertical translation up by 5 units can help convey that changes to a function impact all values. You can then generate the graph of $y = 3x + 5$ by first stretching $y = x$ so each output value is 3 times what it was and then translating the resulting graph up by 5 units.

Moving on to quadratics, you can generate $y = x^2$ by squaring the output values along the entire function $y = x$, noting that the point at $x = 0$ is invariant under the squaring operation. This view of a function that squares values can be applied to other linear functions such as $y = x + 5$ and armed with the idea that the value of the function at the point where $y = 0$ will not change, students might be more likely to figure out whether $y = (x + 5)^2$ is a parabola that has been translated left or right (compared with $y = x^2$).

Using this approach, students can learn to perform the necessary transformations to sketch graphs of functions as complicated as $y = 3(x + 5)^2 - 7$. And, armed with the skills needed to "complete the square," a wide range of functions can be rewritten in a form where these techniques of transformations on $y = x$ will support good graph sketching.

As you will see in future chapters (see, in particular, the Derivatives and Integration chapters), to learn and really understand many of the core ideas in the undergraduate curriculum students need a strong command of *covariation*. In the context of functions, this refers to the ways in which the output of a function varies as the input changes. As noted above, students get many opportunities to consider what the value of a function is for a specific input or for finite sets of inputs, but calculus is apt to be the first time learning and understanding are dependent on this continuously changing, across-time, all-points-at-once way of thinking. Helping students develop their abilities to think about and visualize situations involving covariation will help ensure that they get the most out of your lessons on derivatives and integrals. You will see some specific suggestions for how to do this in the Integrals chapter but in any context where you talk about a function you can ask students questions about how the function varies as various things happen to the input to help develop their abilities to think in these ways. Showing how a function changes with some computer-generated animation can also help students learn how to visualize covariation.

Lexical Ambiguity

English words frequently have different meanings in different contexts. For example, when someone asks you to "set the table" that likely conjures up images of forks and plates and not mathematical objects (or volleyballs!) That this *lexical ambiguity* exists is just something we learn to deal with, but being aware of (and acknowledging) this as your students are encountering new uses of familiar words can help their learning. Here we give just a small sampling of some of the lexically ambiguous words (and representations) that your college math students may be in the process of sorting out.

Average

6: You are taking a trip on a train. When you start the trip, you are 10 miles from home. After 3 hours, you are 130 miles from home. What was the average rate of change of distance with respect to time for the train during this portion of your trip?

Kate:

$$\frac{10+130}{2} = \frac{140}{2} = 70\text{mph}$$

Early in Calculus I students are introduced to the idea of "average rate of change" and given various problems like Problem 6 that often involve computing things like total distance traveled divided by total time it took to travel that distance. Here that would involve computing $\frac{130-10}{3} = 40$. These computations generate values such as miles per hour, gallons per minute, etc., and knowing what the units are of these values is key to understanding precisely what an average rate of change is.

Students, however, likely have much more experience with a different kind of average. On day one of your class when you were going over the syllabus you may well have said something like, “Your quiz average will be worth 20% of your final grade.” Every one of your students likely understood exactly what you meant: To figure out their quiz average, they should add together all of their quiz scores and divide by the number of quizzes. That’s their quiz average. We don’t normally say, “Your quiz average is 86.5 points per quiz” because the units seem obvious. Quiz scores (and most of the things they have encountered prior to calculus) involve a finite number of discrete values. The average rate of change examples involve continuous values most of the time.

When students encounter average rates of change for the first time, they are likely trying to make sense of the concept and calculations by applying their understanding of average as “add things up and divide by the number of things.” This can interfere with their learning of what average means in the context of rates of change. This can cause them to do things like what Kate does. She finds the arithmetic average of the speed of a car at the start and end of a trip ($(10 + 130)/2$) when asked about the average rate of change in position. Acknowledging to your students that their understanding of “average” is totally correct for some things but that there are other meanings of the term will go a long way in helping them. If you work a few examples of the discrete value type of average when talking about average rates of change you can raise their awareness of the differences and help them build a solid understanding of what it means to have an average of rates of change.

Inverses

7: Find the derivative of $y = \sin^{-1} x$.

Larry:

$$y = \sin^{-1} x = \frac{1}{\sin x}$$

$$y' = \frac{0 - \cos x}{\sin^2 x} = \frac{-\cos x}{\sin^2 x}$$

Larry encounters some difficulty while trying to carry out the computations needed to find the derivative of $y = \sin^{-1} x$. And the trouble begins with the first step. Before digging into Larry’s specific work, let’s consider just what that exponent of -1 means in particular contexts.

If you see -1 after a mathematical term, what kind of object might it be indicating? The additive inverse of 7? The multiplicative inverse of 2? The function inverse of sin? The answer of course is any of them. We often say “inverse” and assume that students will understand what kind of inverse we mean by the context. For experts, this becomes nearly automatic and the ambiguity rarely causes confusion. For students, however, they may be working hard to make use of their understanding of (multiplicative) inverse when we try to teach them about the inverse of the sine function and this can cause unnecessary confusion and frustration (and written work that includes things like that in Problem 7). These mathematical objects have both first and last names (e.g., multiplicative inverse, function inverse, additive inverse) but in casual conversation we sometimes drop the first name. Acknowledging this to students, consistently using both the first and last names when you speak (and helping them deepen their understanding of what it means to “inverse” something) can help them develop fluency with this aspect of mathematical language.

Absolute value bars and norms

This is an example of something that might seem as if it is lexically ambiguous to your students, but really isn’t. When students see absolute value bars in some arithmetic or algebra situation (for example, $|5 - 9|$) that might not cue the same kind of thinking as when they see those same bars used to indicate the norm of a vector. But it should. Read on.

A Bridge Too Far?

8: Prove that $\lim_{x \to 2}(-3x - 4) = -10$

Mindy:

$$|f(x) - L| < \varepsilon$$
$$|-3x - 4 - (-10)| < \varepsilon$$
$$\Rightarrow |-3x + 6| < \varepsilon$$
$$\Rightarrow 3x + 6 < \varepsilon$$
$$\Rightarrow x < \frac{\varepsilon - 6}{3}$$

In her efforts to start this delta-epsilon proof, Mindy has done many appropriate things. The initial epsilon expression is correct, the substitution was carried out appropriately and subtraction of a negative number was done accurately. But then something went wrong. Mindy states that $|-3x + 6| = 3x + 6$. The root of this issue may be Mindy's early experience with absolute value bars where she was likely exposed to many exercises that asked her to simplify expressions like $|-7|$ or $|-7 - 2|$. These experiences can lead students to believe that when they see absolute value bars, they should just drop the negative sign. In addition to generating incorrect answers, this view of the absolute value function also fails to do justice to key ideas about subtraction and distance that play important roles in various mathematical situations (for example, delta-epsilon proofs).

Early introductions to subtraction often make use of physical objects to illustrate the idea of "taking away." For example, back in elementary school you probably answered quite a few questions like, "Mindy has 7 cookies. She eats 2 for snack. How many cookies does Mindy have now?" These kinds of examples were used to illustrate basic subtraction computations and as such, they have many strengths. However, there is another model for subtraction that is more powerful and generalizable, especially for situations involving negative numbers: distance.

Instead of asking about cookies, we can ask, "You are 7 blocks from home and then you walk 2 blocks towards home. How far away from home are you now?" And then this can be made more abstract by utilizing a number line and asking questions such as, "How far apart are the numbers 7 and 2?" This model has two very nice features. First, it can accommodate negative numbers and second, we can exploit students' intuitive understanding of distances as positive to explain the absolute value function. Expressions like $|-9 - 3|$ can be interpreted as "How far apart are -9 and 3?" If distance is understood as a positive-only quantity then this can be understood by even young students. It also means that $|-9 - (-3)|$ can be interpreted as "How far apart are -9 and -3?" That gives meaning to the rule "subtracting a negative is like adding a positive." This approach also provides a basis for understanding why $|9 - 3| = |3 - 9|$: The distance between two numbers on the number line doesn't change if you move from 9 to 3 or from 3 to 9. In addition, $|-3|$ can be understood as the distance between zero and 3 ($|0 - 3|$)—or the reverse—instead of just the number without the negative sign (as Mindy seems to be thinking in Problem 8).

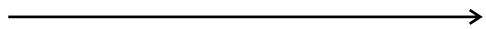
With this understanding in place, students are more likely to understand the seemingly mysterious expressions found in delta-epsilon proofs. The expression $|f(x) - L|$ is "The distance between $f(x)$ and L." And when you finish it off with "$< \epsilon$" it becomes a statement about that distance being less than ϵ. In a similar vein, $0 < |x - a| < \delta$ can be translated into "The distance between x and a is greater than zero and less than δ." Armed with this distance interpretation of subtraction and the absolute value function, there is hope that students won't make the errors that we saw in Mindy's work, and that they can understand the powerful ideas conveyed in a delta-epsilon proof.

In later classes, when they are dealing with vectors and encounter expressions like $|x|$, the notation is consistent because they are being asked to find the length of the vector x (or the distance of its tip from its tail at the origin).

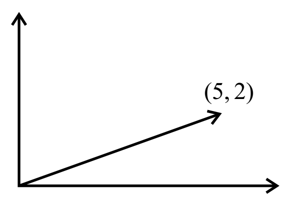
In this and all cases of lexical ambiguity, it is best for you to acknowledge the ambiguity instead of pretending that it doesn't exist or assuming that students can use the contextual clues to decipher the appropriate meaning. For absolute value and norm situations in particular, a brief foray into the distance interpretation of subtraction when more sophisticated uses of it come up in class can improve students' understanding of the new, more complex ideas you are helping them learn.

Rays and Vectors

Quick question: What does this figure represent:



A vector? A ray? Part of a numberline? The start of $xy-$ coordinate axes? All of these? A line with an arrow on one or both ends is a representation you have been working with for years. You probably first saw it early in elementary school in the form of a numberline. Then when you started learning geometry, you saw these arrow/lines used to represent the sides of an angle. Somewhere along the way you saw one perpendicular to another to represent a plane. Eventually you encountered them being used to represent displacement, likely in two dimensions and now they were called vectors. This can be quite confusing for students as they try to use their understanding of (for example) rays from geometry to make sense of this new vector idea you are presenting to them. Vectors and rays are known to be quite challenging for students to interpret and use [5, 10, 14]. Consider for a moment the image of a portion of a coordinate plane and a vector going from the origin to $(5, 2)$:



To appropriately work with this representation, students need to understand the similarities and differences between what the x-axis and y-axis lines-with-arrows mean and what the slanted line-with-arrow means. Again, acknowledging that this *representational ambiguity* exists can help them build a new, contextually-appropriate, understanding of the term instead of trying to reconcile what they know with the new ideas you're presenting to them.

And while we have you focused on representations involving axes, we call your attention to the fact that "x-axis" may be lexically ambiguous to some of your students. Although we in the mathematics world treat "x-axis" as synonymous with "horizontal axis," our colleagues in physics do not. Physicists routinely create position versus time graphs with x as the variable for position. This means that, to them, the vertical axis is the x-axis. So, when we say "x-axis" some students may think we are referring to (what we call) the "y-axis." The potential for confusion extends to situations where we discuss $\frac{dx}{dt}$ in the context of parametric equations as the rate of change in the "x" direction. But for students in physics, that expression refers to the derivative of position (in whatever direction is being modeled in the problem) with respect to time. And while we are on this subject, in the context of spherical coordinates we play a cruel trick on our students. In our classes we refer to the polar angle as phi and the azimuthal angle as theta. But when these quantities are discussed in their physics classes, the theta and phi will be swapped.

Definite vs. Indefinite

$$\int x^2 + 1\, dx \qquad\qquad \int_0^2 x^2 + 1\, dx$$

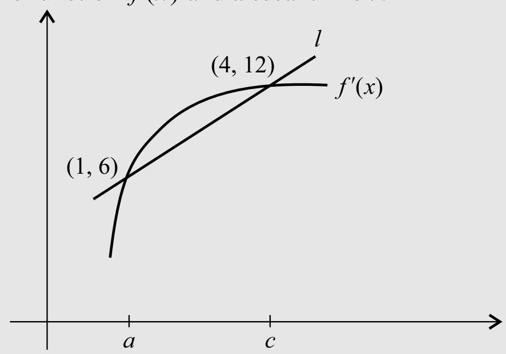
Question: What's the difference between these two expressions?

When asked to describe what an indefinite integral is, research findings demonstrate that some calculus students give answers that suggest that they are using their understanding of the word "indefinite" as they try to make sense of the ideas [4]. For example, one student said that when you compute an indefinite integral, "you really won't get a really precise answer ... you get an unfinished thing." Another student said that when you do an indefinite integral computation, "...you're not sure, you have to, like, find out." Similarly, when discussing definite integrals, students said things consistent with the English language definition of "definite." For example, one student said, "...like, you *know*, like, it's in-between, like, these two, like it's definite." Another student, when asked what a definite integral was stated, "I guess you're certain where it's ending and starting so that instead of just a general thing..."

Despite the similarities in how we talk about them (and write them), definite and indefinite integrals are quite different. An indefinite integral is really just an anti-derivative, while a definite integral is, by definition, a limit of Riemann sums. Of course the lexical ambiguity of "definite" and "indefinite" is only a small part of what students find difficult in these two concepts—but it certainly doesn't help! See the chapter on integrals for more discussion of how students think about antiderivatives and definite integrals.

Graphs and Visualizing Dynamic Change

9: Below is the graph of the function $f(x)$ and a secant line l.



a. Find the slope of the secant line l.

b. If point a moves along the x-axis toward point c (and the secant line l moves accordingly), does the slope of l increase, decrease or stay the same? NOTE: a refers to the x-coordinate of the left-hand point of intersection of the graph of $f(x)$ and line l (adapted from [9]):

Nicholas:

a. $\frac{12-6}{4-1} = \frac{6}{3} = 2$

b. increasing. Slope: +2, uphill

Odette:

$\frac{12-6}{4-1}$ ← const. so it stays the same.

Peter:

a $\frac{12-6}{4-1} = 2$

b. Try pt.: (2, 9)

$\frac{12-9}{4-2} = \frac{3}{2}$ Decreases!

Rose:

a. $\frac{12-6}{4-1} = 2$

b. Decreases - getting flatter and flatter

Both parts of Problem 9 tap into students' understanding of slope, and for this reason it might seem likely that most will perform similarly on the two parts. It turns out, however, that although both parts require an understanding of slope to get the correct answer, there is a lot more going on in (b) than you might think. To get the correct answer to (a) you need to know that slope is a measure of the ratio of change in y-values to change in x-values, then you need to use the given points to do the calculation. This comes fairly easily to most college students, with upwards of 80% getting such problems correct [9]. And the errors typically come from misusing the formula.

When examining a problem like (b), we think the notation (which we have reproduced faithfully from the original publication) is a little ambiguous. Somehow the left intersection point is supposed to be both fixed (at the point $(1, 6)$) and variable (when named $(a, f(a))$). Odette's response suggests she struggled with this notational issue. If students interpret the problem as intended, they need a very different sort of reasoning than what was required for part (a). When you were figuring out the answer to (b), did you mentally examine the slope of l as it slid along $f(x)$ in an animated movie in your mind? Did that movie automatically start playing the instant you read the question? Chances are that it did and that you were able to envision what happens as $f(x)$ and l initially intersect at $(1, 6)$ and then intersect at values represented by $(a, f(a))$ as a increases. In contrast, although Peter's answer is correct, you may have found the approach a bit odd, and unnecessary, preferring instead what Rose writes. Peter looks at one pair of points, calculates the slope and then compares that value to the calculation of slope from two other points.

To succeed in K–12 mathematics, looking at things point-by-point, or in a pointwise fashion, is all that students need. When students arrive in calculus, however, the questions and mathematical models are now all about change and those pointwise skills are no longer sufficient. We talk about rates of change, how rates and slopes change over time or as some other value changes, implicitly using students' abilities to see things in all their dynamic, animated glory. Important note: This across-time view into mathematics has been unnecessary for success up until students reach calculus.

And it turns out that running those movies in our minds is not a practice or skill that develops automatically just because you can calculate the slope values for particular lines. Although over 80% can correctly answer the part of problem that does not utilize this kind of animation/visualization, only about 45% of students get both parts (a) and (b) correct [9]. Looking at the data differently, very few (less than 10% of) students get (b) correct but (a) incorrect, whereas quite a few more (in the range of 34–40%) get (b) wrong but are correct in their computations for (a).

So, just because a student has mastered the pointwise calculations needed in courses prior to calculus does not mean that student will automatically be able to do the visualization needed to interpret the various representations we use to teach and test calculus students. Findings similar to those with this task have been reported with other tasks in the study [9] and in other studies [1].

Students who haven't yet learned to run the animations in their heads will face significant trouble when asked to do things that include sketching graphs of derivatives by reading off slopes of lines tangent to the original function [11]. In general, these difficulties fall into the larger category of challenges that students face with *covariational reasoning* where one must figure out how one quantity changes as another is varied. See the section on Function earlier in this chapter for more about the key role that this kind of reasoning plays in learning college mathematics.

What You Can Do

You can help your students develop strong visualization skills for covariational situations by providing bridges between what is evident in a graph representation (what they see on the page or whiteboard) and what **you** see in your mind/imagination. You can do this with dynamic graphing software (Google will help you find many open-source Geogebra applets) or by drawing in a set of still images from the movie that is playing in your mind. You can be on the lookout for situations that implicitly or explicitly call on these kinds of visualization skills and use those as opportunities to strengthen your students' abilities to see in diagrams what it is that you see.

A Good Place to Start

When you see what seem to be "careless mistakes," take a moment to consider why the student's flawed approach may have *made sense* to them. Until you probe their thinking, you are better off viewing such mistakes as symptoms of an underdeveloped understanding, not merely momentary lapses.

Bibliography

[1] M. Carlson. A Cross-Sectional Investigation of the Development of the Function Concept. In Ed Dubinsky, Alan H Schoenfeld, and J Kaput, editors, *Research in Collegiate Mathematics Education*, volume 1, pages 114–162. American Mathematical Society, Washington, DC, 1998.

[2] L. Chalouh and N. Herscovics. Teaching algebraic expressions in meaningful ways. In A. E. Coxford, editor, *The Ideas of Algebra, K–12*, chapter 1988 Yearb, pages 33–42. National Council of Teachers of Mathematics, Reston, VA, 1988.

[3] Eugenio Filloy and Teresa Rojano. Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics*, 9(2):19–25, 2013.

[4] William Lloyd Hall Jr. *Language and Area: Influences on Student Understanding of Integration*. PhD thesis, The University of Maine, 2010.

[5] Karen F Hollebrands. High school students understandings of geometric transformations in the context of a technological environment. *The Journal of Mathematical Behavior*, 22(1):55–72, 2003.

[6] C. Kieran. Concepts associated with the equality symbol. *Educational Studies in Mathematics*, 12(3):317–326, 1981.

[7] A. C. Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens. Middle School Students' Understanding of Core Algebraic Concepts: Equivalence & Variable. *Zentralblatt für Didaktik der Mathematik*, 37(1):68–76, 2005.

[8] D. Kuchermann. Children's understanding of numerical variables. *Mathematics in School*, 7(4):23–26, 1978.

[9] S. Monk. Students' understanding of functions in calculus courses. *Humanistic Mathematics Network journal*, 27:21–27, 1994.

[10] Sarah Rizzo. *College students' understanding of geometric transformations*. Unpublished Master's thesis, The University of Maine, Orono, ME, 2013.

[11] John Stahley. *Students' qualitative understanding of the relationship between the graph of a function and the graphs of its derivatives*. (Unpublished Master's thesis) PhD thesis, The University of Maine, Orono, ME, 2011.

[12] S. Trigueros, M. & Ursini. First-year undergraduates' difficulties in working with different uses of variable. *CBMS issues in mathematics education*, 8:1–26, 2003.

[13] Paul White and Michael Mitchelmore. Conceptual Knowledge in Introductory Calculus. *Journal for Research in Mathematics Education*, 27(1):79–95, 1996.

[14] A. Yanik, H. B., & Flores. Understanding rigid geometric transformations: Jeff's learning path for translation. *The Journal of Mathematical Behavior*, 28(1):41–57, 2009.

1

Limits

Anatomy of a Chapter

Before we dive into all the fascinating ways students think about limits, we want to give an overview of the structure of each chapter (if you read the Preface, this will serve as a review). Following a short introduction to the chapter topic to showcase student thinking, chapter sections are organized around examples of student work. There are sets of problems and student answers, each of which has the same format: a mathematical task, sample student responses to that task, and analysis/discussion of those responses. The content of the chapters is based on **findings from research**. Many of the tasks that appear in the problem and student answers sections are taken directly from research papers about student thinking and learning. In these cases, the sample student work may be reproduced verbatim. For some tasks discussed in the literature, we generated sample student work based on research findings, informed by our own experiences working with students. Other tasks and student responses were created to illustrate ideas from research but were not modeled directly on any specific research study. Citations indicate whether we took a problem or student work directly from the research or adapted it for our purposes.

As you read, you are **strongly encouraged** to try out the problem yourself and to anticipate how students might answer prior to looking at the sample student solutions. This will help you get the most out of the discussion of student thinking and is the best we can do to create an active, engaged learning experience for you via text!

Many researchers suggested ways to address known student difficulties and to help students develop strong understanding of the ideas. Those are discussed in the second part of each chapter in **What You Can Do**. While most of the suggestions are supported by citations to research, occasionally we have drawn from our own experiences and from those of our reviewers.

Each chapter concludes with **A Good Place to Start**, a single suggestion, informed by the research, that you could use when you teach that chapter's material. These suggestions are entirely our own, based on our reading of the available research.

Now back to our regularly scheduled chapter on . . .

Limits

Limits form the theoretical basis for all of calculus. From the perspective of the mathematician, they are absolutely essential. From the perspective of some students, they're a waste of time. One such freshman timidly approached a professor who had diligently spent several class periods getting students to calculate derivatives using limits. "Professor, sorry if this is out of place, but in high school they showed us a much easier way of doing these derivative things." For her, not only was the cart in front of the horse—the horse was completely unnecessary!

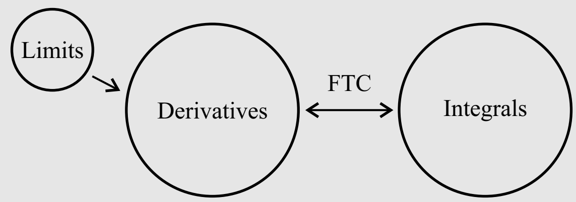
Although limits are vital to the theory of calculus, as with some other topics, we are really good at giving students answers to questions they do not yet have. From our knowledge of calculus (and beyond), we know many ways that limits are vital tools enabling us to solve problems. Students do not yet have that information and they may fail to appreciate the utility of limits, contributing to the views described above.

These challenges are compounded by the fact that limits also extremely challenging for students to understand. Most students struggle to gain the intuitive understanding that Newton and Leibniz would have used—and are left completely baffled by the modern ε-δ definition [8]. They typically exhibit an amazing ability to seemingly understand limits very well in one context and have no clue in another. Fortunately, many researchers have inquired into student understanding of limits. Their work ranges from a seven-step decomposition of the learning of limits [3] to studies of how to improve students' journeys through this difficult topic [2]. Here's what that research says about what your students might be thinking:

How does it all fit together?

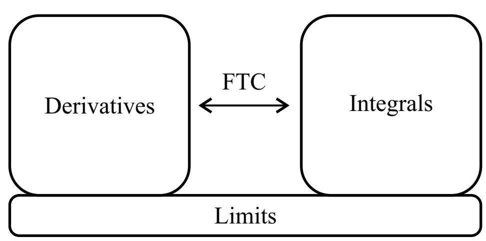
1. Draw a concept diagram that shows how the main concepts of Calculus I are connected.

Alex:



A few years ago, one of the authors (Dave) concluded a Calculus I course by asking students Problem 1. Alex's response is typical of many students. Although one main point of the semester shines through (that the Fundamental Theorem of Calculus connects derivatives and integrals) his positioning of limits suggests that they are of minor importance and only connect to derivatives. It's easy to see where such a conception comes from, since it exactly mirrors many typical textbooks. Limits are introduced in a functional setting (i.e., $\lim_{x\to a} f(x)$) and used as a tool to define derivatives. But after you figure out all of the derivative rules, limits fade into the background. The appearance they make in Riemann Sums (taking the limit as the number of boxes goes to infinity—or more precisely, the width of the largest box goes to zero), feels like a mere cameo role. Hence for Alex, and many others, the concept of the limit isn't connect to integrals in any substantial way.

In contrast, a mathematician might view the same topics as connected differently:



In this model, limits form the foundation for calculus, with both derivatives and integrals resting on limits as their logical precursors. In calculus teachers' dreams, students leave the course sequence understanding limits in a deep way that connects all of the important topics: derivatives, integrals, sequences and series. Even if Alex moves past all the naive unproductive ways of thinking discussed in the rest of this chapter, even if he understands the ε-δ definition and can produce beautiful proofs, he has work to do before limits take their proper place in his understanding of calculus.

The way limits are introduced to students sometimes fails to do justice to why the idea is so important and useful. As a result, students may not have a sense of the rationale for bothering with a seemingly esoteric idea.

One remedy for this is to give students the opportunity to feel the **need** for limit as a tool and idea. More details about this approach are given at the end of the chapter, but the essence is to start with the idea of average rate of change and then pose the challenge of how a rate of change can be calculated at a particular instant, thus motivating the need for an approximation to zero. This gives students the opportunity to understand limit as an idea that helps us solve problems.

Getting students to this big picture understanding of a concept as fundamental as the limit vitally impacts their understanding of calculus. Without this grounding concept to tie everything together, a year's worth of calculus might quickly dissolve into a bunch of disconnected facts. The limit connects all those topics together—it's the bow that ties them up into the pretty package that is, or should be, calculus.

All Functions Are Continuous

2. Find the limit and explain your answer: $\lim\limits_{x \to 2} 3x^2$

Bonnie: 12. Plug in 2, get 12.

Is Bonnie's thinking on Problem 2 correct? Hard to tell. It's possible that she is really thinking "You could make $3x^2$ arbitrarily close to 12 by ensuring that x is close but not equal to 2." More likely, she's following a simple algorithm for limit problems: plug in the number—or possibly the more advanced algorithm: plug in the number, if that doesn't work, do some algebra and try again.

The underlying problematic thinking here, from a mathematical perspective, is that all functions are continuous. While this idea is ridiculous to any mathematician (and probably to most students, when posed in such a way), it's an understandable mistake given that most functions students have seen are, in fact, continuous on their domains. This way of thinking is reinforced by the algebraic manipulations needed to do many basic limit problems. That the original function wasn't continuous is seen as an anomaly—an inconvenience that a little algebraic work can overcome [8].

Examples of this common problematic way of thinking come up again and again in calculus. When given that $f(2) = 1$ and $f'(2) = -2$, and then asked to find $\lim\limits_{x \to 2} f(x)$, many students get the answer correct for the wrong reasons: "If 2 is substituted into $f(x)$, the table shows that the answer is 1" or "The value of the function exists at $x = 2$, therefore the limit must be equal to that value of the function" [1, p. 492]. Very few students made the logically necessary argument that differentiability implies continuity, forcing the limit to equal the value at the point.

Limit = (Unreachable) Bound

3. Students talking about limits …

Charley: 1, 1, 1, 1, … *doesn't converge.*

Danielle: *A function can't touch its asymptote.*

The first few limits that students see leave many of them with the incorrect idea that the properties of these easy examples apply to all limits. Looking at limits like $\lim_{x \to \infty} 1/x$ and sequences like $\frac{1}{n}$ or $\frac{1}{n^2}$, students notice that the limit is never actually reached. This leads students to think that this is a defining property of limits, leading to Charley's and Danielle's comments above. (Interestingly, historians tell us that in making the "mistake" that limits are "unreachable," students are actually agreeing with Newton [6]!) Examples that don't satisfy those properties (e.g., $\lim_{x \to \infty} \frac{\sin x}{x}$, the sequence $\{0, \frac{1}{2}, 0, \frac{1}{4}, 0, \frac{1}{8}, \ldots\}$, or the constant sequence shown above—all of which are equal to their limiting values infinitely many times) leave some students befuddled ("I thought it had to 'go to' the limit, not jump around!") The idea that a limit is a bound can be more absurd; for instance, the sequence $.9, .99, .999, \ldots$ has 1 as a bound—but π and 10^{1000} are also bounds! (Some students' informal idea of "bound" matches more closely with the Least Upper Bound.)

What is somewhat surprising is that research suggests that these under-developed ways of thinking might be unavoidable. Even after teaching specifically to avoid these ideas, students still demonstrated these ways of thinking [4].

This leads some researchers to label these problems as naive conceptions—incorrect ideas that serve as rest areas on the road to a more complete understanding of limits.

"All Limits are Monotone"

4. Does this sequence converge or diverge:

$$0,\ 1,\ 0,\ 1/2,\ 0,\ 1/3,\ 0,\ 1/4, \ldots ?$$

Earl: *This doesn't converge, it jumps around.*

Earl's response to Problem 4 could be a direct consequence of having seen the sequence $(-1)^n$ earlier. He was told that $(-1)^n$ diverges, and now he thinks that all sequences that aren't monotone—either strictly increasing or strictly decreasing—diverge. Given that the language we use in describing limits is mostly the language of motion (e.g., a_n "goes to" L; $f(x)$ "approaches" 3) this sort of conceptual error shouldn't be surprising. When asked to describe in their own words what a limit is, some student responses reflect this way of thinking: "The limit of a sequence is the point that all terms in the sequence are leading toward" [4, p. 297]. Such students are likely to view non-monotone limits (e.g., this sequence or $\lim_{x\to 0} x \sin(1/x)$) as anomalies—not counterexamples to a false statement but exceptions to a rule.

A Few Points Are Good Enough

5. Calculate $\displaystyle\lim_{x\to 3} \frac{-2x^2 + 18x - 36}{x - 3}$.

Frances:

f(3.001) = 5.998

answer: 6

Given the nature of most calculus problems, Frances' answer to Problem 5 is most likely correct, but her reasoning leaves something to be desired. In theory, using a single point to determine a limit is ridiculous—in practice, it's a very effective strategy. With mathematical maturity comes the judgement necessary to understand the applicability and limitations of such a method. Students take years to develop this maturity and many, instead of doing any deep thinking about the problem, fall back on an easy procedure: plug in one (or several) points. This is another example of a naive conception, one that has an element of truth, but fails the test of generality. Using examples where the eventual limit is irrational can help, as can showing examples where a calculator's precision gives a wild answer. However, this procedural view of limits has proven very difficult to overcome [9].

What Are Those Squiggly Greek Letters?

6. Use the delta-epsilon definition to prove that

$$\lim_{x\to 2} x^2 - x + 1 = 3$$

Gaston: *Choose* $\epsilon > 0$ *and* $\delta = \epsilon/(x+1)$. *Then* $|x-2| < \delta$ *and* $|f(x) - 3| = x^2 - x + 1 - 3 = x^2 - x - 2 = (x-2)(x+1) < \delta(x+1) = \epsilon$.

Ah, deltas and epsilons, the toughest part of many calculus classes (at least by modern standards). As anyone who has taken an analysis course knows, epsilons and deltas are as important as they are challenging. For calculus students, the hurdles are almost insurmountable—an implication, absolute values (see Chapter Zero for more), and not just quantifiers (which they haven't studied yet) but **nested** quantifiers (more in the Proofs chapter)—and part of it is literally Greek to them!

At the calculus level (as opposed to the analysis level), students are frequently overwhelmed by epsilons and deltas—as demonstrated by Gaston's confused response. It's almost as if he has little puzzle pieces of correct thinking but doesn't have the tools to put them together in a logical way. If we look closer at his work, we see that the structure of the proof is actually quite good, starting with an epsilon and then choosing a delta before proving the correct implication. What keeps his work from being a correct proof is that his choice of δ depends on x. What Gaston writes actually proves (modulo some absolute value bars and a few wording changes) a different statement:

$$\forall \epsilon > 0 \;\; \forall x \;\; \exists \delta > 0 \;\; (|x-2| < \delta \Rightarrow |f(x) - 3| < \epsilon),$$

instead of the subtly different correct version:

$$\forall \epsilon > 0 \;\; \exists \delta > 0 \;\; (0 < |x-2| < \delta \Rightarrow |f(x) - 3| < \epsilon).$$

Gaston's mistake makes it clear that he is missing some critical conceptual ideas of limits. Making δ depend on x shows a lack of understanding of the underlying logical structure of the definition of the limit and hints that he might simply be blindly following a template. Also, if he had connected the steps of the proof to a graphical understanding of the situation, it would be clear that δ should depend on ϵ and not on x.

Not that these sorts of mistakes are surprising given the complexity of the ε-δ definition. Let's take a closer look at the hurdles faced by students grappling with this difficult concept.

On a superficial level, students are intimidated by the Greek letters that they may not have seen before. On top of that, absolute values (that they learned once but might not be fluent with—see discussion of this in Chapter Zero) are intermingled with an if-then statement. Standing alone, a simple implication would be within students' reach (think of the test for divergence of series), but here it just compounds their confusions. Finally, it's not one quantifier but *nested quantifiers* that start the definition. It's no wonder students struggle! Even for students who succeed in Calculus, each one of these ideas presents significant challenges that are often apparent when they are in an Introduction to Proof course. A more complete discussion of the complex nature of these issues can be found in the Proofs chapter.

While calculus students can be taught (with some difficulty) how to do easy ε-δ proofs, even the most advanced will struggle with putting all of these challenging elements together in a way that makes sense to them. Students typically produce garbled responses such as Gaston's, which show some hints of correct thinking but little attention to the logical structure of the proof. And even those students who produce correct proofs are typically on shaky conceptual ground.

What You Can Do

While some of the challenges students face in learning limits have proven difficult to overcome, student thinking about the ideas has been the focus of quite a bit of research. A variety of ideas and instructional materials for addressing the challenges can be found in the publications cited in this chapter. Here, we provide suggestions based on one of the most significant underlying causes for such challenges: the language we use [4, 5].

"As x gets closer and closer to a ..." certainly evokes a monotone, unreachable sense—and even the colloquial use of the word "limit" gives a sense of bound (e.g., speed limit, limited supply, etc.). Also, these phrases often deal with motion, while the rigorous definition is paradoxically static. This tension between the language we use, which influences students' intuitive understanding of limits, and the formal definition can produce the confusion seen above [8]. Add to that the less–than–compelling reasons students have for the whole limit enterprise and it can be an especially challenging section of the course for both students and their instructors.

Fortunately, recent work indicates that changing the language we use can lead to better student understanding [7, 2]. Phrasing limits in terms of approximation (e.g., "precision," "error," and "tolerance") leads students to have a more usable (and more sophisticated) concept of limits. With this language, "a_n goes to L" becomes "a_n approximates L to within any arbitrary error."

If you do want to dive headfirst into the deep pool of ϵs and δs, be aware that students won't necessarily see the same moving pictures that might be in your head. Like the sliding secant problem you read about in Chapter Zero and will hear more about in the Derivative chapter, students' *covariational reasoning* might not be developed enough to visualize how δ might change continuously as ϵ changes.

Even if they do have the ability to visualize these changes, the definition is likely to overwhelm them with its complexity. One way to build intuition about the definition is to use analogies to situations that have the same structure. One researcher suggested the following as a way to build an intuitive understanding of limits which supported the formal definition.

> Suppose we run a bolt manufacturing company. We have lots and lots of contracts, with lots and lots of different companies. As you might expect, everybody's needs are a little different. Bolts that we provide for home construction have to be of good quality, whereas bolts that we provide for NASA to be used on the space shuttle have to be of exceptional quality. For the sake of simplicity, let's look at only one variable that goes into the quality of our bolts: length. Bolts for home construction that are supposed to be four inches long can be a little more or a little less than four inches. But bolts for the space shuttle that are supposed to be four inches long have to be within a much smaller target range in order to be acceptable. The length of the bolt depends directly on how much raw material we put into the bolt making machine. How do we create bolts that we know will be of a length that falls within our target range? If a new customer requires an even smaller tolerance, can you take on the job? —Adapted from [2].

This situation mirrors most of the conceptual difficulties of the limit definition, in a context that may feel more familiar to the students than a purely mathematical one. The implication (as well as the input-output sense of a function) is mirrored by the production process. Both the input and the output have possible errors, and one impacts the other. Finally, the nested quantifiers are mirrored by the possibility of having to satisfy any request. In other words, your output ($f(x)$) can be forced to be within a particular tolerance (measured by ϵ) of the desired value (L) by making your input (x) within a small enough input tolerance (measured by δ). By explicitly drawing the connections between situations like this and the definition of the limit, we can leverage students' informal understanding of this situation to help them understand limits. In addition to attending to the language-related issues discussed above, incorporating analogies such as the factory one can be very valuable for students [2].

Viewing limits through the lens of approximation, we lay better groundwork for the myriad uses of limits in the calculus sequence. For derivatives, the slope at a single point is being approximated by slopes of secant lines, with the error being controlled by how close the two points are. For Riemann sums, the area under the curve is being approximated by sums of areas of rectangles—with the error controlled by the width of the widest one. In each case (and in others), students can see that one quantity is being approximated to within an arbitrary error by making an input sufficiently precise. One way to impress this on students is to introduce (and return to) the concept of the limit as the following open sentence:

You can make ______ arbitrarily close to ______ by taking ______ sufficiently close to ______.

Then each new example of a limit is seen as a different way to fill in the blanks of this sentence. For instance, the derivative might be viewed as

You can make <u>the slope of the secant line</u> arbitrarily close to <u>the slope of the tangent line</u> by taking <u>the parameter h</u> sufficiently close to <u>zero</u>.

A close alternative, more philosophical, approach to this fill-in-the-blank strategy is to define them more vaguely:

A limit is the end result of an infinite process.

As with the various ways of filling in the blanks, this approach applies easily to all the main uses of limits in a first-year calculus sequence. For integrals, the infinite process involves approximating an area using an increasing number of boxes, and the end result is the area under the curve. For derivatives, the infinite process involves looking at average changes over intervals that are smaller and smaller, and the end result is the instantaneous change.

Both of these softer ways of introducing limits (fill-in-the-blank or "end result of an infinite process") avoid many of the conceptual difficulties shown by the student work in this chapter, and both also situate limits as a foundational concept for all of calculus (not just derivatives). Both essentially rely on the language of approximation, using the knowledge students already have instead of fighting against it. And even if you don't teach the ε-δ definition, by bringing their intuitive understanding of limits closer to the rigorous definition, we have better prepared them for the rigors of later math classes.

One last thing you can do to help students understand limits is to better motivate them to **want to** understand limits (as described in the opening section of this chapter). You might tackle average rate of change first (see Chapter Zero for insights into what makes this difficult for students), then ask how one would get an approximation for a particular instantaneous rate of change. The task of making better and better approximations (using smaller and smaller intervals around the instance in question) leads to the idea that the "best" approximation would come if the change occurred over an interval of width zero. However, this creates the dreaded "division by zero" problem. But lucky for us, we've got a solution for that—and presto—your students have a need for limits! Repeat this approach when introducing definite integrals ("We wish we could have rectangles of width zero, but we can't...") and students have opportunities to view limits as valuable ideas and tools for solving problems. This ordering of topics is backwards from many popular textbooks, but matches many newer ones. By creating the intellectual need to learn about limits, students are more motivated to dive into the definition and start calculating and proving things.

A Good Place to Start

After giving students opportunities to see the need for limits, shift your language away from metaphors of motion toward the notions of approximation that more accurately reflect the mathematical concept of the limit. Practice looking at $\lim_{x\to 3} f(x) = 5$ and saying "the function f becomes approximately 5 when x is approximately 3."

Bibliography

[1] J. Bezuidenhout. Limits and continuity: some conceptions of first-year students. *International Journal of Mathematical Education in Science and Technology*, 32(4):487–500, 2001.

[2] T. Boester. Student understanding of epsilons and deltas: a metaphorical approach. PhD thesis, in preparation.

[3] J. Cottrill, E. Dubinsky, D. Nichols, K. Schwingendorf, K. Thomas, and D. Vidakovic. Understanding the limit concept: Beginning with a coordinated process schema. *Journal of Mathematical Behavior*, 15(2):167–192, 1996.

[4] R.B. Davis and S. Vinner. The notion of limit: some seemingly unavoidable misconception stages. *Journal of Mathematical Behavior*, 5(3):281–303, 1986.

[5] J. Ferrini-Mundy and K. Graham. Research in calculus learning: understanding of limits, derivatives, and integrals. In J. Kaput and E. Dubinsky, editors, *Research Issues in Undergraduate Math Learning: Preliminary Analyses and Results (MAA Notes No. 33)*, pages 31–45. Mathematical Association of America, Washington, D.C., 1994.

[6] J.V. Grabiner. Who Gave You the Epsilon? Cauchy and the Origins of Rigorous Calculus. *The American Mathematical Monthly*, 90(3):185–194, 1983.

[7] M. Oehrtman. *Collapsing Dimensions, Physical Limitation, and other Student Metaphors for Limit Concepts: An Instrumentalist Investigation into Calculus Students' Spontaneous Reasoning*. PhD thesis, The University of Texas at Austin, 2002.

[8] D. Tall and S. Vinner. Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2):151–169, 1981.

[9] S.R. Williams. Models of limit held by college calculus students. *Journal for Research in Mathematics Education*, 22(3):219–236, 1991.

2

Derivatives

Unless you're teaching from Apostol [1], derivatives come early in first semester Calculus. And while most students eventually become quite good at *computing* derivatives, many (even those getting As) still aren't sure why they're doing what they're doing. While developing a strong understanding of the derivative is especially challenging, even the road to computational fluency is paved with numerous speed bumps.

When it comes to understanding the concept of the derivative, researchers have decomposed students' learning into three separate stages [29]. First, students need a basic understanding of slopes and how to compute them in order to grasp fully the connection between a secant line and the difference quotient that represents its slope. (Of course, more fundamental ideas about functions and change are packed into these concepts—see Chapter Zero for more.) Second comes understanding the limiting process (typically, $h \to 0$ or $x \to a$) as the secant lines approximate the tangent line. In this stage, students need to grasp the idea—arguably the central idea of calculus—that one quantity (the slope of the tangent) can be approximated arbitrarily well by another quantity (the slope of a secant). This gives the derivative at a single point. The final stage involves carrying out the first two stages at *all* points *simultaneously*, resulting in the derivative being a function itself and not just an isolated value. (While calculus usually stops here, functional analysis takes the abstraction another level, thinking of the derivative as an operator on a space of functions.) Permeating these stages is the concept of covariation—how one quantity changes as another changes.

Think how complex the ideas are in each of these stages. It's no wonder that some students give up trying to understand what's going on and focus instead on mastering procedures!

Student difficulties with derivatives range from deep conceptual problems with understanding the nature of real numbers and functions to more procedural difficulties applying the various derivative rules. Here we lay out the most common ways of thinking students demonstrate while learning, or trying to learn, about derivatives.

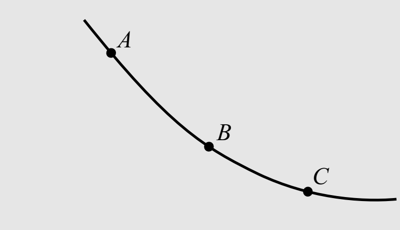
Still Shots but No Animation

1. For the function f shown to the right, put these three quantities in order from least to greatest:

I - The average rate of change of f between points A and C.

II - The instantaneous rate of change of f at point C.

III - The average rate of change of f between points B and C.



Arthur:

$\mathrm{II} < 0$ and $\mathrm{I} > \mathrm{III} > 0$

so $\mathrm{II} < \mathrm{III} < \mathrm{I}$

The first few weeks of Calculus typically center around questions of average rates of change (i.e., slopes of secant lines) and instantaneous rates of change (which we eventually call the derivative at a point). Problem 1 deals with both concepts, in an abstract setting where no computation is required. In this case, all three quantities in question are negative, with the segment $\overline{AC}$ the steepest, $\overline{BC}$ the next steepest, and the tangent line at C the least steep. Since all three slopes are less than zero, a correct ordering would be $I < III < II$.

So what is Arthur thinking when he gets the order exactly reversed? He added additional information, showing that he thinks II is negative and the others are positive; otherwise we might guess that he had ordered them by some intuitive sense of "steepness." Not taking into account the fact that all three are negative, he might not see that the steepest (with the largest **absolute value** of the slope) is also the smallest (farthest left on the number line). Such errors are fairly common as students struggle to coordinate the concept of steepness with the numerical values obtained by finding slopes.

But with the additional information Arthur provides, it's clear he's thinking something else. He has the correct sign for the instantaneous rate of change at C, so it's possible he's picturing the tangent line at C correctly and seeing that it has a slightly negative slope. What thinking might lead him to conclude that the two average values are positive?

One possible answer goes back to the lexical ambiguity of the word "average" discussed in Chapter Zero. By the time students reach you, they've taken hundreds of averages and almost all of them involved adding values and dividing by the number of values. That framework for understanding *average* can be an impediment to understanding *average change*. Arthur might be reading "average change of f" as "average of f" or more precisely "average output of f." If you look only at the y-values of the points, the average of A and C is clearly positive (both are above the axis) and larger than the average of B and C, leading to Arthur's answer.

Even a student who avoids these difficulties and answers the question correctly might not have the same mental image as you do. An expert has the ability to "see" how continuously changing one variable affects the other parts of a picture. In this case, such a dynamic understanding might involve seeing the secant line that connects A and C slowly pivoting around the point C until it becomes the line BC, and then continues until it becomes the tangent line at C. Figure 2.1 provides a still shot from an applet (this one written in Geogebra) that illustrates this motion.

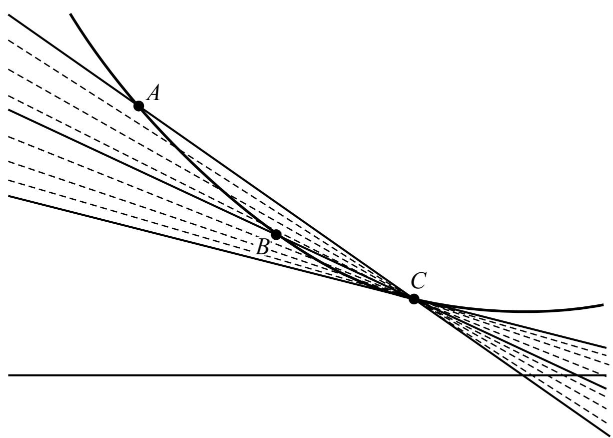


Figure 2.1. Sliding Secant Becomes a Tangent

While experts can coordinate in their minds how one point moving in one direction affects the slope of the secant line, calculus students may be stuck with only a few still shots of that movie. This type of concept arises repeatedly throughout college mathematics, enough that researchers have given a name to the idea of coordinating the change in an input variable with the change in the output: **covariation** [16, 4]. Arthur, like many students, may not be able to visualize how the slope of a secant line changes as one point moves along a curve. Helping your students develop their covariation skills at the beginning of the calculus sequence could prove enormously important for their future success. (For additional discussion of covariational reasoning and strategies for helping students develop it, refer to the Functions section of Chapter Zero.) Similar thinking is vital when you imagine how the area inside a set of boxes changes as the number of boxes increases, how a Taylor series starts to resemble a sine curve as you take more and more terms, or (looking ahead to differential equations) how changes in the populations of a predator and a prey affect each other.

Non-commutative Algorithms

2. If $f(x) = \sin(2x)$, find $f'(\frac{\pi}{4})$.

Bertha:

$$f(\pi/4) = \sin(2\ \pi/4)$$
$$= \sin(\pi/2)$$
$$= 1$$
$$f'(\pi/4) = (1)' = 0$$

The logical implications of Bertha's work on Problem 2 might be a bit shocking. Instead of differentiating and then plugging in $\frac{\pi}{4}$, she plugs in and then differentiates. So the derivative of *every* function at *every* point is zero! Calculus is over. We're out of work.

Before we dismiss her work as ludicrous, it's worth pointing out that nothing in the notation $f'(a)$ suggests that the order of the two operations (take the derivative, plug in a) should be completed in that order. Only with the benefit of a rich understanding does one interpretation become ludicrous and the other obviously the right one.

Turning to what Bertha writes, you may think that she is just confused about the order of operations when differentiating and only needs to be reminded that you have to differentiate before you plug in a value. This type of error, however, may reveal a deeper issue with her understanding of the derivative itself which is not apt to be fixed with just a reminder. Bertha may have mastered the procedure of taking derivatives—her work does show quite a bit of correct thinking—but she hasn't tied that procedural understanding into the concept of the derivative as a *function* representing the *rate of change*. Putting her in positions where she has to explain what she means to her peers might help her see the logical flaw in her thinking.

Overgeneralization of Procedures

3. If $f(x) = (3 - x^2)^3$, find $f'(x)$.

Cristobal:

$$f'(x) = 3(3-x^2)^2(-2x)(-2)$$

4. If $g(x) = e^{x+x^2}$, find $g''(x)$.

Dolly:

$$g'(x) = e^{x+x^2}$$
$$g''(x) = e^{x+x^2}$$

Edouard:

$$g'(x) = e^{x+x^2}(1+2x)$$
$$g''(x) = e^{x+x^2}(1+2x)^2$$

In Problems 3 and 4 we have two examples of a classic error: *overgeneralizing* a procedure—that is, using a procedure correctly, but in a circumstance where it isn't appropriate [5]. For example, in the first problem, Cristobal overgeneralizes the chain rule, by "multiplying by the derivative of the inside" correctly once and then doing it again (that's where the (-2) on the end of the expression came from). The first application of the chain rule is great, but the second goes a step too far—applying the rule to the term $(-2x)$, which isn't the "inside" of the original function. The form of this function probably triggers a very procedural level of thinking—"multiply by the derivative of what's inside" [27, 29]. Given Cristobal's work, what he views as the "inside" indicates an incomplete understanding of composite functions and how the chain rule applies to them.

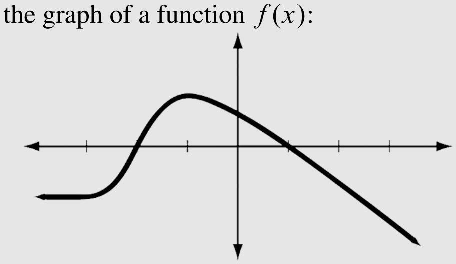
For Problem 4, both students interpret the fact that e^x is its own derivative in novel ways, a fairly common occurrence [12]. Dolly overgeneralizes this rule by viewing the more complex composite function $g(x) = e^{x+x^2}$ as behaving the same as the basic function $g(x) = e^x$. Her answer could indicate a lack of understanding of the behavior of e^x or of the chain rule, or both.

On the other hand, Edouard realizes the need to use the chain rule, but overgeneralizes, seeing $g'(x) = e^{x+x^2}(1+2x)$ as sharing the basic property of functions like e^x and ignoring the fact that the first derivative is now the *product* of an exponential and a linear term. To generate the second derivative from $g'(x) = e^{x+x^2}(1+2x)$ he probably figures he just needs to do the same thing he did earlier ("you just multiply by $(1+2x)$"), leading to the $(1+2x)^2$ term at the end of the expression.

The answers to both of these examples might be merely the result of the student's momentary lapse in concentration, in which case pointing out the error might be all that is necessary to ensure that similar errors don't occur in the future. However, researchers have documented that these answers can indicate confusion about the nature of functions, composite functions, and the chain rule [6]. For instance, some students might be confusing compositions of functions with products of functions. Without more evidence about what this student is thinking, we'd cautiously assume that the issues exhibited in this work form the tip of a larger iceberg, and we would try to put the students in positions where they would explain their thinking (to their peers or to us).

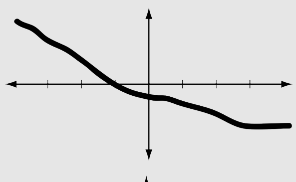
Graphical Gaffes

5. Pictured here is the graph of a function $f(x)$:

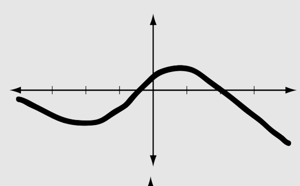


Sketch a graph of the derivative, $f'(x)$.

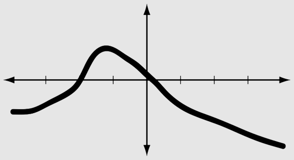
Fay:



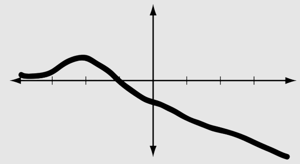
Gustav:



Hanna:



Ike:



Problem 5 is a type of graphical question that has become common in recent years and has proven to be quite challenging for students. If you haven't already, take a minute to look at all four of the responses to Problem 5, none of which is exactly right. What was each student thinking?

Two of the students, Fay and Ike, appear to understand the connection between the increase/decrease nature of the function and the positive/negative value of the derivative. However, it seems that Fay's knowledge stops there, with none of the subtleties of the graph reflected in her answer. Importantly, she ignores the *magnitude* of increase or decrease of the function. Hanna's graph looks suspiciously like the original function, a well-documented error [17, 20]. She probably moved along the graph of f, analyzing the situation at each point, but instead of focusing on the *value* of the slope at each point, she uses small segments of the tangent lines themselves to compose the graph of the derivative [8].

Gustav's work might be the most confusing to analyze, as he appears to have gone in the wrong direction, drawing something close to an antiderivative for f instead of drawing f'. The underlying cause of this thinking might stem from misremembering or not understanding the first derivative test. For example, he might think that the *function* changes from positive to negative where the *derivative* has a local maximum. Looking closely at his graph reveals that the local extrema of his graph match up with the zeroes of f. Gustav and Ike's graphs also fail to correctly account for the linear nature of f for large values of x. While algebraically, it is clear that the derivative of a linear function is a constant, the fact that the tangent line to the function lies *directly on top of* the function leads students to make the same mistake that Hanna does, using the function's tangent line as its derivative [8].

The Derivative is a Function

6. Fill in the blanks:

Josephine: If f' is positive, f is increasing

Josephine: If f'' is positive, f' is ???

Most students we work with eventually get that a positive derivative means a function is increasing. But do they really understand? If Josephine got the full generality of that statement, she would answer both parts of Problem 6 without difficulty. However, she's not alone in getting the first problem correct and the second one wrong.

What underlies such student thinking is actually somewhat simple. While on some level they understand that the connection between a positive derivative and an increasing function applies to any function, research has shown that many students fail to see that *the derivative is itself a function*, one to which *this same connection would apply* [4]. Combined with the fact that f'' might be seen as the second derivative of f, but not as the first derivative of f'—some students are left baffled by the second problem but not the first. In terms of the three-part learning of the derivative (difference quotient, slope at a point, slope at all points), Josephine might actually have a reasonably good grasp on all but the last part, understanding that the derivative is itself a function [29]. If she views the derivative as a process leading her to an answer, then seeing that in this case, her answer is actually a **function** might still be elusive. This issue returns in differential equations, where solutions are **always** functions! (See the Differential Equations chapter for more.)

What You Can Do

Students' difficulties with derivatives can largely be split into two categories: procedural and conceptual. However, contrary to what you may think, it is not actually the case that conceptual understanding will come with enough procedural practice. Research suggests that they need to be built together [24] and that it is also possible that focusing exclusively on procedural understanding first can *actively interfere* with developing conceptual understanding (whereas focusing on concepts first appears not to compromise acquisition of procedures later on) [23].

With that in mind, what can we do to help students make the most out of their practice? Several instructional approaches have proven especially effective in college calculus. First, students, even those from historically underserved groups, can succeed when presented with opportunities to practice their skills in environments where the stakes are low and their difficulties will be quickly be addressed [13]. For some, this might take the form of a long list of problems to do for homework; however, we should always be aware that marginalized students will be more likely to do such assignments alone, where simple mistakes will turn into frustration instead of being corrected by friends.

Having students work with each other in class can be particularly productive when trying to build computational fluency [13]. In fact, with so many students coming into college calculus having already taken it in high school, some instructors have the students who already learned the chain rule pair up with those who are learning it for the first time. The "teachers" will understand the procedure better for having taught it [26], and the "learners" will benefit from the more individualized attention to their mistakes [25, 11].

It may also be useful to check on students' understanding of variables and functions to make sure that underlying confusion about those are not contributing to difficulties with procedures. As discussed in the section about variables in Chapter Zero, students may not yet have mastered the conventions for how we name constants and variables. This can contribute to what appear to be procedural errors but are actually symptomatic of something more complex. Dealing fluently with problems that involve terms such as $f(x)$, $f(5)$, c, and $f'(c)$ requires knowing which terms are numbers and which are functions. Giving students a list of such terms and getting them to state whether they are numbers or functions is a useful diagnostic tool. This may reveal the (underdeveloped) thinking that is contributing to the errors you see when students work with derivative (and other) expressions and equations.

Students' difficulties that are purely conceptual in nature are harder to grapple with and require different kinds of questions to detect. Many students will be able to correctly take any derivative you throw at them, and still not understand the underlying concepts [28]. Innovative (and effective) approaches to developing the concept of the derivative have taken several forms.

One approach is to build on a specific example of a rate of change that is most familiar to students: velocity (see [9] for a discussion of this approach). Motion provides a common experience that, when properly harnessed, can form the basis for students' conceptual understanding of limits. Physical activities can help to make connections between the abstract nature of the algebra and their previous knowledge [18]. For instance, having students "walk out" a position versus time graph can provide a visceral lesson in how position and velocity are related (especially if you have access to motion sensors that convert motion into graphs). And students' own experiences walking slowly, quickly, and backwards will allow them to graph the velocity graph—even before the derivative has been introduced.

While using the idea of motion can prove to be highly effective, any classroom demonstration is fraught with its own set of difficulties [7]. Converting from spacial to temporal variables is difficult. For instance, students who are shown a parabolic height versus time graph frequently mistakenly conclude that the object itself travelled in a parabolic arc, while it might have only moved straight up then straight down [14, 15]. Students who witness a demonstration also have a difficult time focusing on the key aspects of what they see (e.g., "I know he's going at a constant rate, but his head is bobbing up and down so I don't see how the graph can be a straight line").

Physics education research suggests that the effectiveness of using demonstrations in class is greatly improved if you do several things. Students who predict beforehand what will happen, then watch a demonstration, then have the opportunity to discuss with each other why the result did or did not conform to their predictions are much more likely to see what you want them to see and to remember it later [7]. Without that structure to process observations, humans are likely to "see" things in the demonstration that confirm their current thinking instead of seeing the demonstration as providing a counterexample to their views. A technological alternative to live demonstrations is the use of a computer program to model the situation. Such programs have proven successful at improving students' conceptual understanding of functions and change [3], possibly by helping them visualize the movie instead of just still shots (see the first problem in this chapter). Free versions of many such visualization/demonstration tools can be located with an internet search and many publishers now include such tools along with etexts and instructor resources for their textbooks.

Another technological approach is to use graphing devices to help students visualize functions, secant and tangent lines, and derivatives. Such strategies have been shown to help students develop robust understandings about functions, especially to augment the highly restrictive pointwise view of functions (see [22] for a review of this literature). To help students translate static images into motion, more advanced graphing calculators, computer algebra systems, and well-designed java applets can literally animate secant lines as they approximate tangent lines. Several studies have shown that the use of such technological tools can increase students' conceptual understanding without sacrificing much of their procedural fluency [21, 10]. Others have gone further, using elementary programming languages to allow students to construct the derivative (and even discover the Fundamental Theorem of Calculus) [2].

Technological tools can even help students with tasks like Problem 5, which asks for the graph of the derivative given the graph of a function. Marc Renault's Try to Graph the Derivative Function has proven both popular and effective with students.

For teachers who are inclined to take a more abstract approach to derivatives, there are other ways to ensure that students develop the underlying conceptual understanding. Some researchers suggest that students will be more likely to see the limit as the underpinning of the derivative if the language used in both cases is the language of approximation [19]. For instance, as we suggested in the section on limits, you might introduce the derivative as the following way of filling out a canonical limit sentence:

> You can make the slope of the secant line arbitrarily close to the slope of the tangent line by taking the parameter h sufficiently close to zero.

Finally, simply being aware that understanding the derivative involves a multistep process with increasing levels of abstraction can help us get inside students' heads and figure out how to help them move towards a robust understanding of the ideas.

> **A Good Place to Start**
> Make sure students build on a solid foundation, with a deep understanding of rates of change, before helping them see the derivative at a point as a movie, with secant lines approximating the tangent line.

Bibliography

[1] T.M. Apostol. *Calculus Vol. I.* J. Wiley & Sons, 1967.

[2] M. Asiala, J. Cottrill, E. Dubinsky, and K. Schwingendorf. The development of students' graphical understanding of the derivative. *Journal of Mathematical Behavior*, 16(4):399–431, 1997.

[3] N. Balacheff and J.J. Kaput. Computer-Based Learning Environments in Mathematics'. *International Handbook of Mathematics Education*, pages 469–504, 1996.

[4] M.P. Carlson. A cross-sectional investigation of the development of the function concept. In E. Dubinsky, Schoenfeld A. H., and Kaput J., editors, *Research in Collegiate Mathematics Education*, volume 1, pages 114–162. American Mathematical Society, Washington, DC, 1998.

[5] B.A. Cipra. *Misteaks... and How to Find Them Before the Teacher Does*. AK Peters, Ltd., 2000.

[6] J.M. Clark, F. Cordero, J. Cottrill, B. Czarnocha, D.J. DeVries, D. St John, G. Tolias, and D. Vidakovic. Constructing a schema: The case of the chain rule? *Journal of Mathematical Behavior*, 16(4):345–364, 1997.

[7] C.H. Crouch, A.P. Fagen, J.P. Callan, and E. Mazur. Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics*, 72:835, 2004.

[8] J. Ferrini-Mundy and K. Graham. Research in calculus learning: understanding of limits, derivatives, and integrals. In J. Kaput and E. Dubinsky, editors, *Research Issues in Undergraduate Math Learning: Preliminary Analyses and Results (MAA Notes 33)*, pages 31–45. Mathematical Association of America, Washington, DC, 1994.

[9] K. Gravemeijer and M. Doorman. Context Problems in Realistic Mathematics Education: A Calculus Course as an Example. *Educational Studies in Mathematics*, 39(1):111–129, 1999.

[10] M.K. Heid. Resequencing Skills and Concepts in Applied Calculus Using the Computer as a Tool. *Journal for Research in Mathematics Education*, 19(1):3–25, 1988.

[11] A. H. Herzig and D. Kung. *Cooperative learning in calculus reform: What have we learned?*, volume 5. Conference Board of Mathematical Sciences: Issues in Mathematics Education, Washington, DC, 2003.

[12] Aladar Horvath. Looking at calculus students' understanding from the inside-out: The relationship between the chain rule and function composition. In *Eleventh Conference on Research in Undergraduate Mathematics Education (RUME)*, 2008.

[13] E. Hsu, T.J. Murphy, and U. Treisman. Supporting High Achievement in Introductory Mathematics Courses: What We Have Learned from 30 Years of the Emerging Scholars Program. In M. Carlson and C. Rasmussen, editors, *Making the Connection: Research and Practice in Undergraduate Mathematics Education* (MAA Notes 73). Mathematical Association of America, 2008, pp. 205–220.

[14] G. Leinhardt, O. Zaslavsky, and M.K. Stein. Functions, Graphs, and Graphing: Tasks, Learning, and Teaching. *Review of Educational Research*, 60(1):1–64, 1990.

[15] L.C. McDermott, M.L. Rosenquist, and E.H. van Zee. Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics*, 55:503, 1987.

[16] G.S. Monk. Students' Understanding of Functions in Calculus Courses. *Humanistic Mathematics Network Journal*, 9:p21–27, 1994.

[17] R. Nemirovsky and A. Rubin. *Students' Tendency To Assume Resemblances between a Function and Its Derivative.* TERC Communications, 2067 Massachusetts Avenue, Cambridge, MA 02140 ($5)., 1992.

[18] Ricardo Nemirovsky and Chris Rasmussen. A case study of how kinesthetic experiences can participate in and transfer to work with equations. *International Group for the Psychology of M athematics Education*, page 9, 2005.

[19] M.C.F. Oehrtman. Strong and Weak Metaphors for Limits. *PME CONFERENCE*, 3:397–404, 2003.

[20] A. Orton. Students' understanding of differentiation. *Educational Studies in Mathematics*, 14(3):235–250, 1983.

[21] J.R. Palmiter. Effects of Computer Algebra Systems on Concept and Skill Acquisition in Calculus. *Journal for Research in Mathematics Education*, 22(2):151–156, 1991.

[22] M. Penglase and S. Arnold. The graphics calculator in mathematics education: A critical review of recent research. *Mathematics Education Research Journal*, 8(1):58–90, 1996.

[23] Dolores D. Pesek and David Kirshner. Interference of instrumental instruction in subsequent relational learning. *Journal for Research in Mathematics Education*, pages 524–540, 2000.

[24] Bethany Rittle-Johnson, Robert S Siegler, and Martha Wagner Alibali. Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology*, 93(2):346, 2001.

[25] Elizabeth C Rogers, Barbara E. Reynolds, Neil A. Davidson, and Anthony D. Thomas. Cooperative Learning in Undergraduate Mathematics Issues that Matter and Strategies that Work, 2001.

[26] Rod D. Roscoe and Michelene Chi. Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4):534–574, 2007.

[27] Glenn Gordon Smith and David Ferguson. Diagrams and math notation in e-learning: Growing pains of a new generation. *International Journal of Mathematical Education in Science and Technology*, 35(5):681–695, 2004.

[28] P. White and M. Mitchelmore. Conceptual Knowledge in Introductory Calculus. *Journal for Research in Mathematics Education*, 27(1):79–95, 1996.

[29] M.J. Zandieh. A theoretical framework for analyzing student understanding of the concept of derivative. *Research in Collegiate Mathematics Education. IV. CBMS Issues in Mathematics Education*, pages 103–127, 2000.

3

Integration and the Fundamental Theorem of Calculus

> It appears very unlikely that the introduction of integration can be made easy, assuming that we wish students to grasp more than the ability to integrate and obtain the answers to simple applications. [23, p. 9]

Research on the teaching and learning of integration suggests exactly this sentiment: there is no silver bullet that makes integration easy for students. The quote also captures how instructors may feel when attempting to help students appreciate the wonderful ideas associated with integration, especially if these big ideas are introduced near the end of a long term packed full of big ideas. However, researchers have detailed the many ways that students find the ideas challenging, what their productive and unproductive thinking can be, and some of the things we can do to help guide them toward our goal: a conceptual understanding of integration, how it relates to differentiation, and the ability to apply that knowledge in computing integrals.

Most calculus texts begin with differentiation and then move on to its inverse, integration. Packed within the integration portion of the text are a large number of challenging concepts, pieces of the puzzle that perfectly fit together to produce the true heart of calculus:

The Fundamental Theorem of Calculus (FTC)

Or at least that's the way mathematicians see it.

Many students don't share this view. One author (Dave) once started a spring Calculus II course by asking students to write down everything they'd learned in the first semester of calculus. They filled the board with examples, computations, differentiation rules, limit laws, anti-derivatives, but not one of them wrote the FTC which was supposed to be the culmination of the whole semester's work! Aside from the obvious jokes ("What part of *Fundamental* wasn't clear? Did your professor call it *A Marginally Important* Theorem of Calculus?") these students' responses should give any calculus instructor food for thought. Why don't more students understand how the FTC perfectly ties together the ideas of integration and differentiation? Where do we lose students as we progress through the ideas in those sections of their calculus textbook? What earlier ideas are key to developing a robust understanding of what integrals are? What sense are they making of Riemann sums, limits of sums, antiderivatives, accumulation functions, and finally both parts of the FTC?

We want students to not only be able to calculate anti-derivatives and integrals correctly, we also want them to understand how the integral is defined as accumulating quantities that can be modeled as areas; how the value of the integral changes as the right endpoint varies; how the limit idea that they first saw with derivatives is now used to approximate an exact answer for the area under the curve; and finally, how the two foundational ideas of calculus, the derivative and the integral, actually turn out to be inverse processes in one of the truly momentous achievements of human thought: the Fundamental Theorem of Calculus. (Sorry, we got a little carried away there.)

Here we begin with findings from research into students' thinking about the various topics that come together in our study of integrals.

What Is an Integral?

> **1.** In your own words, define $\int_a^b f(t)\, dt$.
>
> **Alvin:** It's the area under f between a and b.
>
> **Barbara:** You undo the derivative, plug in b and a, and then subtract the two.
>
> **Cosme:** You have to put boxes both above and below f, and then let the number of boxes go to infinity.
>
> **Dalila:** It's the limit of the Riemann sums.
>
> —[14]

Before we analyze these students' responses to Problem 1, take a minute to answer this question for yourself. If you look at a sampling of textbooks, their answers range nearly as much as the answers above. However, most books word the description of the definite integral more carefully than these four students.

Leaving aside any hypotheses on f, the four responses above show very different views of what the integral is. Alvin's response indicates that he has the general idea of one of the problems that the Riemann integral solves, but might not understand the mechanisms used to solve it. (We should note that some books we've seen do give this as one definition of the definite integral.)

Barbara's response is common among those who have already learned the FTC, and indicates two problems that resurface repeatedly through college mathematics [12, 1, 11]. First, many students depend on an algorithmic view of mathematics learned through years of coping with math classes by figuring out how to do the problems without really understanding them. Here, Barbara sees the integral as asking her to carry out a particular set of steps, and describes those steps. Secondly, Barbara's answer indicates a lack of understanding of how mathematical definitions differ from the implications of those definitions. The FTC is an implication of the definition of the integral, not its definition. (More on troubles with definitions can be found in the chapter on Proofs, where this misunderstanding comes to the surface more frequently.)

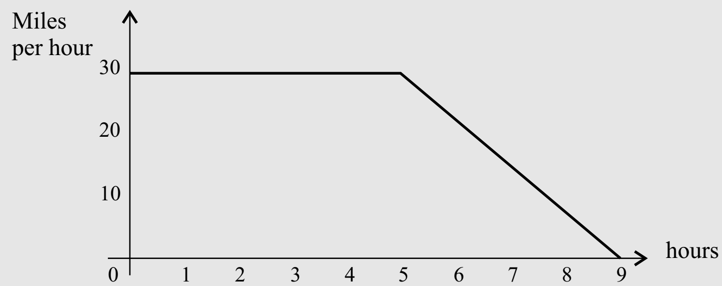
Cosme and Dalila both respond in ways that hint at the definite integral as a limit of sums of areas of boxes. Despite Dalila's textbook-perfect answer, Cosme's answer is more convincing in that he has a grasp of the underlying ideas. While Dalila's words could be memorized, Cosme's slightly awkward phrasing is less likely to be a rote response. Cosme's also hints at a more sophisticated understanding of integration, since the upper and lower sums are typically used to show that the Riemann integral exists.

Because many instructors reserve a detailed discussion of the existence of the Riemann integral until an upper-level course in analysis, it's somewhat understandable that many students, like the four above, completely ignore the conditions under which the integral exists [13]. Many books choose to simplify the situation, requiring f to be continuous and initially restricting attention to the case where $f \geq 0$ and a and b are finite. Many students who mention hypotheses get them wrong, for instance saying that f must be bounded. This suggests that many of those who actually do write "f must be continuous" may actually be guessing, instead of understanding why the continuity of the function implies the existence of the integral. (Though not suitable for a calculus class, we can't help but mention Lebesgue's amazing theorem, giving a necessary and sufficient condition for the existence of the Riemann integral: that the discontinuities of the function have measure zero.)

These four responses capture many of the problems students have in understanding what the definite integral means. All of the views of integrals that we want students to develop are predicated on (1) area as a model for multiplication and (2) an understanding of definite integrals as a measure of *accumulation.* These most basic ideas may not be emphasized clearly nor be well-understood by all students. And, without this firm foundation, the ideas and computations associated with integrals may lack any meaning for students.

Integral as Area

2: Here's the graph for the rate versus time for a car:



How far does the car travel between hour 0 and hour 9?

Erick: Looks like it's averaging about 20 mph. Doing that for 9 hours would get you . . . about 180 miles—what does this have to do with calculus?

Buried in this problem is a deep connection between integrals and areas, a connection Erick seems not yet to have developed. Before we dive into the calculus of integrals, let's step back and take a look at how students' conceptions of area can influence their understanding of integrals.

As discussed in more detail in Chapter Zero, area is a powerful and very useful model for multiplication. Although we are typically introduced to the connections between area and multiplication for the purposes of computing the area of some physical object (the top of a desk, carpet needed for a particular room, etc.) where answers are given in square units of some sort, this area-multiplication connection can also be utilized in other ways. For example, if we use a rectangle and represent time (in seconds) on one of the sides and rate (in meters per second), the "area" of the rectangle is still obtained via multiplication but now the answer is in units of distance (meters in this case). As you can see in Figure 3.1, planar areas can be used to model the product of various quantities. Taking this approach requires an understanding that there are *physical* areas (tops of desks, carpet covering floors, etc.) and *graphical* areas. Graphical areas are *area representations* of quantities different from those we encounter in the physical world as areas (e.g., distance, pounds, etc.). If a student understands that the area of a rectangle can represent the product of any two quantities, summing up areas of rectangles to obtain the accumulation of some quantity may seem a whole lot less mysterious.

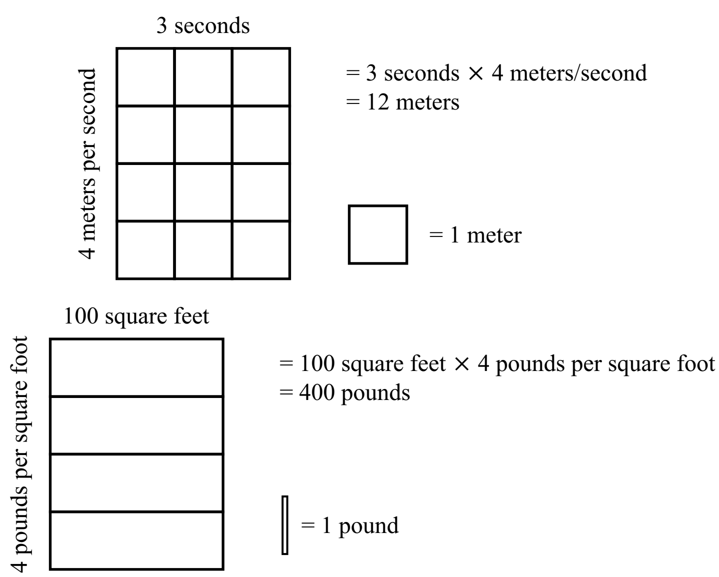


Figure 3.1. Area Models for Distance and Pounds

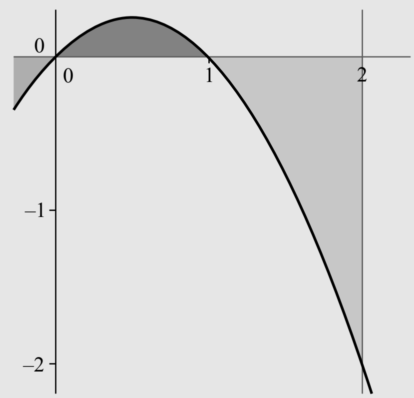
Unfortunately, this way of thinking about and (importantly) *visualizing* multiplication as an area is not necessarily at the disposal of all calculus students [8] but is worth the time to help them develop.

In Problem 2, Erick answers in a way that leaves us wondering if he's really thinking about area at all. He might be estimating the average height of the function ("about 20 mph,") looking at the total length of the trip ("9 hours") and then using the fact that $distance = rate \times time$ to get his answer.

As for what this has to do with calculus, the quantity we seek is the accumulation of distance and that is modeled with the area under our (in this case, piecewise linear) function. That's exactly what an integral does! The Riemann sum underlying that integral would involve products $f(x_i^*)\Delta x_i$ representing the area of a rectangle and that area gives us the distance traveled over that small interval. Students who don't see $f(x_i^*)\Delta x_i$ as the area of a rectangle have little hope of understanding why the Riemann sum expression represents the accumulation of the area of many rectangles. If they don't see that the units of that expression are simply *miles*, they are unlikely to fully understand the power of the integral.

As powerful as this *multiplication as area* model is, it also introduces some confusion . . .

3. Find the total area of the bounded regions enclosed by the x-axis and the function $f(x) = x - x^2$, between $x = 0$ and $x = 2$. Explain your reasoning.



Flossie:

$$\int_0^2 x-x^2\,dx = \frac{x^2}{2}-\frac{x^3}{3}\Big|_0^2$$

$$= \frac{4}{2}-\frac{8}{3} = \frac{12}{6}-\frac{16}{6} = -\frac{4}{3} = -\frac{2}{3}$$

$\frac{2}{3}$ (must be positive)

Gil:

Split it at $x=1$

$$\int_0^2 x-x^2\,dx = \int_0^1 x-x^2\,dx + \int_1^2 x-x^2\,dx$$

$$= \frac{x^2}{2}-\frac{x^3}{3}\Big|_0^1 + \frac{x^2}{2}-\frac{x^3}{3}\Big|_1^2$$

$$= \left[\frac{1}{2}-\frac{1}{3}\right] - 0 + \left[\frac{4}{2}-\frac{8}{3}\right] - \left[\frac{1}{2}-\frac{1}{3}\right]$$

$$= \frac{1}{6} + \left(-\frac{2}{3}-\frac{1}{6}\right) = \frac{1}{6} + \left(-\frac{5}{6}\right)$$

$$= \frac{1}{6}+\frac{5}{6} = 1$$ Should be +

In an effort to convey the idea of area being a physical covering, elementary school instruction often includes the declaration: areas are always positive! While this is obviously true when we are measuring a wall or a tablecloth, students may overgeneralize this to situations that do not involve *physical areas* and conclude that even in situations where area is being used as a *representation of multiplication*, the answer must be positive.

For many students, the first few days of working with integrals can leave them with the impression that integral = area [2, 23]. They then conclude that because area is always positive, integrals must always be positive.

We then muddy the waters for students when we ask them questions about the physical area and questions where they need to utilize the sign of the quantity represented by the area—a distinction that can take time to understand.

Problem 3 asks for the physical area. We see symptoms of the challenges such problems create reflected in Flossie's work. Remembering the (over-simplified) idea that integral = area, she carefully computes the integral, getting $-\frac{2}{3}$. Believing that areas are never negative, she quickly "corrects" her answer, possibly without questioning the methods that got her a negative number in the first place.

Gil makes a more successful attempt at the problem, even getting the correct answer. He knows to split the integral where it hits the axis, computing two separate integrals. Though his calculations are correct, the way he writes his answer is a little problematic. His initial integral, $\int_0^2 (x - x^2)\ dx$ equals $-\frac{2}{3}$, as Flossie's computation demonstrates, not 1, as his string of equalities suggests.

Does Gil really understand why he needs to split the integral? Could he explain what question he'd *actually* have answered if he *had not* split it? Orton reported that many students know what process to go through to get the right answer to problems like this without knowing why [23]. A student who understood the underlying ideas could explain that if positive y-values represent something happening (or moving in a particular direction), then the negative y-values represent the opposite process (or movement in the opposite direction). They would then add that in this case, the problem asks for the total (physical) area and so it is the (positive) magnitudes that need to be combined.

Importantly, the student would also be able to describe when one needs the combination of the areas as the accumulation (positive plus negative) of these things over the interval of integration. With this understanding, combining the integrals for different sections of the graph in different ways (utilizing the different signs in some cases, utilizing only the magnitudes in others) can make sense to students. They would also be able to correctly write down a formula for the area, in this case, $\int_0^1 (x - x^2)\ dx - \int_1^2 (x - x^2)\ dx$, before going through the calculations. In contrast, Gil only notices the need for a negative sign after he has calculated the second integral.

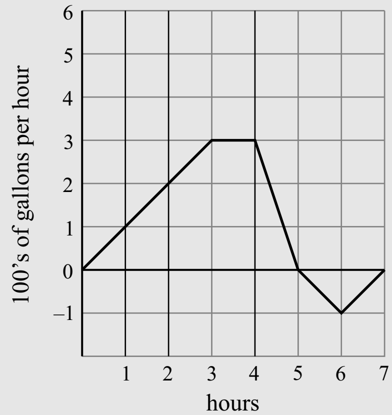
If you understand that area can be used to represent or model multiplication, then the sign of the product depends on the signs of the things being multiplied. A rectangle with time as the width and velocity as the height can indeed have an area that is negative if that velocity is, for example, -50 meters per second.

In addition to negative areas that are formed when the area in question is located below the x-axis, an area can be negative if we are reversing the conventional direction (for accumulating areas) and going instead from right to left. This creates a negative change which can be modeled by making dx negative, thus generating areas with the opposite sign from what is obtained when a "regular" integral is computed. Above we described the difficulties students can have when the height of the area (or rectangle being used to approximate it) represents a negative quantity; here we see that students also have to grapple with the width representing a negative quantity. This comes up in physics applications when some process (for example, in thermodynamics) is being run backwards, thus making the sign of the change in the horizontal direction opposite of the left-to-right direction that gives us a positive value for dx. A student with this understanding would be able to explain why if $f < 0$ on [0, 1], then $\int_0^1 f dx < 0$ and $\int_1^0 f dx > 0$. However, without a solid foundation in the idea of area as a model for multiplication ideas, students may memorize but not understand these important points.

Integral as Accumulation

Before you understand the FTC, you have to view the definite integral as an accumulator function, with an output that changes as the upper bound changes. While many texts devote little space to viewing an integral in this way, researchers generally agree that this step is crucial to getting students to understand the FTC (see, e.g., [28]). Students find it very difficult to reason about the accumulation of one quantity, even in a realistic setting such as the one presented in Problem 4.

4. Let $f(x)$ represent the rate at which the amount of water in a water reservoir changed (in 100s of gallons per hour) during a 7 hour period starting at noon. At noon, the reservoir was empty. Here is the graph of f:



The function $g(x) = \int_0^x f(t)\,dt$ gives the water level in the reservoir x hours after noon. Describe the graph of g.

Henriette: Hmm. I guess g would be a graph of water and time, but I don't know what it would look like.

Ivo: It goes up until 5 and then down until 7.

Juliette: It goes up slowly at first, getting faster and faster until it's steady at 3. It keeps going up until 5, but at a slower and slower rate. After five it falls back a little bit until 7.

Kiko: For the first three hours, as each hour goes by, the amount it goes up by keeps getting bigger. From 3 to 4, it just increases at a steady rate. From 4 to 5 it increases, but more slowly. At 5, it peaks, going from gaining water to losing it. Between 5 and 6 it goes down—slowly at first, then faster. From 6 to 7 it keeps going down, fast at first and then slower.

Lorena: The function f tells you how quickly the function g increases or decreases, so f tells you how steep to make g—when f is big, g is steep and when f is negative, g decreases. Here, f is biggest between 3 and 4, so g is steepest in that range.

—(Adapted from [4, p. 5].)

As discussed above, to understand accumulation, one has to first understand why some areas represent positive quantities and some represent negative quantities. In addition, students have to deal with covariation—how one variable changes as another changes. For example, in Problem 4, the top limit on the integral (x) is changing and the value of the integral ($g(x)$) is changing as well. Difficulties with covariation reared their head in the Limit and Derivative chapters. Research suggests that underdeveloped covariational reasoning explains many of the difficulties students face throughout calculus [**?**].

The water reservoir problem shown here comes directly from the Carlson group's work on covariation [4, 150], which reported on student difficulties as they struggled to understand situations that involved coordinating changes in one variable with changes in another. Students were well-skilled and well-practiced at determining the value of a function for a particular input (e.g., "Evaluate $f(x) = 3x^2$ at $x = 4$") but success on those tasks did not ensure that they could successfully figure out how the values of $f(x)$ change as x took on values from 1 to 4, for example [19].

After studying student responses to questions and how students eventually came to understand situations such as the one in Problem 4, Carlson and her collaborators developed a framework with five different levels of student thinking [3]. Arranged in order of increasing sophistication, these levels describe the types of thinking that students engage in when working on questions like Problem 4. The five responses to the question illustrate those five levels. As described in [3, 357], the students like those above would be capable of the types of thinking shown in Figure 3.2. These levels are meant to be cumulative so that III would be capable of thinking in ways similar to I and II, but maybe not IV and V.

These levels allow us to characterize how students are thinking about these ideas. It turns out, however, that students

I:	Coordinating the value of one variable with changes in the other variable.
II:	Coordinating the direction of change of one variable with changes in the other.
III:	Coordinating the amount of change of one variable with changes in the other.
IV:	Coordinating the average rate of change of the function with uniform increments of change in the input variable.
V:	Coordinating the instantaneous rate of change of the function with continuous changes in the independent variable for the entire domain of the function.

Figure 3.2. Levels of Student Thinking About Covariation

can sometimes write or say things that are consistent with one level of reasoning, but when their thinking is probed, it becomes clear that they don't possess the understanding needed to actually support that level of thinking. The student answers given in Problem 4 illustrate these five levels (Henriette at level I, Ivo at level II, etc.), but a student who states an answer like the one Juliette gave may actually have only a tentative grasp on the ideas. One might find, upon further questioning, that she actually places only at the second level. As with all categorizations of student thinking, short written or spoken statements can sometimes give only the illusion of substantive understanding. These descriptions of levels are nonetheless a very useful lens through which you can view and preliminary diagnose student thinking.

A deep understanding of the FTC requires exactly the thinking described by level V. Before you take the derivative of an integral, you must be able to coordinate how the value of the integral changes as the upper bound changes. Many students, including those who completed second semester calculus with top grades, had serious difficulties grasping covariation at either level IV or V [3]. Returning to the specifics of the question, one researcher reported that on a similar question, just 28% of students could accurately sketch a graph of the accumulation function $g(x)$ [10].

What's "Definite" About a Definite Integral?

5: What is the difference between the indefinite integral $\int x^2\,dx$ and the definite integral $\int_a^b x^2\,dx$?

Manuel: The first one: you really won't get a really precise answer ... you get an unfinished thing. For the definite integral . . . I guess you're certain where it's ending and starting so that instead of just a general thing...

—[16]

We all have such a solid understanding of what definite integrals and indefinite integrals are that we probably don't even think about the fact that the words "definite" and "indefinite" have meanings in English that aren't exactly compatible with their use in mathematics. But as we saw in Chapter Zero, students who are trying to understand complicated ideas for the first time grasp onto any familiar ideas or words they can find. This means that some students bring their English language definitions of "definite" (something known, precise) and "indefinite" (something that can't be known, imprecise) to their understanding of integration ideas [16, 15]. (See Chapter Zero for more examples of these kinds of ambiguities.)

Depending on the textbook you are using, students might be expected to understand the terms "antiderivative" and "indefinite integral" as synonymous or the use of one term might be heavily emphasized. In either case, when students are then exposed to the idea of a "definite integral," they are apt to make comparisons to the term "indefinite integral" they've seen earlier.

This can all just compound the varied challenges that students face when learning about integration. Thinking of a definite integral as a more precise version of an indefinite integral really is not going to help students understand the FTC or Riemann sums.

Riemann Sums

6. Explain why the formula

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x_i$$

calculates the area under a continuous, positive function f.

Narda: Hmm. I think the formula is missing something. If $n \to \infty$, then the $\Delta x_i \to 0$. Zero times anything is zero, so you'd just be adding up a bunch of zeros.

Octave: As n gets really big, you're just looking at a bunch of line segments that go from the x-axis up to f. Add them up and you get the area under f.

Priscilla: Each box has height $f(x_i^*)$ and width Δx_i. When n goes to infinity, you're adding up an infinite number of those boxes, and that gives you the area.

—[28]

Many students do come to a good understanding of how the $f(x_i^*)$ term measures the height of the ith box at x_i^*, the Δx_i measures the width, the product of the two gives the area, the sum adds up all n areas, and the limit as $n \to \infty$ (or more properly, as the maximum Δx_i approximates 0) ensures that the error approximates 0. (However, if we were resorting to the language of motion, we would say "the error approaches 0.") These three responses, however, highlight the most common unproductive ways students think about Riemann sums detailed in the literature.

Much of the difficulty of the Riemann integral boils down to the challenge of taking the limit of a sum—essentially the limit of a sequence of finite sums. Several researchers have pointed out that asking students to understand this complicated sequential limit before tackling sequences themselves (which usually come later in the second semester) is logically backward [27]. All three of the students whose work is shown in Problem 6 struggle to understand exactly how the limit interacts with the other parts of the formula.

Narda makes a classical analysis mistake: after bringing the limit inside the sum (problematic, since the sum itself **depends** on n), she sees that the width of the boxes goes to zero, making the sum trivial [2]. If asked to justify her work, she might appeal to the limit laws, saying that the limit of the sums is the sum of the limits. Such a limit law, probably learned earlier in the semester and remembered verbally more than symbolically, applies to the sum of two terms, not when the number of terms changes as n increases. A discussion with peers after appropriate prompting might help her toward the realization that if she were correct, every integral would equal zero.

Octave's answer exhibits what Oehrtman calls students' tendency to "collapse dimensions" [20]. Here, he collapses the 2-D boxes into 1-D line segments, and then "adds" those up. He appears untroubled by (and may not even understand) the fact that each line segment would have zero width and thus zero area or that adding up lengths (in, say, *meters*) will never yield area (which would be in *square meters*). The way Octave is thinking about area is not something that just materializes as students learn about Riemann sums. Elementary school students sometimes believe that adding up a series of linear measurements will give them a measure of area. When figuring out the area of a rectangle, they repeatedly move a ruler short distances parallel to one side, moving across the figure and adding up the linear measurements [18]. Students who display this thinking might not understand area as an array of unit squares or how multiplying two lengths gives the number of unit squares contained in the figure. Calculus students demonstrate some of the same difficulties with area as elementary school students [8, **?**].

As problematic as it is, Octave's thinking is common among many students and some experts [24]. Prabhu and Czarnocha researched the history of such thinking and report that Archimedes, Cavalieri, and John Wallis all used similar thinking to make sense of situations that would now fall under the ideas of integration. They hypothesize that students make use of "collapsing dimensions" to bridge a gap in their understanding. They want to see the infinite process (taking increasingly thin boxes) as reaching a conclusion, which in their minds means that the boxes become lines.

Octave's flawed thinking might be exposed if the quantities had units attached to them. The change in position can be found by integrating velocity, but if you follow his thinking and "add up velocities" (say in meters per second) you would get a velocity, not a distance. Only by multiplying a velocity by a (possibly small) amount of time do you get a distance, which you can then add up to get a displacement. Similarly, work comes from a force being applied over a distance, so the total work done is the integral of the force over that distance. Applying Octave's thinking to this situation would mean that work is the sum of an infinite number of forces with no mention of the distances [25].

As with some of the other difficulties and ways of thinking described in this chapter, this relates to the challenges students face when area is used as a representation for multiplication instead of the more familiar *physical area*. To see things appropriately, Octave needs to understand that work is the product of force and distance, modeled with area where the units of that area are appropriate for work. Unfortunately, units are something that undergraduate students struggle with, sometimes providing two-dimensional units in answers to volume tasks and one-dimensional units in answers to area tasks [8]. The issues that show up in students' answers to integral problems may actually be rooted in how they have been thinking about dimensions and units for years.

This reasoning leads to difficulties, similar to those found with derivatives. In that case, the tangent line derives from secant lines which connect a fixed point and a nearby point. Students want to see the infinite process of one point approaching the other as reaching a conclusion, leaving a tangent line that "touches the curve at only one point." The vital importance of the two points *never* coinciding, so that the secant line is always defined, is lost on many students. In both the case of the derivative and of the integral, the root of the problem lies in the complexity of the idea of the limit, and both situations provide excellent opportunities to help build students' understanding of this foundational concept. (See the Limits chapter for more.)

Priscilla's response shows a slightly different naive idea of integrals, in some sense mirroring Octave's response. While he concentrates on the dimensions of the boxes, Priscilla focuses on the number of boxes, claiming that you eventually add up an infinite number of them. (She should, of course, add the areas of the boxes, not the boxes themselves, as her "adding up an infinite number of those boxes" comment suggests!) With a more developed sense of area and limit, and more precision in her words, Priscilla might be in a position to give a more complete answer, saying instead that the eventual area is approximated with an arbitrarily small error, provided the number of boxes is large enough (though still finite.)

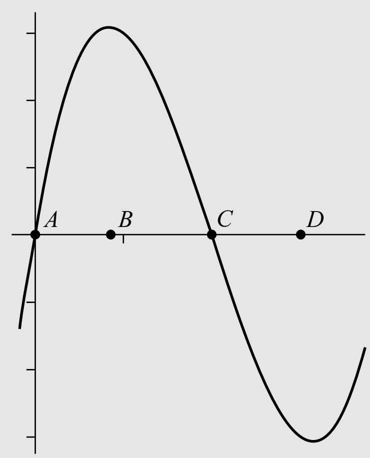
These conceptual roadblocks are only part of students' problems; many also have trouble with the algebraic steps needed to actually calculate the Riemann sums for linear and quadratic functions.

FTC I

7. Here is a graph of $h(t)$:
At what point does the related function

$$p(x) = \int_0^x h(t)\, dt$$

attain a local maximum? Explain.



Qadir: B, where the function reaches its highest point.

Raymond: B, where it's gaining the most area. After that, it gets less and less.

Sonia: C, where it goes from gaining area to losing area—the area is negative after C.

Tico: C, where p' switches from $+$ to $-$

— Adapted from [2].

Here we see a question that probes the connection between accumulation functions and the first part of the FTC. Before we tackle these four students' responses, take a minute to examine possible solution paths. Two basic methods emerge, the first applying the FTC to get $p'(x) = h(x)$. Applying a First Derivative Test at the point $x = C$, where h (and thus p') goes from positive on the left to negative on the right, establishes $x = C$ as giving a local maximum.

A second method focuses on p as accumulating the area bounded by h. As x increases from 0, p accumulates more and more area above the axis until x reaches C. After that, p is accumulating negative values, and thus decreasing. Hence p has a local maximum at $x = C$. The point $x = B$ represents the place p is accumulating **most quickly**, and thus represents an inflection point of p.

Even before students get to either of these solutions, they have to correctly interpret the definition of $p(x)$, a non-trivial task for many. In fact, if presented with the expression $F(?) = \int_0^x h(t)\ dt$ and asked what variable should F depend on to make it a non-constant function, few students will correctly answer x, opting instead for t or even h.

As for these four students, it's not entirely clear what sense they have made of the definition of p. Qadir gives an incorrect answer that indicates "the function" he's referring to in his response is probably h, not p as the question intends. This mistake, confusing the function and its derivative, comes up frequently throughout calculus [2]. Raymond arrives at the same incorrect answer with more complex reasoning. His use of the word "gaining" indicates that the "it" he refers to is the accumulation function p. Unfortunately, he mistakes the place where the accumulation begins to slow down as the local max. Sonia attempts a similar accumulation argument, arriving at the correct answer despite the problematic choice of wording regarding "negative area."

Tico's answer is intriguing in how convincing it is to some—it might even be an answer some mathematicians would give. Still, it leaves open the possibility that he actually understands very little! A close look at his wording reveals that his justification ("where p' switches from + to $-$") just restates the First Derivative Test, without any reference to the specifics of the function p—or to h. Had he instead written "where $p' = h$ switches from + to -" we would be much more convinced that he was correctly applying an argument based on the FTC.

Computational Difficulties with Antiderivatives

One of our beloved senior colleagues opined to his students every semester: "Differentiation is a skill that a monkey can learn. Integration is an art." As with other arts, computing integrals is one that takes a while to master. Here we show you what you're likely to see from your student apprentices, along with a brief description of what they're probably thinking.

Basic Antiderivatives

8. Find:

$$\int x^3\ dx \qquad \int x^{-1}\ dx \qquad \int \sqrt{x}\ dx$$

Ugo:

$$\frac{1}{4}x^4 \qquad \frac{-1}{2}x^{-2} \qquad ???$$

The Power Rule for derivatives seems simple enough, and students quickly learn how to invert it when finding antiderivatives of positive integer powers of x. Negative and fractional powers give them more trouble, as these examples suggest [23]. Although such students almost certainly can do basic arithmetic, the context triggers ideas that push them off track. Here, Ugo's first response indicates some facility with the Power Rule (although he neglects to add an arbitrary constant).

In his second response, Ugo's incorrect answer may come from one of several possible sources. After computing many derivatives with the power rule, sometimes students are so used to carrying out that procedure, they do it in the one case where it doesn't work. In other words, sometimes in the process of generalizing ideas from one situation to another (for example, applying the power rule to a new function with a power they have not already applied the rule to), students *overgeneralize* the rule and apply it in places where it does not belong. Here, Ugo may have applied the "subtract one from the exponent" procedure to a situation where it is not appropriate.

Alternatively, his work might be symptomatic of an issue with negative exponents. In particular, he may have intended to add one to the exponent but in doing that, he added one and got 2 instead (and let the negative sign remain). Yet another possibility is that he first added 1 to -1, realized that would give him the dreaded x^0, wanted to avoid that and came up with -2 instead. Although we all see x^{-1} and $\frac{1}{x}$ as identical, some students may actually react very differently to those expressions.

If the question had been posed as $\int \frac{1}{x}\, dx$ instead, that may have triggered Ugo to recall facts about $\ln(x)$. (Note: This overgeneralizing of algorithms is often characteristic of students' work when they initially encounter new ideas. It can be very useful to learn to recognize differences between difficulties that result from overgeneralizing and those that come from an underdeveloped understanding of the basic algorithm so they can be addressed with students.) When it comes to the fractional power in the third question, he might not realize that the square root translates to a one half power, or he may be unclear about how to add one to that power. Many students who make these kinds of mistakes will catch them if the class norm is to take the time to check work by differentiating, or if they work in groups.

Harder Antiderivatives

9. Compute the following integral: $\int x \cos x \, dx$

Velma: $\frac{1}{2}x^2(-\sin x)+C$

Wallis: $= x(\sin x)+\frac{1}{2}x^2\cos x + C$

—[17]

Both of these students apply an algorithm, and do so pretty well. Unfortunately for them, the algorithms they use are inappropriate in these cases! Neither of these students correctly implements the Integration By Parts argument that will quickly finish off this integral.

Velma's mistaken algorithm, common among calculus students, simply integrates both factors and leaves them as terms in a product [23]. Such a method works perfectly for the sum of two functions, but not for the product. In this case, it's actually possible that Velma checked her work. Applying a similarly flawed Product Rule for derivatives would "confirm" her incorrect answer.

Wallis's work shows a much more complicated (but still flawed) algorithm. Correctly remembering the product rule for derivatives ($(fg)' = fg' + f'g$), he thinks that a similar rule exists for integrals:

$$\int fg\, dx = f \int g dx + \int f dx\ g.$$

Again, these are examples of overgeneralizing: applying correct algorithms in inappropriate ways.

Much Harder Antiderivatives

10. Find:

$$\int \frac{1}{1+x^2} dx \qquad \int \cos(x^2) dx$$

York:

$$= \frac{\ln(1+x^2)}{2x} + C \qquad = \frac{\sin x^2}{2x} + C$$

Here we see a couple of more complicated over-generalizing issues. Both of these questions require undoing the Chain Rule, and these mistakes mirror many of the issues students struggle with when differentiating (see the Derivative chapter for more on this).

To see what York is thinking, imagine replacing the x^2 in both problems with $3x$. After integrating the outside function, you would have to divide by the derivative of the inside which means dividing by 3. She applies the same reasoning to both problems, dividing by the derivative of x^2 both times. She might be working backwards in her head: "When you apply the Chain Rule, you'll pop out a $2x$, so we have to divide by it so they'll cancel." Had she checked her work by differentiating, or consulted with a peer, she might very well have seen the need for the quotient rule and gone back to the drawing board. With some careful prompting, she might even have determined that her "divide by the derivative of the inside" rule only works for linear functions.

Creating answers in this way is likely another case of a student's overgeneralization of an algorithm. York may have successfully used this approach to problems in the past but not yet recognized when a totally different, trig-based approach is called for.

One simple thing you can do to help students move beyond these rookie errors is to institute the following simple rule (and follow it yourself!): *An integration problem isn't done until you've taken the derivative to check your answer.* In addition to supporting procedural fluency with integrals, doing this can also reveal which derivative procedures students are still working to master.

Improper Integrals

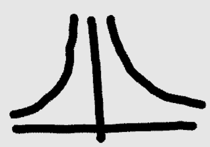
11. Find $\displaystyle\int_{-1}^{1} \frac{1}{x^2}\,dx$. Explain your work.

Zelda: $= -x^{-1} = \frac{-1}{x}\Big|_{-1}^{1} = -1-(-\frac{1}{-1}) = -2$

Akoni: $= -\frac{1}{x}\Big|_{-1}^{1} = -1-(-\frac{1}{-1}) = -2$ $+2$

(it's an area, must be +)

Ema:



You have to split it at 0. Since both sides go to infinity, both integrals are infinite.

Hone: I get -2, but I think I forgot something about the constant of integration.

—[13, p. 80]

Researchers who worked with students on improper integrals reported some dispiriting results. Few students who worked on this question understood that the function's asymptote at $x = 0$ presents a problem, and fewer still knew what to do about it [13]. Many students, it seems, take the attitude that if you can write it down, it must exist (and be finite).

Zelda approaches the problem from a purely computational, algebraic perspective, producing the wrong answer and leaving it at that. Akoni takes this approach one step further, nominally tying the computation back to the idea of area and changing the sign of his answer. In addition to missing the improper part of this integral, he might also be suffering from the positive area syndrome discussed earlier. Had the function (and its integral) been negative, he might have said exactly the same thing, switching a correct answer into an incorrect one!

Ema acknowledges the challenge presented by the asymptote, and correctly wants to split the integral into two parts. Unfortunately, she then reasons that "both sides go to infinity" implies that both integrals diverge. While the conclusion is true in this case, her logic is flawed, as the example $\frac{1}{\sqrt{|x|}}$ shows. Her incorrect thinking, intimately tied up with other issues surrounding infinity, boils down to the idea that if you add up things that have some infinite property, the result must be infinite.

The classic example to challenge this thinking is Gabriel's horn ($\frac{1}{x}$ rotated around the x-axis between $x = 1$ and ∞), which paradoxically has finite volume but infinite surface area. In this paradox, as with Ema's work, the unbounded nature of one attribute (surface area in the case of the horn / height of the function in Ema's case) does not imply that other attributes are also infinite (volume of the horn / integral of the function).

Finally, Hone seems disturbed by his negative answer, but mistakenly draws on ideas of *indefinite* integrals and their arbitrary constants as a possible fix.

What You Can Do

While the research on integration makes it painfully obvious how difficult these concepts are for students, it also gives us some hints for improving the situation. Some suggestions probably stretch a bit beyond the scope of what you're looking for. Orton, for instance, suggests that we should be teaching ideas of limits as early as grade school [23]. Others suggest something less drastic than changing the grade school curriculum: cover limits of sequences before tackling integrals [6]. Below, we distill researchers' suggestions, organized into four categories. First, we describe ways to utilize area models to help students develop foundational ideas about accumulation. Then we look at ways to improve students' background knowledge of covariation and limits, both of which are vital to understanding integrals. Third, we focus specifically on suggestions for improving the teaching (and learning) of Riemann sums. Finally we tackle the most important, and most conceptually challenging, topic: the Fundamental Theorem of Calculus.

Exploit Area Models

As discussed in Chapter Zero, area works extremely well as a representation for products, but it is often underutilized in instruction, including in the teaching of integrals. One way you might utilize it is by starting students' foray into integration with basic examples of how area can represent the product of a rate and time. This might help them have a fighting chance of understanding what all those $f(x_i^*)\Delta x_i$'s represent. Instead of using (as most textbooks do) a non-linear rate function as the context for area representing accumulation, start instead with a situation students can easily make sense of. Figure 3.3 shows a graph representing water being added to a swimming pool at a rate of 4 gallons per minute.

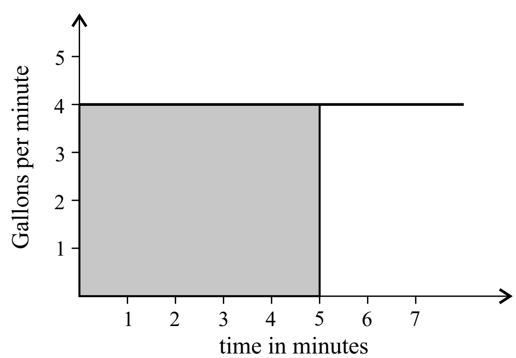


Figure 3.3. Area Model for Gallons

Working on an example like this, students will be able to figure out that the amount of water in the pool at 5 minutes is 4×5 gallons, and that the shaded area is a rectangle with area 4×5. This can set up an interesting discussion of the units. They'll know the answer is gallons and if you help them see how the area is really $4 \frac{\text{gallons}}{\text{minute}} \times 5$ minutes $= 20$ gallons, they can begin to get a sense of why the product of a rate of some quantity per unit time and time gives the accumulation of that quantity.

Increasing the level of complexity, you might have students work with non-constant (but linear) rates of change and generate expressions for the accumulation over any length of time. For example, if the rate of change of water in the swimming pool is given by $W(t) = 3t$, the accumulated water can be found by calculating the area of various triangles. For the situation involving a constant rate of 4 gallons per minute, students will likely be able to give the answer $4t$ gallons to the question, "How much water has accumulated after t minutes?" The one with a rate of $W(t) = 3t$ is just slightly more complicated but you can get students to see that the area of one of the triangles with base t will be $\frac{1}{2}t(3t) = \frac{3}{2}t^2$ gallons. Adding in an example where the rate is negative for some values (e.g., something like $W(t) = -5t + 20$) can give a context for talking about negative accumulations, negative areas, etc. in ways that may make those often challenging ideas more accessible to students. In particular, when the rate of change is negative, water is draining from the swimming pool and students can reason about the amount of water in the pool at particular points in time.

With that groundwork in place, you might get fewer blank stares when you introduce a non-linear example and generate methods for approximating the area, eventually getting to the idea of using an infinite number of rectangles. By starting with the constant rate example that can be modeled with a rectangle, the idea that the area of each of the rectangles in our Riemann sum can be found by multiplying Δx_i by $f(x_i^*)$ is apt to seem a whole lot less mysterious.

You could also use the swimming pool example to motivate the derivative-antiderivative connection between the accumulation function and the rate function. Students will recognize (sometimes even without prompting) a derivative-based relationship between $3t$ and $\frac{3}{2}t^2$ (or other examples of rate and accumulation expression pairs you generate based on area calculations). This can help set the stage before you formally arrive at the FTC.

Build a Better Foundation: Understanding Limits and Covariation

As the problems above illustrate, students need to have a solid understanding of a number of different topics before they are even in a position to understand integrals. According to the researchers cited throughout this chapter, the three most important topics are covariation (how a dependent variable changes as an independent variable changes), limits, and accumulation.

Limits are important enough to warrant their own chapter, which we hope you enjoyed already. One suggestion discussed there in greater detail is to carefully use wording that mirrors the idea of limits as approximations (as opposed to the language of motion, such as "x goes to a"). Applying that suggestion to integrals, you might describe the limit of the Riemann sums for a positive function as follows:

> You can make ***the sum of the areas of the boxes*** arbitrarily close to the ***area under the function*** by taking ***the width of the widest box*** sufficiently ***close to zero***.

If students are familiar with this general format for limits, they will be more likely to see integrals as another application of the idea of limits, using their previous knowledge of limits to help them understand this new concept. Seeing yet another example of how limits are used might help them come to the important realization that limits are central to calculus [22].

As for covariation, one might argue that the entirety of calculus falls under this idea: understanding how the changes in one variable affect the changes in another. Here we mean a more elementary concept, described by Carlson et al. as "coordinating two varying quantities while attending to the ways in which they change in relation to each other" [3, p. 354]. Integrals have numerous ways in which covariation plays a pivotal role:

- $\int_a^x f(t)\,dt$ changes as x changes.
- the Riemann sum changes as the number of subintervals changes.
- the Riemann sum changes as the points x_i^* change.
- the error of the approximation changes as the width of the subintervals changes.

Researchers generally agree that students who have a better-developed sense of covariation will be more likely to understand integrals (see, e.g., [21]).

So what types of activities help students build better understandings of covariation? Question 4 described earlier, modeling the water level in a reservoir, represents one type of question that can help. Newer calculus texts have adopted many problems like this, sometimes including scaffolding questions that ask students to work their way up through increasingly complex ideas. For instance, asking students to calculate the accumulation over a portion of the graph where the function is constant, then over a linear portion, and so on, helps students make connections between accumulation and area. After students can flexibly work with these elementary functions, they are in a better position to understand more complicated functions.

A technology-driven option for getting at the idea of covariation, suggested by Carlson's group, is to make use of motion sensors to tap into students' intuitive understanding of their own motion [4]. Asking students to walk a given position vs. time graph allows them to build on their intuitive understanding of speed and acceleration, and the immediate feedback provided by the sensors (in the form of a graph on a computer screen) lets them adjust their thinking on the fly, repeatedly and quickly refining their ideas.

A question that strikes at the core idea of both covariation and accumulation is now known as "the bottle problem" (see, e.g., [5]). The now-famous task shown in Figure 3.4 is a deceptively simple question.

Imagine this bottle filling with water. Sketch a graph of the height of the water level as a function of the volume of water that's in the bottle.

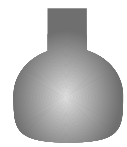


Figure 3.4. Carlson's Bottle Problem

After thinking through this problem for yourself (warning: it has tripped up some brilliant mathematicians), you might take a minute to think through how a student at each of the five levels (I through V) in the discussion of Problem 4 might respond. As with the other questions mentioned here, it's vital to coax students into generalizing their ideas so that they are not distracted by the particulars of any specific function (or in this case, bottle). The goal is not to have students understand this particular question, but to understand the underlying ideas of accumulation. An example of how to help students do this comes from Carlson and company. After having students work through a couple of specific bottles, they suggest asking them to write up an instruction manual for how to complete this task for **any** bottle.

As Thompson cautions, these types of accumulation questions differ from the Riemann integral in two important ways [28]. First, in a Riemann integral, what's being added up is the product of two quantities, making the task more complicated than the accumulated volume seen in the bottle problem. His second point is that technically, the expression $\int_a^x f(t)dt$ does not "accumulate area" in the same way that the bottle question suggests. Instead, it provides a way to approximate, using boxes, an area which changes as the value of x changes. He reports that this distinction between two different models of the definite integral (one as accumulation and one approximated by boxes) represents an important hurdle for students in their quest to understand integrals and one to watch for in your own students. (More on this in a second.)

Learning Riemann Sums

Students understand Riemann sums on many different levels. On one extreme, elementary school kids can understand the idea of filling up a curved area with smaller and smaller rectangles. On the other end, a rigorous treatment requires taking the *sup* or *inf* over the set of all possible finite partitions of the interval $[a,b]$. Most calculus students lie somewhere in between these two, and a reasonable goal is to help them move a significant distance in the right direction. Researchers suggest a number of different approaches.

Given the difficulty of calculating Riemann sums directly (go ahead and admit it: it's a real pain doing those by hand!) many authors suggest using technology to assist in the process (e.g., [23, 12]). Devices as simple as calculators or spreadsheets can allow students to see how finer approximations approximate the eventual area with smaller errors. Recent textbooks provide a wealth of problems in this vein. Carrying out these calculations can help students better understand how the various components of a Riemann sum fit together.

Other researchers have suggested using different types of questions to get at the ideas that underpin Riemann sums. Before addressing these sums explicitly, Oehrtman and Sealey had students develop the underlying ideas for themselves by working on applications like the one shown in Figure 3.5 [25].

A uniform pressure P applied across a surface area A creates a total force of $F = PA$. The density of water is 62 pounds per cubic foot, so that underwater the pressure varies according to depth, x, as $P = 62x$.

a) Draw and label a large picture of a dam 100 feet wide and extending 50 feet underwater.

b) Approximate the total force of the water exerted on this dam.

c) Find an approximation accurate to within 1000 pounds.

d) Write a formula indicating how to find an approximation with any predetermined accuracy, ϵ.

Figure 3.5. An Application Problem

The natural strategy for students working on this problem is to divide the dam up into horizontal strips, calculating the force on each one—as in a Riemann sum. After working on this problem, their students were more likely to remember and understand the details of the definition of the Riemann sum [25].

Another problem asked students to approximate the energy required to compress a spring, again asking students to make the approximation to within a specified tolerance and within any arbitrarily tolerance. In both cases, the task gets students to approximate a useful quantity by summing products of two simpler quantities—exactly what a Riemann sum does. The benefits of getting students to see applications of Riemann sums in a variety of contexts are echoed by others as well [3]. Orton suggests using similar questions, though in the more elementary setting of area, where the quantity being approximated is the area of a planar region and the two terms in each product are length and width [23].

While questions like those in Figure 3.5 may help students understand Riemann sums themselves, Thompson and Silverman argue that this approach does not directly address the main reason we study this topic: the FTC [28]. To understand how Riemann sums are used to get the FTC, they suggest that students need to see Riemann sums as varying with the right-hand endpoint of the interval. Only then will they be able to understand that complicated expression for the derivative of the integral ($\frac{d}{dx}\int_a^x f(t)\ dt$).

To help students develop these ideas, they suggest using a slightly different version of the Riemann sums, one that emphasizes the right-hand endpoint as being a variable. For calculating the area under f between 0 and x, this version would be

$$\sum_{i=1}^{\frac{x}{\Delta t}} f(i\,\Delta t)\Delta t.$$

After experimenting with various ways of asking students questions about this formula, they settled on a slightly different formulation, shown in Figure 3.6.

Although the notation here is quite complex (and ignores the issue of making sure the summation stops at an integer) this research suggests that having students explicitly use Riemann sums as accumulating functions helps them bridge the gap between seeing Riemann sums as approximating area, and seeing the integral as an accumulator of area. The goal of approaching Riemann sums from this perspective is to better prepare students to understand the FTC. Speaking

Suppose f is continuous real-valued function on (a, b) and $x > 0$. Let g be defined as

$$g(x) = \sum_{i=0}^{\frac{x}{\Delta x}} f(i\,\Delta x + a)\Delta x,$$

for $a < x < b$. Explain why g is a step function.

Figure 3.6. Riemann Sums as Accumulating Functions

of which . . .

Fundamental Theorem of Calculus

Most every suggestion so far has the goal of making the FTC understandable to students. In addition to those ideas, several researchers have suggested specific ways to work through the FTC itself to make it understandable. All three of the ideas described below ask students to understand two ideas separately (corresponding to the two sides of the equation in the FTC), and then realize that they are in fact the same.

Smith suggests a number of innovative questions, including one that asks students to consider a closed disk that grows over time [26]. This area can be viewed as accumulating (like an integral), and a series of questions lead students to see that the rate of change of this accumulation is actually equal to the circumference of the circle. Thus the derivative of the accumulation is equal to the function being accumulated, or symbolically, $\frac{d}{dx}\int_0^x f(t)dt = f(x)$.

Thompson suggests having students work through a similar set of ideas in a slightly different setting [27]. Suppose that water is filling the cylindrical tank shown in Figure 3.7. With some effort, one can calculate the volume of the water as a function of the depth of the water. This represents the accumulation of water. How does this quantity change as the depth changes? The average change can be calculated, both algebraically and geometrically, and then the instantaneous change can be found through a limiting process. This last quantity can also be realized in a different way, as the surface area of the water. Thus the derivative of the accumulation of water (with respect to the depth) is equal to the surface area on which it accumulates.

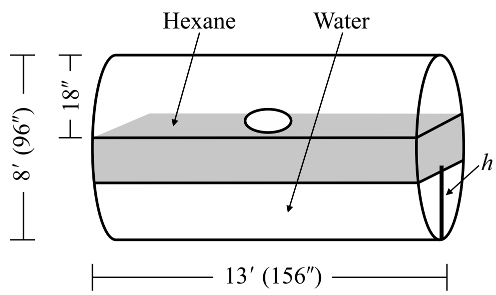


Figure 3.7. Applied Accumulation Problem

Finally, Dubinsky and his colleagues took a more technologically-based approach [9]. After teaching their students a basic programming language called ISETL, they guide students through the process of programming a handful of modules. One module calculates the accumulation of a function; another calculates the derivative. According to these researchers, having students program these modules into a computer helps them to think through these processes, creating mental images that mirror the relationships they are programming. When the differentiation module is applied to the accumulation module, the original function is magically recreated, and students have an opportunity to understand the inverse relationship of the derivative and integral that lies at the heart of the FTC.

A Good Place to Start

Make sure students have a strong understanding of an area model of multiplication (for example, a rate of change times time) which will allow them to understand integration not just as a procedure but as a meaningful counterpart to differentiation.

Bibliography

[1] Shafia Abdul-Rahman. Learning with examples and students' understanding of integration. In *Proceedings of the Eighth International Conference of Mathematics Education into the 21st Century Project on Reform, Revolution and Paradigm Shifts in Mathematics Education. Johor Bahru: UTM*, 2005.

[2] J. Bezuidenhout and A. Olivier. Students' conceptions of the integral. In *Proceedings of the Conference of the International Group for the Psychology of Mathematics Education (PME)*, volume 2, pages 73–80, 2000.

[3] M. Carlson, S. Jacobs, E. Coe, S. Larsen, and E. Hsu. Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education*, 33(5,352–378), 2002.

[4] M. Carlson, S. Larsen, and S. Jacobs. An investigation of covariational reasoning and its role in learning the concepts of limit and accumulation. In *Proceedings of the 23rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, volume 1, pages 145–153, 2001.

[5] Marilyn P. Carlson. A cross-sectional investigation of the development of the function concept. *Research in collegiate mathematics education*, 3:114–62, 1998.

[6] B. Czarnocha, S. Loch, V. Prabhu, and D. Vidakovic. The Definite Integral: A Coordination of Two Schema. In *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics, Utrecht, Nederlands*, 2001.

[7] A. Dorko. Calculus students' understanding of area and volume in non-calculus contexts. Master's thesis, The University of Maine, Orono, ME, 2011.

[8] Allison Dorko and Natasha Speer. Calculus students' understanding of volume. *Investigations in Mathematics Learning*, 6(2):48–68, 2014.

[9] E. Dubinsky. A learning theory approach to calculus. In *Symbolic Computation in Undergraduate Mathematics Education*, MAA Notes 24, pp. 43–55, 1992.

[10] European Society for Engineering Education - Maths Working Group. *Visualization, Conceptual and Procedural Knowledge in Mathematics for Engineers. The Case of the Definite Integral*, 2004.

[11] J. Ferrini-Mundy and Karan Graham. Research in Calculus Learning: Understanding of Limits, Derivatives, and Integrals. In James J. Kaput and Ed Dubinsky, editors, *Research Issues in Undergraduate Mathematics Learning: Preliminary Analysis and Results*, MAA Notes volume 33, pp. 29–45. Mathematical Association of America, Washington, DC, 1994.

[12] J.L.L. Fuster and F.J.S. Gómez. Una Interpretación de las Dificultades en el Aprendizaje del Concepto de Integral. *Divulgaciones Matemáticas*, 5(1/2):61–76, 1997.

[13] A.S. Gonzalez-Martin and M. Camacho. What is first-year Mathematics students actual knowledge about improper integrals? *International Journal of Mathematical Education in Science and Technology*, 35(1):73–89, 2004.

[14] T.A. Grundmeier, J. Hansen, and E. Sousa. An Exploration of Definition and Procedural Fluency in Integral Calculus. *PRIMUS*, 16(2):178–191, 2006.

[15] W. Hall Jr. Language and area: Barriers to student understanding of integration. Master's thesis, The University of Maine, Orono, ME, 2010.

[16] W. Hall Jr. Student misconceptions of the language of calculus: Definite and indefinite integrals. In *Proceedings of The 13th Conference on Research in Undergraduate Mathematics Education*, 2010.

[17] K. Hirst. Classifying Students' Mistakes in Calculus. In *Proceedings of the 2nd International Conference on the Teaching of Mathematics (at the undergraduate level)*, 2002.

[18] Richard Lehrer. Developing understanding of measurement. In *A Research Companion to Principles and Standards for School Mathematics*, pp. 179–192. 2003.

[19] George S. Monk. Students' understanding of functions in calculus courses. *Humanistic Mathematics Network Journal*, 9:21–27, 1994.

[20] M. Oehrtman. Collapsing Dimensions, Physical Limitation, and Other Student Metaphors for Limit Concepts. *Journal for research in mathematics education*, 40(4):396, 2009.

[21] M. Oehrtman, M. Carlson, and P. W. Thompson. Foundational reasoning abilities that promote coherence in students function understanding. In M. Carlson and C. Rasmussen, editors, *Making the Connection: Research and teaching in undergraduate mathematics education, MAA Notes volume 73*, pp. 27–42. Mathematical Association of America, Washington DC, 2008.

[22] Michael Oehrtman. Layers of abstraction: Theory and design for the instruction of limit concepts. In M. Carlson and C. Rasmussen, editors, *Making the Connection: Research and teaching in undergraduate mathematics education*, MAA Notes volume 73, pp. 65–80. Mathematical Association of America, Washington, DC, 2008.

[23] A. Orton. Students' understanding of integration. *Educational Studies in Mathematics*, 14(1):1–18, 1983.

[24] V. Prabhu and B. Czarnocha. Paradox and Proof. 2002.

[25] V. Sealey. Definite integrals, Riemann sums, and area under a curve: What is necessary and sufficient? PME-NA, 2006.

[26] Nanci Newman Smith. *Student's Emergent Conceptions of the Fundamental Theorem of Calculus*. ProQuest, 2008.

[27] P.W. Thompson. Images of rate and operational understanding of the fundamental theorem of calculus. *Educational Studies in Mathematics*, 26(2):229–274, 1994.

[28] P.W. Thompson and J. Silverman. The concept of accumulation in calculus. In M. Carlson and C. Rasmussen, editors, *Making the Connection: Research and teaching in undergraduate mathematics education*, MAA Notes volume 73, pp. 43–52. Mathematical Association of America, Washington, DC, 2008.

4

Sequences and Series

It goes to one, so it diverges.

—a possibly confused student talking with her group about a power series.

We know that students get very confused when working with sequences and series. If we wrote the way they talk, we might have opened this chapter like this:

> They have long thought of this one as harder than its predecessor, in large part because of these two. While some newer ones have rearranged them, they remain a serious challenge for most of them.

Wow, that's confusing! The previous paragraph includes many pronouns with ambiguous antecedents. When students talk about sequences and series, they exhibit similar tendencies, as in the opening quote. This student might correctly be saying that because the **terms** go to one, the **series** diverges—or maybe not. The members of her group might be just as confused as you were when you read our (thankfully fake) paragraph. Clarifying to whom or what one is referring can clear up many confusions. Here's a clarified version of that previous paragraph:

> **Students** have long thought of **Calculus II** as harder than **the first semester**, in large part because of **sequences and series**. While some newer **curricula** have rearranged **the topics**, **sequences and series** remain a serious challenge for **most students**.

Much better!

As students struggle to understand sequences and series, they will say and do things that make little sense from the perspective of rigorous mathematics. Sometimes their confusion stems from not understanding that sequences and series are different things and referring to them interchangeably—after all, their colloquial meanings are nearly identical. Having used these words as synonyms for well over a decade, calculus students naturally continue to do the same when they reach our classrooms (for additional examples of these kinds of lexically ambiguous terms and how they interact with learning, see Chapter Zero).

Although the ambiguity of the terms contributes to student difficulties, many troubles are less linguistic in nature. In the end, many students emerge from a second semester of Calculus with only a minimal (and fleeting) grasp of what sequences and series are, and no understanding of how fundamental they are to mathematics. In this chapter we give you some insights into what's going on inside your students' heads as they tackle these important topics.

Before reading any further, it might help to review several closely related sections of the book, especially the Limit chapter. If you plan to tackle sequences and series from a more theoretical perspective, the Analysis chapter might provide interesting reading as well. In this chapter, we refrain from repeating too many of the ideas covered elsewhere, concentrating instead on specific issues that arise when teaching sequences and series.

It's Just Around the Next Bend ...

1. Is there a smallest real number which is strictly greater than $\frac{1}{2}$? Explain your reasoning.

Alma: Sure, 1. Any smaller number is less than $\frac{1}{2}$.

Boris: Yes, .5000 ... 01 with an infinite number of 0s.

Cristina: No. The sequence .501, .5001, .50001, ... gets smaller than every number bigger than $\frac{1}{2}$.

Douglas: No. $\dfrac{\frac{1}{2}+x}{2}$ is always between $\frac{1}{2}$ and x, so you can keep finding numbers closer and closer to $\frac{1}{2}$.

[5, p. 413]

Problem 1 doesn't really ask anything about sequences or series, but it elicits some of the ways of thinking about the real line that cause difficulties when dealing with sequences and series. Here we see a variety of answers, several encouraging, but all of which leave us with questions about the students' thinking.

Of the answers in Problem 1, Alma's is most obviously wrong as she appears to be interpreting "real number" as "integer" or "natural number." She ignores all non-integers despite the fact that one such number is given in the problem! The implications of working with the integers instead of the reals might shock her; convergent sequences in completely disconnected spaces are rather boring. That said, she's likely to quickly see the issues with her answer when they are pointed out.

Boris's response sounds a little less naive, though still incorrect. He has some sense of a continuum of real numbers, where "greater than" might mean only a little bit greater than. His problem lies with his imprecise description of a number that is "as little as possible" greater than $\frac{1}{2}$. While his thinking is in the right ballpark, it clearly needs refining. Students responding like this typically talk in phrases that strike modern mathematicians as odd, but ironically would have been perfectly acceptable to Newton. When asked how his number differs from $\frac{1}{2}$, Boris might respond with "It's 0.0000 ... 1 bigger." This inifinitesimally small positive number is exactly the sort of beast that Newton and Leibniz used to invent the subject! In fact, Robinson's work on non-standard analysis gave their infinitesimals a rigorous foundation (see [12] for a treatment of calculus based on this work).

Boris, whose class is probably working in the world of standard analysis, needs to understand an important mathematical idea: just because you can write something down doesn't mean that the object exists. Understanding this fact will help students when they get to more complex topics, such as Cantor's proof that the cardinality of a set and its power set are never the same.

In Cristina's work we see what we might hope for. She appeals to an infinite sequence of numbers getting arbitrarily close to $\frac{1}{2}$. This shows more sophisticated thinking than Boris. Her thinking essentially boils down to the Archimedean principle, that for any positive number x there is a natural number n so that $\frac{1}{n} < x$. While Cristina's work is exactly correct, her conception of real numbers might still leave some details out. For instance, it's not clear that she would be able to refute Boris's argument, and might actually be convinced that 0.500 ... 1 represents a number that is both greater than $\frac{1}{2}$ and less than every number in her sequence.

Like Cristina, Douglas completely and correctly answers the problem. His response, however, indicates that he might not fully appreciate the power of his own method. By finding a number halfway between x and $\frac{1}{2}$, he disproves the existence of a smallest such number. Only one such counterexample is needed, but his answer hints that instead of this more advanced argument (proof by contradiction), he might be thinking along the lines of Cristina's work, envisioning an infinite sequence of numbers converging to $\frac{1}{2}$. His calculation allows him to construct such a sequence, a constructive argument that avoids the conceptual difficulties of the proof by contradiction. (See the Proofs chapter for more on this.)

Problem 2 shows another mind-bender that elicits these ways of thinking about the real line.

Number 9, Number 9 . . .

2. True or false, $.\overline{9} = 0.999999\cdots = 1$

Elida: False, it's always less than one, no matter how far out you go. So it's just . . . like infinitely close to 1.

Here we have a truly classic problem, one given to students of mathematics at many different levels. In fact, it's been used with kids in middle school, and features prominently in Burger and Starbird's popular book for liberal arts students [6], many of whom never see the beauty of calculus or the power of limits. Students who struggle with this problem are uncomfortable with the number 1 taking such an unfamiliar, yet still decimal, form. All students are aware, however, that a single number can take more than one form, for instance $\frac{1}{2} = 0.5$. In fact, most will agree that $\frac{1}{3} = 0.\overline{3} = .333333\ldots$, twisting themselves in painful contortions to avoid the inescapable conclusion about $0.\overline{9}$ when you triple both sides of this equation [20].

The distinction students implicitly make here goes back at least to Aristotle. According to his theory, there are two types of infinities: the *potentially infinite* and the *actually infinite*. The first involves a sequence of events or items which is endless. Actual infinity, on the other hand, encapsulates a potentially infinite process into something that is completed, but still consists of an infinite number of elements. Analyzing Elida's thinking through this lens, she may understand $0.\overline{9}$ in a potentially infinite manner (say, as the process of continuing to write down or say 9s), but may not be able to view that process as *actually* infinite, for instance, having written out all infinitely many 9s. Without being able to see the end, she concludes that it never stops and never reaches 1.

Similar thinking happens in some students who, in other contexts, are clearly able to conceptualize actual infinity. For instance, Tall reported that 13 of 36 calculus students correctly concluded that

$$\lim_{n\to\infty}\left(1+\frac{9}{10}+\frac{9}{100}+\frac{9}{1000}+\frac{9}{10000}+\ldots\right)=2,$$

but still thought that $0.\overline{9} < 1$ [19].

Wording the problem slightly differently produced even more uniformly confusing responses. Li and Tall [14] reported what happened when they asked two problems of college mathematics students:

I. Can you add $0.1 + 0.01 + 0.001 + 0.0001 + \cdots$ and go on forever and get an exact answer?

II. $\frac{1}{9} = 0.\overline{1}$. Is $\frac{1}{9}$ equal to $0.1 + 0.01 + 0.001 + 0.0001 + \cdots$?

Interestingly, at the end of a semester that included significant work on sequences and series, 14 of 21 students answered "no" to the first problem and "yes" to the second, nearly identical problem.

In both of these cases, students were sometimes able to conceive of actual infinity (the end of the infinite process) in one setting, but not in a slightly different, though mathematically equivalent, context. For most students, a repeating decimal evokes ideas of an infinite process and not, as we might hope, ideas of calculus and limits. Clearly, setting mathematics in a slightly different context or using slightly different wording or notation can trigger different ideas students hold. If calculus students are struggling to conceive of actual infinity, they are also likely to struggle when asked to work with sequences and series as objects, as opposed to as infinite processes. Unable to understand what's going on behind the symbols, they might settle for just moving the symbols around.

How Many Sequences Am I Holding?

3. Does this sequence a_n converge or diverge? $a_n = \begin{cases} 0\,, & \text{if } n \text{ is odd} \\ \frac{1}{2n}\,, & \text{if } n \text{ is even} \end{cases}$

Fausto: They both converge. One keeps getting closer to zero and the other is already there.

Genevieve: Yeah, but you can't **reach** the limit, so the whole thing doesn't converge.

—Adapted from [20].

In the responses to Problem 3 we see two common naive conceptions about sequences [20]. While Fausto's error may seem like a simple, almost harmless error (who cares, after all, if you call it one sequence or two—it converges either way) it reflects a deeper misunderstanding about the very nature of sequences. Like functions, sequences take a variety of forms, including algebraic and graphical representations. Many students, like Fausto, misread more complex algebraic representations, like a piecewise-defined function or this sequence. He sees two different algebraic expressions and treats them as two separate sequences.

A complete understanding of sequences requires moving beyond viewing sequences only in one form, and seeing them in a more abstract light. Only when you see a sequence as an infinite ordered list of numbers, no matter how that list is defined, can you understand the definitions, examples, and theorems in their full force. Helping students understand how the various representations are connected can help them progress toward this goal.

Unlike Fausto, Genevieve seems to see only one sequence. She might think, however, that a convergent sequence can never equal its limit. This thinking might stem from conflating the convergence of a sequence with a functional limit, transferring a restriction on the input ("x never reaches a") into one on the output. Another more likely cause stems from the fact that the language of limits is largely borrowed from the language of motion. We might say that a sequence "converges to," "goes to" or "gets closer and closer to" a value. When you "go to" the store or "get closer and closer to" an answer, it's presumed that you **don't** arrive at your goal until the very end of the action. (Imagine the absurdity of going to the store by passing by it infinitely many times, each time overshooting a little bit less!)

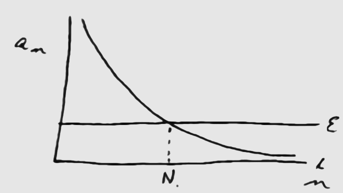
Students take this language of motion and apply it to the sequences, concluding that sequences can't reach their limit until the very end. But for a sequence "the end" is never reached, and thus a sequence never reaches its limit. (Note that such reasoning relies heavily on potential infinity.) Reinforced by a few examples that fit this pattern (after all, $\frac{1}{n}$ never reaches 0), this way of thinking becomes ingrained in students' heads, leading to responses like Genevieve's.

To be fair to students like Genevieve, however, we should note that historically, mathematicians went back and forth about whether or not a limit can be attained. D'Alembert expressed sentiments similar to Genevieve's when noting, "To speak properly, the limit never coincides, or never becomes equal to the quantity of which it is the limit, but is always approaching and can differ by as small a quantity as one desires." (Quoted in [7, p. 162].)

What's in a Definition?

4. Define what it means for a sequence a_n to converge to L.

Hernan: If $a_n \to L$, then there exists $\varepsilon > 0$ such that $|a_n - L| < \varepsilon$ for all $n \geq N$, where N is a large positive integer. [16, p. 285]



Iselle: Let $\{a_n\}$ be a sequence where $\varepsilon > 0$ and N is a positive integer. Then $|a_n - L| < \varepsilon$, $\forall n \geq N$.

Julio: Eventually the dots are all within any ε band around the limit.

Karina: $a_n \to L$ means a_n gets close to L as n gets large, but does not actually reach L until infinity. [20]

Few definitions in mathematics carry as much difficulty and importance as the definition of a limit. In the case of the sequence definition, different teachers, texts, and students approach it differently, as the four responses to Problem 4 suggest.

Hernan's picture and accompanying response contains all the right ingredients, but the ideas haven't simmered long enough to meld into a tasty dish. In the drawing, the horizontal axis appears to be labeled both l and n, whereas the limit l should really be a point on the vertical axis. His verbal description switches the correct universal quanitfier ($\forall \varepsilon$) with the existential one. That said, his half-cooked ideas are part way to a tasty, full-flavored concept.

Taken together, Iselle's algebra-based response and Julio's conceptual one complement each other quite well! Unfortunately for Julio, some applications of sequential limits require algebraic manipulation, and his answer gives no indication of his understanding of this aspect of the definition. While Julio's short sentence does reference the quantifiers (with "any" signifying the universal quantification of ε and "Eventually" indicating the existential quantification of N), Iselle's response gives no such indication that she has linked her algebraic symbols with the correct logical connectors. She may be simply parroting what she's heard in class, or, like many students, she may not yet appreciate how N is supposed to depend on ε. This confusion has also been well documented in the case of ε-δ proofs (see for instance [9, 10, 13, 17]).

Finally, Karina's answer is filled with the motion metaphors discussed above and exhibits some of the problematic ways of thinking associated with those, including the unreachability of a potential infinity.

As with many definitions, understanding the negation of the definition of limit helps elucidate the definition itself.

What's Not in a Definition?

5. What does it mean if a sequence a_n **does not** converge to L?

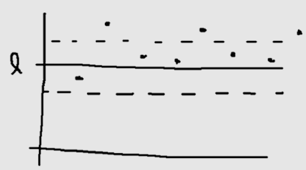
Lowell: It never gets really close to L.

Marie: It goes to infinity.

Norbert: The a_n jump around or go to infinity.

Odile: For any $\varepsilon > 0$, there exists a positive integer N s.t. $|a_n - L| > \varepsilon$ whenever $n \geq N$. [16, p. 286]

Polo: Well, for some $\varepsilon > 0$ the picture would have to look like this:



If you've taught a logic course, you know all too well that students have difficulty negating quantified statements, especially those with nested quantifiers. That's exactly what Problem 5 asks of students, though the quantifiers remain hidden in the deceptively simple question. Not surprisingly, calculus students who are just beginning to understand convergence have difficulty with this task.

The first three responses show how students with only a tentative grasp of convergence struggle to make use of their uncertain knowledge. Lowell's response, like others above, suggests a view of convergence that relies on metaphors of motion. For Lowell, $a_n \to L$ might mean that "they get really close to L." A non-convergent sequence would fail this test, or "not get close to L." Note that his imprecise, everyday language makes it difficult to decipher how he is using quantifiers.

Marie's response hints that she might conceive of all sequences as monotone (and possibly increasing). The mental image of such a sequence (maybe a sequence of dots representing the graph of a_n vs. n) probably prompted her response. Norbert uses similarly vague language, but manages to capture the two prototypical behaviors of divergent sequences. An expert mathematician might give a similar response, relying on her intuitive understanding of limits. When pressed, however, any mathematician would certainly be able to give a more formal, precise answer. We should also note that experts' use of prototypical examples can vary greatly from how students use them [1]. Norbert might only be thinking of two examples (say $(-1)^n$ and n), while an expert would see these examples as representative of entire classes of behavior.

Both Odile and Polo take more sophisticated approaches to the definition of non-convergence. It appears that Odile sees that the problem involves both the definition of sequential convergence and a negation. She correctly remembers the definition, but applies a flawed procedure in negating the statement. Failing to negate the quantifiers, she chooses instead to only negate the inequality. So while a sequence like $(-1)^n$ doesn't converge, it fails her definition of non-convergence. (Would she have given the same answer if the problem had read "divergence" instead of "non-convergence," which more obviously invokes the negation of convergence?) Like many students, she probably didn't try to make any sense of her answer after she wrote it [16].

Polo's answer indicates that at least in a graphical setting, he has a fair understanding of non-convergence. His verbal answer provides the key clue that, unlike Odile, he switched to an existential quantifier; his graph hints that he might understand the inequality $|a_n - L| > \varepsilon$ need only happen infinitely often, not for every n greater than some N.

All of these answers show how challenging this deceptively simple problem is, and how coming to a deep understanding of sequential convergence involves working through many significant ideas.

Prove It!

6. Prove that the sequence $a_n = \frac{n}{n+1}$ converges to 1.

Qadry: n and $n+1$ both go to ∞, and $\frac{\infty}{\infty}$ is 1.

Rachel: Let's see, the sequence goes $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots$. That gets arbitrarily close to 1, so it converges.

Simon: Take $\varepsilon > 0$ and let $N = \varepsilon$. Then if $n > N$, $|\frac{n}{n+1} - 1| = |\frac{n-1}{n+1}| \approx |\frac{N}{n+1}| < N = \varepsilon$.

Qadry appears to be using infinity as if it were a number. At St. Mary's College of Maryland, students and faculty have the opportunity to earn an Infinity License (see Figure 4.1) which allows them to use infinity in all the informal (and ultimately correct) ways that mathematicians do. Qadry's response shows that he isn't ready to be a responsible, licensed infinity user because he is improperly claiming that $\frac{\infty}{\infty} = 1$. In truth, infinity does possess many of the same qualities that numbers do (variables and functions can approach it, sums can "end" there, etc.). Students like Qadry are still in the process of figuring out the ways in which infinity **isn't** like real number. In this case, the fact that he arrives at the correct answer will probably make it harder to convince him that his argument contains serious logical flaws. Changing the sequence to $\frac{2n}{n+1}$ might help raise his awareness about his thinking.

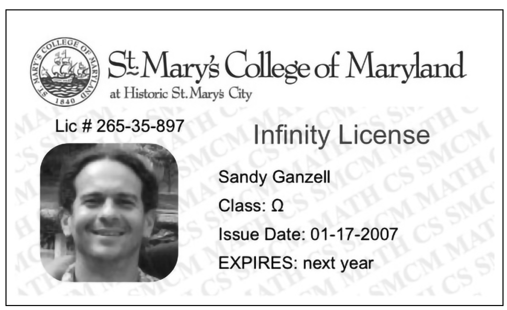


Figure 4.1. SMCM Professor Sandy Ganzell's Infinity License

Rachel's work, on the other hand, shows quite a bit of mathematical sophistication. She has correctly written down the first few terms, and draws a correct conclusion from the pattern she sees. Her phrase "arbitrarily close" suggests an awareness of the $\forall\varepsilon$ part of the definition, although she may be using that phrase in its colloquial not mathematical sense. Unfortunately, without noting that the sequence is monotone (so that it is arbitrarily close to 1 for all terms past some point), her argument remains incomplete.

Simon's work is fatally flawed from the very beginning since the choice of $N = \varepsilon$ only works for constant sequences! Still, it's possible that his intuitive understanding of the situation may have been as good as Rachel's. We can't know for sure, but seeing the word "prove," he may be choosing to tackle what he knows is the gold standard, the ε-N proof. Unable to figure out the details, he writes something down that looks plausible and hopes for some partial credit (or possibly an inattentive grader). Attempting to simply copy a template proof is a common coping strategy among college math students (see the Proofs chapter for more on this, or [11]).

Whether right or wrong, the responses to Problems 4, 5, and 6 show how asking students for information in different ways elicits different types of responses. Some teachers prefer to request informal explanations in an effort to build intuitive understanding of sequences and saving rigor for a more advanced course. Others think it vital to push students toward more formal arguments in the calculus sequence. While this decision is partly a matter of choice, you should understand the range of options, the cognitive difficulties involved in understanding something as complex as the ε-N definition, and how students' responses can hide their confusion about this topic.

What's in a Series?

7. What does it mean to say that the series $\sum a_k$ converges?

Trudy: It means the terms go to zero.

Umeka: If you add up all the terms, you get closer and closer to some number.

Vance: If you keep adding up more and more terms, and call that a new sequence, and **that** sequence converges, then the series converges.

Winnie: The sequence of partial sums converges.

As we've seen, students can already be fairly confused about sequential convergence, before we ever talk about series! Here we see several students with various degrees of sophistication in what they understand about series convergence. Before reading on, ask yourself what each of these students might and might not understand about this topic.

Trudy's incorrect answer tempts us to presume she has not understood much yet. Remembering the test for convergence (sometimes called the nth term test or divergence test), she might be mistakenly thinking that the terms "go to zero" **if and only if** the series converges. Unfortunately for her, only one of those two implications is true. While this is a fairly serious mistake, it still leaves open the possibility that she also knows the definition of convergence in some other way. Something about the way the problem was asked (or maybe her desire to write succinctly) may have prompted her to write what she thought was an equivalent formulation of series convergence.

Umeka appears to be thinking more along the lines of the correct definition, although we see in her response indications ("closer and closer") of many of the naive conceptions about sequential convergence discussed above. Her response becomes more interesting when viewed from the perspective of potential versus actual infinity. (Recall, potential infinity involves an infinite, unending process; actual infinity involves that process reaching a conclusion.) When Umeka says "add up all the terms," she hints that she may have moved beyond the potentially infinite idea of adding on more and more terms, to the actually infinite concept of completing all the infinite additions. However, the second half of her answer, talking about getting "closer and closer," indicates that she still hasn't fully moved beyond thinking of the infinite series as a process. Her response shows how difficult it can be to generalize an idea (in this case thinking of infinite processes as being completed) to a new context (here, the sequence of partial sums).

As for the last two responses, Vance seems to awkwardly get through a wordy version of the definition that you might expect from an expert; Winnie gives the expert's answer. Before we label Winnie an expert, we should note that sometimes the perfect answer can be too perfect. She might just be parroting a line her professor used, with little understanding of what it really means. Paradoxically, even though Vance's wording is jumbled and less coherent, we view it as providing more evidence that he really understands the definition.

That Depends on What "It" Means ...

8. Does the series $\displaystyle\sum_{n=1}^{\infty} \frac{2^n}{3^n}$ converge or diverge?

Xavier: Let's see. Doesn't 3^n grow a lot faster than 2^n?

Yolanda: Yeah that's right! So it goes to zero! So it converges!

In the dialogue about Problem 8, there a chance that Yolanda knows exactly what she's talking about and that Xavier understands her perfectly. More likely, these two students are mixing up sequences and series and throughout the above dialogue, aren't entirely clear which they are discussing. They might even have an entire conversation where one student is discussing the sequence and the other the series, missing the fact that they are talking past each other. These types of communication difficulties sometimes stem from the use of pronouns without clearly understood antecedents. Is Yolanda correct? That depends on what the definition of "it" is, and goes back to the ambiguous referents in the confusing opening paragraph of this chapter.

Xavier makes a correct, though very informal, analysis of the numerator and denominator as separate sequences. Yolanda takes this and draws conclusions, which are correct (and somewhat justified) if she's referring to the sequence but completely unjustified if she's referring to the series. Do they know that a sequence converging to zero does not in general imply that the related series converges? Do they see the much more subtle point that for geometric series this implication is correct? Do they understand that the converse *always* holds? The dialogue leaves us guessing. Being precise yourself, and getting your students to follow your lead, can help clarify many issues.

Many students will take some time before making the distinction between sequences and series. The challenges comes partly from the fact that these two words are actually used interchangeably in everyday English. (It doesn't much matter if Lemony Snicket has a *Sequence* or a *Series* of Unfortunate Events.) Here are the primary definitions, according to one online dictionary:

Sequence: the following of one thing after another; succession.

Series: a group or a number of related or similar things, events, etc., arranged or occurring in temporal, spatial, or other order of succession; sequence.

Add in modifiers to get "absolute convergence" (more certain than regular convergence?) and "conditional convergence" (only when it wants to?) and differentiating (pun intended) between everyday use and math-specific use becomes even more confusing. (See Chapter Zero for discussion of other lexically ambiguous terms that show up in mathematics.)

Even beyond linguistic challenges, series convergence tests force students to hold many things in their heads at once. Many of these tests rely on using a **sequence** to determine the convergence of a **series**. In the case of the "test for divergence" or "nth term test," the complexity ends with just these sequence and series. For the root test or ratio test, there are at least four different mathematical objects to be reckoned with: the original series, the sequence of terms, the sequence of partial sums, and the sequence to be tested (either $\|\frac{a_{n+1}}{a_n}\|$ or $|a_n|^{1/n}$). If the last converges, say to $\frac{1}{2}$, then the second converges to 0 and the first and third converge (though we don't know to what). That's a lot to keep in your brain if you first met sequences and series last week! As with other challenging mathematical topics, students sometimes resort to memorizing algorithms when the ideas become too overwhelming for them to cope with.

What's the Matter With Power Series?

9. When does $\sum_{n=1}^{\infty} \frac{(-x)^n}{n}$ converge?

Zeke: In other words, what is the ... what is the value of ... of x or n — what is the value of n when ... is that what it means? It's not one of these erm ... thingies, is it?
Aalia: Whaties?
Zeke: Like one of those theorems that we've been doing. You know ... it looks approximately like ... You know the, u_n and v_n,
Aalia: Yes. Oh, it might be ... wait, what does it mean for a series to converge
[Pause.]
Aalia: I get so muddled between sequences and series.
[Pause.]
Zeke: Wait, it's basically series converges, when, when you add all its terms. Is that what it means?
Aalia: Yes. Oh, there's all these little theorems—remember the integral one, and the ratio one. I think we can use those.

—[4, p. 93–94]

Ah, power series and Taylor series, the culmination of this section of calculus. The beauty of these topics and their wide applicability (especially in finding approximate solutions to differential equations) justify the need to teach this entire chapter. For a few students these topics function as an exclamation point at the end of a year of calculus; for too many, they hang like a big question mark over an otherwise understandable course.

In Problem 9 we see two students struggling as they fight through the complexities of a power series. Their conversation hints at many of the conceptual challenges students face in this section. Zeke starts out confused about what's being asked, not sure if the "when" in the problem refers to x or n. The problem's vague wording (in this case, the intentional choice of a researcher) could have been more clearly stated as "For what values of x . . . ." As it is, Zeke seems not to understand that only one reading of the problem is reasonable since n is a bound- or dummy-variable leaving x as the only free-variable.

Later in the conversation the students try to find their bearings by appealing to the definition. While only an imprecise, informal definition is given, that is quickly rejected in favor of the use of a series test. These students actually hint at three tests in this dialogue: the comparison test (the u_n, v_n "thingie"), as well as the integral and ratio tests. From a more advanced perspective, we see both of the latter two as versions of the comparison test; for instance, the ratio test implicitly makes a comparison to a geometric series. Such connections are probably not understood by most calculus students since a more theoretical perspective is typically reserved for Real Analysis. For them, these tests provide an algorithm that gets them to the answer, without really having to consider much about the series itself.

It's actually not too hard to imagine Zeke and Aalia settling on the ratio test, and eventually working through it to get a correct answer. They might do this all without ever considering that the object in question is an **infinite sum of functions of** x, and that for x's in the set they've found, that infinite sum converges . . . to another **function**. Understanding power series on that level is a long way from simply being able to chug through the ratio test!

What You Can Do

These examples have hopefully convinced you of something that most calculus students already know: sequences and series are hard. As many of these examples suggest, part of the problem is students' use of imprecise language (and the imprecise thinking behind that language.) So any way you can get students to communicate more precisely will help—hence the effectiveness of active learning methods. Beyond that, what specifically can you do to help students better understand these tough topics? Experienced teachers and mathematics education researchers have a number of suggestions for you to consider.

Language

Mathematicians have developed a precision to our language that is necessary for our work. At the same time, we borrow words that have separate everyday meanings. Students, quite understandably, have trouble distinguishing between colloquial and mathematical meanings (see also Chapter Zero, as well as the Limits, Integrals, and Proofs chapters). This linguistic confusion explains many of the difficulties students face when studying sequences and series, from the language of motion we use to talk about convergence (including the very word "convergence"), to the words "sequence" and "series" themselves, students simply don't always interpret our words the way they we intend them.

What to do about this? Say as little as possible, opening your mouth only when necessary? No, a better alternative is to be explicit about these linguistic challenges, paying close attention to the specific words you use and how they might be misinterpreted by students—and encouraging students to as well. If you do this early in the year (as we suggest in the Limits chapter), you might find yourself replacing the language of motion with the language of errors and tolerances. Instead of saying "x_n gets closer and closer to L" you might say "x_n is really close to L if n is large" or even "x_n approximates L with an arbitrarily small error." Even if you choose to avoid the rigorous ε - N definition, using the language of errors prepares students better for encountering it in a later course. If you do expect students to be rigorous, they will better understand the definition if your informal language better matches with the definition.

Of course, if all you do is change the language you are using, without paying attention to the language your students use, the battle will be lost. Listening to students while they work with each other, or while they explain problems at the board, provides an excellent opportunity for you to gently encourage precision in their language. Doing this repeatedly will eventually produce students who avoid ambiguity and as a result understand the mathematics better. If you have students working in a group, you might appoint one of them to the "Ambiguity Police," doggedly pointing out instances where their groupmates need to be clearer.

Exploring Definitions

As several of the above problems indicate, part of the difficulty of sequences and series comes in understanding the definition of convergence in each case. For something as simple as odd numbers, just giving calculus students the definition probably suffices. For something as complex as the convergence of a sequence, several authors argue that more care should be taken.

After stating the definition of a convergent sequence, a standard lecture course might continue by giving examples of convergent and divergent sequences. A number of authors suggest that this activity should be done by the students, not the teacher [8, 3, 21]. This sort of getting your hands dirty with a new definition is the type of activity done naturally by experienced mathematicians, but not by novices [2]. Giving your students the opportunity and guidance needed to explore a definition might help them build this disposition and skill for themselves.

If the same type of activity were done for series, an added twist could help students go further toward understanding some of the later material. When the student-produced examples show a pattern (every example of a convergent series has terms that converge to zero and every example of a divergent series has terms that don't converge to zero) the natural conjecture emerges. This gives the question of the harmonic series' possible convergence added intrigue! The divergence of the harmonic series then disproves a conjecture that students are invested in, instead of being just another seemingly random fact proven by their math teacher.

Other activities include exploring possible alternative definitions of convergence. For instance, in addition to the correct definition, you might propose two addition options,

$$\forall \varepsilon > 0 \; \exists N \; s.t. \; n > N \Rightarrow x_n < L + \varepsilon$$

$$\forall \varepsilon > 0 \; \exists n \;\; s.t. \; |x_n - L| < \varepsilon,$$

and ask students to find sequences that differentiate among all three—finding a sequence that satisfy one definition but not the others (if such a beast exists). Asking students to grapple with several similar logical statements that have radically different meanings prompts them to think through the relationships between the logical symbols and their meaning.

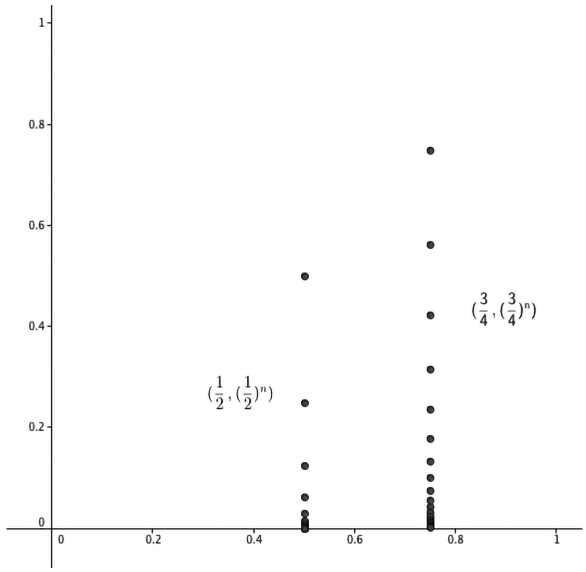
Technology-based Numerical Exploration

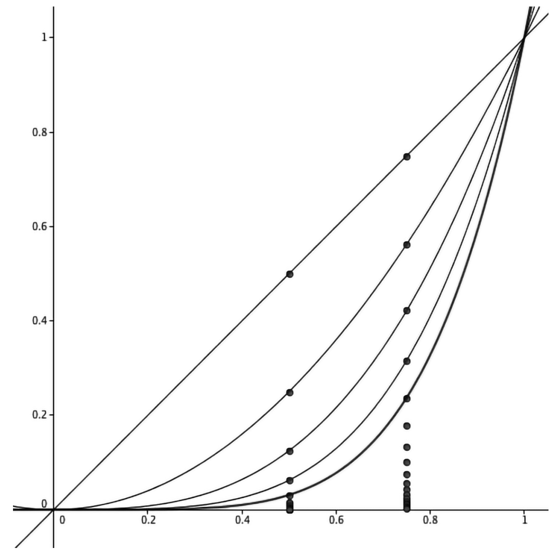
Many teachers and researchers have concluded that technology can enhance the type of explorations students engage in. Hand-held calculators and computer algebra systems allow students to numerically explore sequences and series more extensively and faster than they could with pencil and paper. Tall and Li take this idea further and suggest that students should not just use pre-packaged software, but should program the sequences and series themselves [14]. In this way, they argue, the students will be constructing mental images of sequences and series while they construct the code to calculate them.

In any case, numerically exploring the case of $\sum \frac{1}{n}$ leads to an interesting conundrum. This series diverges so slowly that the semester would long be over before anyone's computer even reached a million. In fact **every** computer would falsely indicate that, **any** series whose terms go to zero would be convergent! (The terms eventually fall below the computer's rounding error; after that, the computer just keeps adding zeroes [18].) Discussing examples like this can help students understand the need for both computation and theory.

Vertical Number Lines

To help students make the conceptual leap from sequences and series of **numbers** to sequences and series of **functions**, one research group suggests introducing vertical number lines [15]. Instead of graphing the sequence with terms stretching out to the right (showing $(n, f(n))$), pile them up along a vertical line (showing $(1, f(n))$). For instance, here's the graph of the sequences $a_n = (\frac{1}{2})^n$ and $b_n = (\frac{3}{4})^n$ strategically placed along the lines $x = \frac{1}{2}$ and $x = \frac{3}{4}$:





Then the sequence of functions becomes an exercise in connecting the dots. In this case, the vertical number lines are then connected by the graphs of the sequence of functions $f_n(x) = x^n$ (on the right).

Their research shows that when students first see sequences graphed on vertical number lines, they are more likely to make the connection that sequences of functions are really just infinitely many sequences of numbers, all stacked next to each other.

Graphical Aids

Technology can also assist in getting students to grapple with the really challenging concepts of power series and Taylor series. Even if you aren't handy with computers yourself, easy online applets show animations of how a sequence of functions can converge to another function. Letting students explore a variety of examples using these tools can help them connect their algebraic work with a graphical representation. For instance, you might ask them to use the graphs of partial sums of the Taylor series for $\ln(x + 1)$ to guess the radius of convergence before having them complete the calculation by hand.

Animated graphs also help students deal with the (sometimes skipped) topic of Taylor series' error estimates. Here even the questions become complicated for the students to understand, with five different variables to keep track of (x, n, the number of terms N, the domain of interest, and the error). You might have students first explore a relatively simple problem graphically, such as "Estimate how many terms you have to take so that the Taylor polynomial for $\cos x$ approximates the function $\cos x$ to within an error of $\frac{1}{4}$." This type of problem is actually quite easy to answer with an accurate graph. Then the exercise of verifying this result through calculation becomes more meaningful.

Of course, technology has its limitations. Students who have reached the limits of their graphing device's resolution may incorrectly conclude that a Taylor series *exactly equals* the function it approximates in some small neighborhood. In truth, this almost never happens.

A Good Place to Start

Be precise in your language, carefully distinguishing sequences, sequences of partial sums, series, and the like, which are easily confused. Thinking back to limits, try to replace the language of motion ("going to") with that of approximation ("within any arbitrary error").

Bibliography

[1] L. Alcock. Uses of Example Objects in Proving. *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, 2:17–24, 2004.

[2] L. Alcock and A. Simpson. Definitions: Dealing with Categories Mathematically. *For the Learning of Mathematics*, 22(2):28–34, 2002.

[3] ———. Two components in learning to reason using definitions. *Proceedings of the 2nd International Conference on the Teaching of Mathematics (at the Undergraduate Level)*, 2002.

[4] ———. Convergence of Sequences and Series 2: Interactions between Nonvisual Reasoning and the Learner. *Educational Studies in Mathematics*, 58(1):24, 2005.

[5] Johnston Anderson, Keith Austin, Tony Barnard, and Janet Jagger. Do third-year mathematics undergraduates know what they are supposed to know? *International Journal of Mathematical Education in Science and Technology*, 29(3):401–420, 1998.

[6] E.B. Burger and Michael Starbird. *The Heart of Mathematics*. Key College Publishing, 2000.

[7] B. Cornu. Limits. *Advanced Mathematical Thinking*, pages 153–166, 1991.

[8] R.P. Dahlberg and D.L. Housman. Facilitating Learning Events Through Example Generation. *Educational Studies in Mathematics*, 33(3):283–299, 1997.

[9] R.B. Davis and S. Vinner. The notion of limit: some seemingly unavoidable misconception stages. *Journal of Mathematical Behavior*, 5(3):281–303, 1986.

[10] E. Dubinsky and O. Yiparaki. On student understanding of AE and EA quantification. *Research in Collegiate Mathematics Education IV, American Mathematical Society, Providence, RI*, pages 239–289, 2000.

[11] G. Harel and L. Sowder. Students' proof schemes: Results from exploratory studies. *Research in Collegiate Mathematics Education*, 3:234–283, 1998.

[12] H.J. Keisler. *Elementary Calculus: an Infinitesimal Approach*. Prindle, Weber & Schmidt, Boston, Mass., 1986.

[13] J. Knapp and M. Oehrtman. Temporal order in the formal limit definition. *International Group for the Psychology of Mathematics Education*, p. 252, 2005.

[14] L. Li and D.O. Tall. Constructing different concept images of sequences and limits by programming. *Proceedings of PME*, 17:41–48, 1993.

[15] Michael Oehrtman, Craig Swinyard, and Jason Martin. Problems and solutions in students? reinvention of a definition for sequence convergence. *The Journal of Mathematical Behavior*, 33:131–148, 2014.

[16] M. Pinto and D. Tall. Student constructions of formal theory: giving and extracting meaning. *Proceedings of the 23rd Conference for the International Group for the Psychology of Mathematics Education*, 3:281–288, 1999.

[17] Kyeong Hah Roh. Students' images and their understanding of definitions of the limit of a sequence. *Educational studies in Mathematics*, 69(3):217–233, 2008.

[18] M. Schechter. Summation of Divergent Series by Computer. *American Mathematical Monthly*, 91(10):629–632, 1984.

[19] D. Tall. Cognitive conflict and the learning of mathematics. *International Group for the Psychology of Mathematics Education, Utrecht, Holland*, 1977.

[20] D. Tall and S. Vinner. Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2):151–169, 1981.

[21] A. Watson and J. Mason. Student-Generated Examples in the Learning of Mathematics. *Canadian Journal of Science, Mathematics and Technology Education*, 2(2):237–49, 2002.

5

Proof Writing

If it is a Miracle, any sort of evidence will answer, but if it is a Fact, proof is necessary.

—Mark Twain

While mathematicians generally share Mark Twain's enthusiasm for proofs, college kids tend to be a little less excited about learning to write a good mathematical argument. As one researcher put it, our students think there are two reasons to write down a proof: "to confirm something that is intuitively obvious or to verify something that is already known to be true" ([19], p. 173). The idea that mathematicians write proofs to *establish* that theorems are true and *communicate* why they are true seems foreign to many students, even to graduating senior math majors!

Getting students to understand the purpose behind proof writing is only the first step (though a large, important, and frequently skipped one) in their path to writing good proofs. Putting that issue aside, the actual writing of a proof is still notoriously difficult for students. Reasons for this abound, including a reluctance to write ("Full sentences?!? It's a *math* class!"), difficulties interpreting mathematical language (as opposed to everyday language), understanding what it means to be mathematically convincing, comprehending (sometimes nested) quantifiers, connecting important mathematical ideas, and grappling with the underlying logical structures of proofs by contrapositive, contradiction, and induction.

Whether your curriculum includes a so-called Bridge Course specifically for teaching proof writing or distributes these ideas across other courses, at some point you will find yourself helping students to write good proofs. Want to know why this can be so challenging and what you can do about it? Read on.

Proof By Example

1: Prove that $1 + 2 + 3 + 4 + \cdots + n = \frac{n(n+1)}{2}$.

Ana:

$n=3$: $1+2+3=6=\frac{3\cdot 4}{2}$ ✓

$n=5$: $1+2+3+4+5=15=\frac{5\cdot 6}{2}$ ✓

2: Prove that the difference between the squares of every two consecutive natural numbers is always an odd number, and that it is equal to the sum of these numbers. [16]

Bill:

$36-25=11 \quad 49-36=13$

$6+5=11 \quad 7+6=13$

So in general:

$A^2-B^2=A+B$

We've all seen the sort of reasoning shown in Problems 1 and 2—using a small set of examples to draw broad conclusions—from math students like Ana and Bill to politicians who argue for broad policy changes based on anecdotes. In both cases the underlying problem is that to the non-skeptical mind, a few examples can be very powerful. And to the students' (and politicians') credit, they have actually proved something; they proved the statement with a $\exists n$ in front, instead of the $\forall n$ that was implied (but not explicitly written.) Even when the universal quantifier is explicit, studies show that large numbers of students are persuaded by a few examples [10]. It's up to us to change that!

Ana and Bill produce work that shows the wide gap between what's convincing to some students and what constitutes a rigorous proof for us. As one researcher put it, "a major reason that students have serious difficulties understanding, appreciating, and producing proofs is that we, their teachers, take for granted what constitutes evidence in their eyes" [6, p. 237]. Unfortunately, what happens all too often in math courses is that instead of students changing their ideas of what is convincing to *them*, they simply learn what is convincing to *us*. By maintaining this evidence gap, they are divorcing the written proof from its main purpose: to convince an informed skeptic that something is, actually, true. Instead, the proof is an end in itself—a means to a good grade.

Proof by Symbols

3: Prove that any homogeneous system of linear equations $A\vec{x}=0$ has at least one solution. (From [6, p. 251].)

Claudette:

Set $x_1A_1+\cdots+x_nA_n=0$

Solve for x_1:

$$x_1=\frac{x_2A_2+\cdots+x_nA_n}{A_1}$$

Do the same for other x_j

If you ignore, for a minute, what the symbols in Problem 3 *actually mean*, Claudette's proof has some great attributes. She appears to understand the problem as well as its conclusion. She starts off in the right place (assuming the hypothesis) and proceeds to the conclusion (finding a solution). The path to get from one to the other, however, detours through the Land of Meaningless Symbols, where everything in sight is by default a scalar! What does she think it means to divide by one of the vector components (A_1) of the matrix A? It also might leave you with another, more disturbing question: if you had seen this work in an algebraic setting where the calculation was correct, would you have assumed that she understood the mathematics underlying the symbol manipulations?

Before we appear too shocked at such work, we should glance back at the previous mathematicians who have visited this world where symbolic manipulations are carried out without justification. Newton and Leibniz used infinitesimals to develop calculus [24] long before Robinson's non-standard analysis brought the theory behind those symbols up to the standards of rigor we use today [18]. Euler's work has long been criticized as being rife with unjustified calculations (see [12] for an excellent analysis of one such result).

What separates those phenomenal mathematicians from Claudette is (in addition to many decades and probably significant mathematical knowledge) that they were eventually proven correct! For many students who use symbols without understanding their meaning, sometimes much more mysteriously than Claudette does, the origins of their difficulties lay deep in their mathematical biographies. As discussed in Chapter Zero, some students respond to the abstract nature of algebra by learning to move symbols around in ways that teachers approve of, without understanding why those rules were chosen. This behavior can continue through college, resulting in answers like Claudette's.

And while we're talking about difficulties with symbols ...

Symbol Abuse

4: Suppose sets A and B (subsets of $\mathbb{R}$) are both bounded above. Prove that $A \cup B$ is bounded above.

Danny:

A bounded above:
$\exists M \ \forall x \in A \quad x < M$
B bounded above:
$\exists M \ \forall x \in B \quad x < M$
If $x \in A \cup B$, $x < M$ so
$A \cup B$ is bounded above $\square$

In his answer to Problem 4 Danny uses the symbol M in three different ways. First as an upper bound of A (which we might label M_A), then as an upper bound of B (M_B), and finally as an upper bound of $A \cup B$ (which might appear as $M = \max\{M_A, M_B\}$ in a correct proof).

This type of symbol abuse, using the same symbol for different variables, frequently gets mathematics students in trouble [21]. We are at least partly to blame; Danny may have sat through a Differential Equations class where the "constant" C changed from line to line! As experts, we see the difference between, for instance, relabeling $2C$ as C and assuming two sets are bounded by the same number. That distinction is actually quite subtle.

Wherever the source of Danny's understanding about symbols lies, the underlying logical problem is quite deep. We frequently omit or obscure the logical dependence of the variables we use. (If M depends on A, shouldn't we write $M(A)$?) This is both a feature and a flaw of standard mathematical notation. If we always wrote out the dependencies of variables, we would drown under a flood of parenthetical lists! (For another example of problems involving logical dependence of variables, see the ε-δ proof in the chapter on Limits.)

A closely related type of symbol misuse occurs when students assume that just because two symbols look different, they are different.

An Element By Any Other Name

5: Prove: If a commutative group has an element of order 2 and an element of order 3 then it must have an element of order 6. (From [21].)

Erika: Let g be the element of order 3 and let h be the element of order 2. Then $g^3 = e$ and $h^2 = e$ where e is the identity of the group.

Consider the subgroup generated by hg. Since $h^6 g^6 = (h^2)^3 (g^3)^2 = e^3 e^2 = e$, this subgroup is $\{hg, h^2g^2, h^3g^3, h^4g^4, h^5g^5, h^6g^6\}$ which simplifies to $\{hg, g^2, h, g, hg^2, e\}$ using $g^3 = e$ and $h^2 = e$. So hg has order 6.

In her response to Problem 5, Erika writes down six different symbols in the subgroup and concludes that the subgroup has order six—true, as long as the six symbols all represent different things. By ignoring this important point, Erika has produced an argument that only really proves the existence of a group of order 6 *or less*. There is even a chance that Erika is well aware that her proof is flawed. Especially on high-stakes exams, students often try to get as close as possible to a correct proof, knowing that their answer isn't quite right, but hoping to score maximal partial credit. Erika may or may not know that her opening assumption implies $g \neq e$, $g^2 \neq e$, and $h \neq e$. With a little work, she could also prove that $hg \neq e$ and $hg^2 \neq e$, completing the proof.

Assume Q is True . . . Hey Look! Q is true!

6: Prove: If m^2 is even then m is even. (Adapted from [22].)

Fabian: We start by assuming the hypothesis. If m^2 is even, that means that there is a number k with the property that $2k = m$. Then $m^2 = (2k)^2 = 4k^2 = 2^2k^2$. Since 2 divides the right side of this equality twice, it must also divide the left side twice. Thus 2 divides m, and m is even.

The classic type of circular reasoning shown in Problem 6 appears again and again in student work. Here if you look closely, Fabian uses m^2 being even to justify the substitution $m = 2k$, which is in fact the conclusion. If only mathematical proofs were so easy! "The conclusion of my thesis is true because I assumed it on page four. I'll take my Ph.D. now, thanks!" Unfortunately, this type of difficulty is one of the most common cited in the literature, and can be extremely difficult for students to overcome [21].

Several researchers traced the origins of these ways of thinking to the fact that working backwards can be a very powerful problem solving strategy [19]. In fact, many trigonometric identities can be proved in this way, the key point being that you start from the conclusion and reach the hypothesis through a sequence of "if and only if" statements [21]. Circular reasoning also indicates a lack of understanding of the logical structure of the statement and the proof.

A closely related problem arises when students are learning how to use induction . . .

Induction-duction What's Your Function?

7: Prove that $1^2 + 2^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$.

Frank: Let's prove this by induction. When $n = 1$ this just says that $1 = \frac{1\times2\times3}{6}$ which is obvious. Let's assume that the formula works for n . . .

Fabien: Wait, you can't do that! Isn't that assuming what you're trying to prove? I learned my lesson in Problem 6. I'm not making that mistake again!

In Problem 7 Fabien is mistaken about a very subtle part of induction proofs. He is exactly right that in a proof of the statement $\forall n \; P(n)$ it would repeat his earlier mistake to assume that every $P(n)$ is true. However, the inductive step here involves proving the statement $\forall n \; [P(n) \Rightarrow P(n+1)]$. The subtle distinction lies between assuming *all* cases of $P(n)$ and assuming a single case with the goal of proving the next one. Confusingly, both statements might be written "Assume $P(n)$."

Most textbooks address this problem, at least on the surface, by using a different letter (typically k) for the inductive hypothesis or being more explicit: "Assume that for one natural number k, $P(k)$ is true.". Along with your favorite choice of metaphor for induction (climbing an infinite staircase, an infinite string of pearls, etc.) this allows students to conceptualize n as a universally quantified variable, and k as an arbitrary, fixed natural number.

In general, students find induction to be rather mystifying. Checking the base case can strike students as a tedious, trivial task. On the other hand, if students check a few base cases, they may be convinced that the statement holds for all n. Then the induction step seems tedious and unnecessary. Those who complete an induction proof correctly often find the proof unsatisfying. For them, it may not be clear why we might want to assume the equivalence of the statements

$$\forall n \; P(n) \qquad \text{and} \qquad P(1) \text{ and } \forall k \; [P(k) \Rightarrow P(k+1)].$$

In fact, some students eventually become quite proficient at producing proofs by induction without really understanding this complex statement, much less the deeper fact that it is must be assumed as an axiom. These types of *proof frameworks* (think of the outline of an induction proof or an ε-δ proof of continuity) sometimes function as a useful way to help students structure a proof, and sometimes serve as a crutch in ways that do not support students to develop deeper thinking [23].

If This Statement is False, Then . . .

> **8:** Determine all the integers n, from 1 to 20 (including 1 and 20), that satisfy the following property:
> "If n is an even number then $n + 1$ is a prime number." (From [2, p. 7].)
>
> **Grace:** 2, 4, 6, ~~8~~, 10, 12, ~~14~~, 16, 18

While Problem 8 isn't actually a request for a proof, the implications for proof writing are hard to overstate. After all, if you don't share mathematicians' view of a simple $P \to Q$ statement, you will be hard pressed to write (or be convinced by) a good proof!

What makes this such a tricky problem? Did you notice what Grace does wrong here? (Don't feel bad if you don't—many mathematicians mess this up.) The statement given has the form $P \to Q$ but she appears to be examining a different property, one with the form P **and** Q. If P is false, the implication $P \to Q$ is true, making the odd numbers 1 through 19 part of the correct answer. Grace is far from alone in her interpretation of such statements. Researchers documented that only three out of 90 university students got this problem correct, with most answering as Grace does [2]. Similar tasks—most famously, Wason's selection task—show that students frequently do not fully understand even the simplest of conditional sentences, and that their understanding is heavily dependent on the situation (see [17] for a review).

Research indicates that these students, like Grace, view the "if . . . then" statement as *causal*, as it is used in everyday language [2]. In this interpretation, the truth of P *causes* Q to be true; if P happens to be false, the statement doesn't say anything (since something else might cause Q to be true). If Dave says, "Natasha, if you go to the party, then I'll go too," there is an implied a cause and effect relationship; Natasha's attendance will somehow cause Dave to go as well. Grace might be reasoning in this way. If $n + 1$ being prime is supposed to be caused by n being even, the statement doesn't say much about what happens when n is odd. She might argue that the statement is neither true nor false for odd values of n.

This stark difference between everyday language and mathematical usage explains many students' difficulties in understanding conditional statements [3]. When a parent says "If you eat your peas, you can have ice cream," it appears to be a statement roughly of the form

$$\text{Eat peas} \Rightarrow \text{Eat ice cream.}$$

It would take a bold (and mathematically sophisticated) kid to ignore the peas and reach for the ice cream, claiming, "You didn't say anything about what happens when I *don't* eat my peas." The parent's meaning, when translated into formal math, is closer to

$$\text{Don't eat peas} \Rightarrow \text{Don't eat ice cream.}$$

Logically, this statement is equivalent to its contrapositive

$$\text{Do eat ice cream} \Rightarrow \text{Do eat peas,}$$

which would strike us as odd in everyday language because of the reversed order of events.

The same is true of other logical constructs [3]. Compare the everyday meanings of the following statements (and the title of this section) with their mathematical meaning:

"Obey the law **or** go to jail."
"Applications will be considered **only if** they are received on time."

Negating logical statements also shows the stark differences between everyday and mathematical language. Suppose someone, probably not familiar with academia, stated, “If you work more than 40 hours a week, you are paid overtime.” Your impulse might be to respond with the everyday negation, “If I work more than 40 hours a week, I am **not** paid overtime!” (which takes the form of $P \Rightarrow \neg Q$). In the world of the mathematical “if ... then”, the correct negation would be of the form P and $\neg Q$ (“I work more than 40 hours a week and I am not paid overtime.”) This statement has a different meaning, with the hypothetical nature of the statement removed. As in the examples above, the colloquial meaning doesn’t match the meaning within the mathematics community.

The point is this: by the time students reach us, they have been using words like “if, and, or, necessary, and sufficient” for over a decade. Getting them to change how they use those words, at least in our classes, takes time, effort, and practice.

“Proof” by Contradiction?

9: Does the argument below prove the statement? Explain.

Statement: *The primes are infinite.*

Proof. Suppose on the contrary that there are only a finite number of primes, $p_1, \ldots, p_n$. Let $M = (p_1 \times p_2 \times \cdots \times p_n) + 1$. As with all natural numbers greater than 1, M must have a prime factor, say p, but p must be larger than all of the p_j (since division by each p_j leaves a remainder of 1). Thus our list of primes was incomplete and the primes must be infinite.

Henri: OK, I get each step but I don’t get why it proves that the primes are infinite. I mean, maybe you just picked the wrong list of primes. And in the end, you prove that M, one more than the product of **all** primes, can’t exist, so how can you use it to prove anything?

Proofs by contradiction (as well as the related proof by contrapositive) are notoriously difficult for students, many of whom respond as Henri does to Problem 9. Such proofs require the ability to temporarily step into a mathematical Neverland where false things are assumed to be true, (“Suppose, for the sake of argument, that P were true and Q were false.”) You can imagine the unspoken thoughts of a young math student: “Why would my teacher want me to think that Q is false when I know that in the end it’s going to be true? If we assume something that isn’t correct, how can it prove anything?”

Historically, some mathematicians have viewed proof by contradiction as different from direct proofs. Some have gone so far as to say that because proofs by contradiction are not *causal* or *explanatory* they should be rejected or at least avoided whenever possible (see [7] for a review of this debate). So while virtually all modern mathematicians have accepted such proofs as rigorous, Henri’s reaction is understandable.

But what keeps students like Henri from accepting proof by contradiction as a valid method? While this question is far from settled among researchers, two theories have been put forward. The first focuses on students’ difficulties with negating logical statements, especially those containing quantifiers [9]. Like many students, Henri might not have learned how to interpret and negate logical sentences yet [23]. Without that ability, he may be unable to follow the logical steps in proving a statement by contradiction.

The second theory concentrates more on the thought processes needed to think through such proofs, especially the need to “enter into a false, impossible world ” ([8], p. 323). Assuming that Q is false in an attempt to prove it true might be causing Henri such cognitive distress that he is unable to think past this point. Both theories offer insights into how we might make such proofs more understandable (and easier to write) for students (see below).

“Every 11 seconds a man is mugged.”

10: True or False:
i. For every positive real number a there exists a positive real number b such that $b < a$.

Isabel: Not necessarily true—it depends on where a and b are located. Who gets to pick where a and b are?

ii: There exists a positive number b such that for every positive number a, $b < a$.

Juan: Part ii) is pretty much the same as i). They’re both true—you can always find a number that’s less than another.

Both statements i) and ii) in Problem 10 involve nested quantifiers, which are notoriously difficult for students—and vital throughout real analysis. The title of this section comes from an old Saturday Night Live skit about crime rates in New York City. “Statistics show that in New York, a man is mugged every 11 seconds. I would now like you to meet that man. His name is Jesse Donnally, and he’s mugged every 11 seconds. Jesse, welcome.”[1] The heart of the matter (and the joke) is the order of the quantifiers. Does one single b work for every a that you pick (one single person getting mugged every 11 seconds), or does the b change depending on what a is (different people getting mugged at different times)?

In Problem 10, Isabel and Juan appear to have different difficulties. Isabel seems to be ignoring the quantifiers altogether, reading it simply as “Is $b < a$?” Juan sees the quantifiers but doesn’t appear to understand how the order in which they appear affects the meaning. His language does indicate that he sees (on some level) both an existential and a universal quantifier. However, he interprets both statements in the same way, as if the universal quantifier comes first. Probably neither Isabel nor Juan has grasped quantifiers enough to understand something as complex as the ε-δ definition of the limit, much less write a proof related to it.

The confusion over these two interpretations has been studied at length, leading to the short-hand $\forall\exists$ and $\exists\forall$ to represent the order of the quantifiers:

$\exists\forall$ Statement	$\forall\exists$ Statement
One man is mugged every 11 seconds.	Every 11 seconds someone is mugged.
$\exists x\ \forall y\ \ P(x, y)$	$\forall x\ \exists y\ \ P(x, y)$
$\exists b > 0\ \forall a > 0\ (b < a)$	$\forall a > 0\ \exists b > 0\ (b < a)$
Addition on the integers has an identity.	Every integer has an additive inverse.

Researchers have found that when faced with situations involving multiple quantifiers, many students will use the context (and not the specific syntax) to interpret the statement [1]. Also, like Juan, students are much more likely to interpret a given mathematical statement as an $\forall\exists$ statement than an $\exists\forall$ statement. In fact, students frequently misinterpret an $\exists\forall$ statement as a $\forall\exists$ statement, but rarely do the reverse.

How to Start

12: Prove: $A \subseteq B \Rightarrow \mathcal{P}(A) \subseteq \mathcal{P}(B)$ (Here $\mathcal{P}(C)$ stands for the power set of C.)

Kate:

Suppose $A \subseteq B$. ~~Take $x \in A$~~
~~Assume $P(A)$~~. ~~Take $x \in P(A)$~~
~~If $x \in A$ then $\{x\} \subseteq A$~~, so $\{x\} \in P(A)$.
ARGH!!!

[1]This clip gets shown every time Dave teaches his proofs class. It’s Season 9, Episode 12, 21:40 into the episode.

For many students, the hardest part of writing a proof is knowing where to start [13, 14]. Watching students in this predicament, like reading through Kate's attempts in Problem 12, can be frustrating both for teachers and for students who **do** know where to begin.

These students appear to not yet have developed an understanding of the logical structure of the statement being proven, and how that logical structure dictates the form of the proof. In this case, Kate apparently knows to assume the hypothesis of the whole statement, confidently writing "Suppose $A \subseteq B$." However, she is unable to apply that same reasoning about logical statements in a more complicated way. Here, the statement is actually of the form $(P \Rightarrow Q) \Rightarrow (R \Rightarrow S)$. But after assuming $P \Rightarrow Q$ (in this case, $A \subseteq B$, or $z \in A \Rightarrow z \in B$), she doesn't see that a proof of $R \Rightarrow S$ should begin by assuming R (in this case, $x \in \mathbb{P}(A)$). Once the logical structure is clear, most students can write out the surprisingly succinct proof, essentially:

$$\text{Assume } A \subseteq B \text{ and } x \in \mathbb{P}(A). \text{ Then } x \subseteq A, \text{ so } x \subseteq B, \text{ so } x \in \mathbb{P}(B).$$

Until students understand the logical structure of a statement (either consciously or subconsciously), they are unlikely to spontaneously develop a method for figuring out how to start a proof or even correctly validate someone else's proof [23].

What You Can Do

Working with students who are in the beginning stages of writing proofs can be both incredibly rewarding and very frustrating. A bevy of research points to ways you can structure your classes and interact with your students that give them a better chance of not only writing better proofs but also understanding why a good proof looks the way it does.

Not to ring this bell too many times, but classrooms in which students are actively engaged lead to better outcomes than passive lectures [4]. When it comes to proof writing, a collection of essays about teaching such courses even has the title *Beyond Lecture: Resources and pedagogical techniques for enhancing the teaching of proof writing across the curriculum* [20].

As for content, recent years have seen an explosion of bridge courses designed to help students make the transition from Calculus to upper-division mathematics coursework. They typically concentrate on writing proofs, sometimes even taking the name "Proofs Course," and have the goal of improving students' performance in higher-level courses. Another approach is to include much of the same content (including proofs by contradiction and induction) in a course covering a particular topic (typically Discrete Mathematics, Linear Algebra, or Number Theory.) Are these types of courses effective? What can you do to make the most of these opportunities to improve students' proof writing?

Research provides some answers to these questions. A bridge course can be an effective way to improve students' performance. Cleveland State instituted a bridge course after seeing students regularly perform poorly in their Real Analysis course [11]. This new course came in two different flavors and they were able to track the students' performance in subsequent Analysis courses. Students from one flavor of course were **twice as likely** as students in the other to pass Real Analysis. What did they do in the more effective course?

The more successful class was structured to give students opportunities to learn different types of proof techniques, one at a time, in a variety of mathematical settings [11]. Students regularly presented their work to the rest of the class and alternative proofs of a mathematical statement were encouraged. In short, the instructors provided opportunities to students to construct their own proofs, analyze and critique each others' proofs, and gradually build and refine their proof techniques.

Other research points to the importance of getting students to see proofs as necessary for verifying (and usually explaining) the truth of a statement [5, 16]. This requires a radical change of thinking for many students who, like Ana and Bill above, see examples as extremely persuasive, or who write proofs solely to satisfy teachers and earn satisfactory grades.

How do you get students to make such a radical change? After studying student proof writing for many years, Harel and Sowder suggest that the key to students learning to write good proofs is to **make proofs tangible** [27]. For a proof to be tangible to a student, it must be *concrete*, *convincing*, and *essential*. A proof is *concrete* if it deals with mathematical objects that they can handle (in the same way they handle numbers.) A proof is *convincing* if a student understands the underlying ideas—and not just the individual steps. Finally, a proof is *essential* if a student sees the need to justify the steps in the proof.

In the problems given in this chapter, the students have clearly missed out on key parts of making a proof tangible. In Problem 3, the vectors Claudette is working with are not yet concrete objects for her. Henri's reaction to the proof by contradiction argument in Problem 9 shows that for him, the proof is not convincing. And we can all imagine a bored geometry student dutifully writing "AB = BA by the symmetric property" where she feels this statement is only essential because of the threat of a poor grade, not because of an intellectual need to justify her reasoning.

Induction proofs are a great place where the standard questions lack the essential aspect. In contrast to questions like Problem 7 (the formula for the sum of the first n squares), you might have students prove Binet's formula for the Fibonacci numbers:

$$F_n = \frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\sqrt{5}}.$$

Students' natural skepticism about the formula ("How is that even an integer!?!") makes them question each step in the proof, looking for the flaw that they incorrectly believe must be hidden. Of course, this brings up the convincing aspect. Even after working through a complete induction proof of Binet's formula, many students won't be convinced of its truth, leading to a productive conversation about whether or not proof by induction is valid and maybe even the fact that induction is an axiom that must be assumed.

Several researchers have reported successes in courses that (to one degree or another) implement the suggestions above. Smith compared two different introductory Number Theory courses which served as bridge courses, one a problem-based "modified Moore-method" course, which might now be called an Inquiry Based Learning (IBL) course, where the students presented, critiqued, and reworked proofs in class, and the other a more traditional (TRAD) lecture-based course [25]. Students in IBL were more active in the classroom and approached proof writing and validating with more attention to the meaning of the mathematical statements than those in the TRAD course. In addition, when presented with a statement to prove, the TRAD students were more likely to try to follow a predetermined format (e.g., proof by contradiction) without attending to the mathematical meaning of the statement.

Years of teaching and studying student proof writing have convinced other researchers of the value of having students validate proofs (both good and bad) to help them to develop their own understanding of what a good proof is [21, 22]. In one study, researchers showed that even senior mathematics majors could be fooled by a good sounding—but logically circular—proof. They argue that students should spend significant time determining the validity of proofs, an activity done in many of the courses mentioned above.

Finally, many experienced teachers suggest that getting students to write *in complete sentences* can improve their understanding of individual proofs and of the enterprise of proof writing more generally. While we don't know of mathematics education research that backs up this claim, most of the mathematicians we've met have had the experience of coming to understand their own theorems better while writing up their results. Certainly it is more difficult for us to see students' possible misunderstandings behind disconnected strings of symbols than in complete sentences.

Researchers have also identified ways of helping students develop strong understanding while avoiding several of the specific challenges and errors mentioned above.

∃∀ vs. ∀∃

Using games to introduce ∃∀ and ∀∃ quantified statements can improve students' understanding of mathematical versions of those statements [1]. For instance, the statement $\forall a > 0\ \exists b > 0\ (0 < b < a)$ could be viewed as a two-player game, where the first player chooses a and the second player then chooses b. Can the b player always win (i.e., make $0 < b < a$)? Then the entire mathematical statement is true. Can the a player find a cleverly chosen number so that the b player can't succeed? Then the statement is false (just to be clear, the statement here is true, not false—player a cannot win). Here the order of the quantifiers, a significant stumbling block for students, is naturally modeled by the players taking turns; they quickly understand that this is a different game from $\exists b > 0\ \forall a > 0\ (0 < b < a)$ where the b player can no longer guarantee the truth of the statement, essentially because she has to choose first. Non-mathematical games, isomorphic to this one, can provide students the necessary insight to understand ∀∃ and ∃∀ statements.

Proof by Contradiction

Research suggests two possible instructional strategies for heading off students' frustration to proofs by contradiction. The first is based on the fact that students have significant difficulty in *unpacking* the logical structure of mathematical statements [23]. To get students to understand proofs by contradiction, it is necessary to first have students understand the equivalence of a statement and its contrapositive, and to be able to translate back and forth between the two.

A slightly different strategy is based on minimizing the amount of time students have to spend thinking in the Neverland where P is true and Q is false [8]. Thus instead of the standard proof of the infinitude of the primes given above, a proof might have two steps:

Lemma: Given any list of primes, $p_1, p_2, \ldots, p_n$, we can find a new prime p which is not in the list.

Proof: Let $M = p_1 \times p_2 \times \cdots \times p_n + 1$. Then dividing M by any p_j leaves a remainder of 1, so the prime factorization of M (which is greater than 1) must contain only primes not in the given list. Let p be one of those primes.

Theorem: The primes are infinite.

Proof: Could they be finite? No! If they were, there would be some number of them, and we could use our lemma to find an additional one.

By putting the bulk of the argument in a logically straightforward, constructive lemma, students only have to face the impossible case where the primes are finite for a brief moment. After students are more used to arguments written in this way, Leron [8] argues that they will better understand more complicated proofs by contradiction.

Where to Start?

While research has identified not knowing where to start a proof as a frequent issue for students, to our knowledge, no teaching methods have been well-tested. Some authors offer their own suggestions, including having students translate back and forth between verbal mathematical statements and those in predicate logic [23] and repeatedly referring back to those logical structures in your own proofs [3]. Others suggest encouraging students to work backwards from the conclusion [26].

Our own personal favorite activity is to have students suggest different ways to start a proof, and then get them to discuss the pros and cons of the five or ten most common suggestions. This activity echoes another suggestion from the research: to teach students some of the common proof frameworks as a way to help them organize their thoughts. By arguing about which starting sentences are likely to lead to successful proofs, they will begin to recognize how the logical structure of a statement determines which proof frameworks might be useful.

Finally, a number of researchers have looked at the stark differences between the thought processes of successful and flailing proof writers. One reports that when proving some theorems, professors have a sense of the proof as centered around a simple, key idea, with the proof flowing as a generalization of that idea [15]. For instance, when asked to prove that the derivative of an even function is odd, one professor said, "Let's see, an even function. There is only one thing about it, and that is its graph is reflected across the axis. Yeah, and you can be quite convinced that it is true by looking at the picture. If you said enough words about the picture, you'd have a proof." In contrast, students took very procedural approaches, not seeing any key ideas within the problem that could form the basis for a proof.

Another researcher took this idea a step further, looking at the specific key ideas needed to complete proofs and concluding that many times undergraduates knew the *facts* needed to prove a statement but lacked the *strategic knowledge* needed to put those facts together into a proof [28]. By emphasizing the key ideas within proofs and the thinking that goes into how those key ideas are put together, students might be more likely to view proofs as elaborations of those ideas—and begin to produce better proofs on their own.

A Good Place to Start

Get your students to read, analyze, critique, and correct proofs that are intentionally flawed. By working to improve others' proofs, they will build the skills needed to improve their own proofs before they make it to your grading stack.

Bibliography

[1] E. Dubinsky and O. Yiparaki. On student understanding of AE and EA quantification. *Research in Collegiate Mathematics Education IV, American Mathematical Society, Providence, RI*, pages 239–289, 2000.

[2] V. Durand-Guerrier. Which Notion of Implication is the Right One? From logical considerations to a didactic perspective. *Educational Studies in Mathematics*, 53(1):5–34, 2003.

[3] S.S. Epp. The Role of Logic in Teaching Proof. *The American Mathematical Monthly*, 110(10):886–899, 2003.

[4] Scott Freeman, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

[5] G. Harel. Two Dual Assertions: The first on learning and the second on teaching (or vice versa). *The American Mathematical Monthly*, 105(6):497–507, 1998.

[6] G. Harel and L. Sowder. Students' Proof Schemes: Results from exploratory studies. *Research in Collegiate Mathematics Education*, 3:234–283, 1998.

[7] Guershon Harel and Larry Sowder. Toward Comprehensive Perspectives on the Learning and Teaching of Proof. *Second Handbook of Research on Mathematics Teaching and Learning*, 2:805–842, 2007.

[8] U. Leron. A Direct Approach to Indirect Proofs. *Educational Studies in Mathematics*, 16(3):321–325, 1985.

[9] F. Lin, Y. Lee, and J. Wu Yu. Students' Understanding of Proof by Contradiction. *Proceedings of the 27th annual conference of PME*, 4, 2003.

[10] W.G. Martin and G. Harel. Proof Frames of Preservice Elementary Teachers. *Journal for Research in Mathematics Education*, 20(1):41–51, 1989.

[11] R.H. Marty. Getting to Eureka!: Higher Order Reasoning in Math. *College Teaching*, 39(1):3–7, 1991.

[12] M. McKinzie and C. Tuckey. Hidden Lemmas in Euler's Summation of the Reciprocals of the Squares. *Archive for History of Exact Sciences*, 51(1):29–57, 1997.

[13] R.C. Moore. *College Students' Difficulties in Learning to Do Mathematical Proofs*. PhD thesis, University of Georgia, 1990.

[14] R.C. Moore. Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3):249–266, 1994.

[15] M. Raman. Key Ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics*, 52(3):319–325, 2003.

[16] A.M. Recio and J.D. Godino. Institutional and Personal Meanings of Mathematical Proof. *Educational Studies in Mathematics*, 48(1):83–99, 2001.

[17] L.J. Rips. *The Psychology of Proof: Deductive Reasoning in Human Thinking*. MIT Press, 1994.

[18] A. Robinson. *Non-Standard Analysis*. Princeton University Press, 1996.

[19] A.H. Schoenfeld. *Mathematical Problem Solving*. Academic Press Orlando, Fla, 1985.

[20] R. Schwell, A. Steurer, and J. F. Vasquez, editors. *Beyond Lecture: Resources and pedagogical techniques for enhancing the teaching of proof writing across the curriculum*, volume 85 of *MAA Notes Series*. Mathematical Association of America, 2016.

[21] A. Selden and J. Selden. Errors and Misconceptions in College Level Theorem Proving. *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*, 3:456–471, 1987.

[22] A. Selden and J. Selden. Validations of Proofs Considered as Texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, 34(1):4–36, 2003.

[23] J. Selden and A. Selden. Unpacking the Logic of Mathematical Statements. *Educational Studies in Mathematics*, 29(2):123–151, 1995.

[24] D. Singh and SK Singh. Pre-Robinson Approach to the Theory of Infinitesimals. *International Journal of Mathematical Education in Science and Technology*, 23(2):249–255, 1992.

[25] J.C. Smith. A Sense-Making Approach to Proof: Strategies of students in traditional and problem-based number theory courses. *The Journal of Mathematical Behavior*, 25(1):73–90, 2006.

[26] Daniel Solow. How to read and do proofs. *AMC*, 10:12, 2005.

[27] L. Sowder and G. Harel. Case Studies of Mathematics Majors' Proof Understanding, Production, and Appreciation. *Canadian Journal of Science, Mathematics and Technology Education*, 3(2):251–267, 2003.

[28] K. Weber. Student Difficulty in Constructing Proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, 48(1):101–119, 2001.

6

Linear Algebra

If you've taught Linear Algebra before, you might remember a particular moment part-way through the semester. You looked out at the formerly-eager faces of your students and realized: they were no longer with you. They were lost. That instant is perfectly captured in the subtitle of a seminal paper on the teaching of linear algebra [2]:

Must the Fog Always Roll In?

Please don't give up or channel Chris Tucker's *Rush Hour II* vaguely-racist rant to Jackie Chan, "Do. you. understand. the. words. that. are. coming. out. of. my. mouth?" Instead, we suggest diving into the content to see why students struggle, and what might help them cut through the fog.

When unpacking the content of the typical Linear Algebra course, researchers have identified three different modes of reasoning students need to engage in to avoid confusion [12]:

- **Visual-geometric** (e.g., visualizing vectors and transformations)
- **Arithmetic** (e.g., matrices as tables of numbers, subject to manipulation)
- **Structural** (e.g., abstract vector spaces with linear transformations).

As you analyze your own thinking about a topic like linear independence, you might notice that you fluidly move among these three modes of thinking. Checking that three vectors are linearly independent might be a very arithmetic process, but you might be simultaneously picturing the vectors themselves (visual-geometric) or deducing that their span must have dimension of at most three (structural).

Students' thinking is rarely so crystal clear. Instead, many see the vague outlines of these same structures, but aren't sure precisely what they are looking at or what any of it means. They are lost in the fog.

Let's try to view the topics of linear algebra from their perspective, and see how we might shine some clarifying sunlight on their Linear Algebra experience.

What kind of object is that?

1: Suppose A is an $n \times n$ matrix, $\vec{v}$ is a vector in n-dimensional space, and λ is a real number. If $A\vec{v} = \lambda\vec{v}$, show that $(A - \lambda I)\vec{v} = 0$. [13]

Amanda:

$$
\begin{aligned}
Av = \lambda v &\Rightarrow Av - \lambda v = 0 \\
&\Rightarrow (A - \lambda) v = 0 \\
&\Rightarrow (A - \lambda I) v = 0
\end{aligned}
$$

Boris:

$Av = \lambda v \Rightarrow Av - \lambda v = 0$
$\Rightarrow (A - \lambda) v = 0$
$\Rightarrow (A - \lambda I) v = 0$

Eigenvalues must have the identity - otherwise can't be expressed.

Cristina:

$Av = \lambda v \Rightarrow Av - \lambda v = 0$
$\Rightarrow (A - \lambda) v = 0$
A - matrix, λ - number
$\Rightarrow A - \lambda$ doesn't work
But $\lambda v = \lambda I v$ so
$\Rightarrow (A - \lambda I) v = 0$

When an expert looks at the letters in this derivation of a key eigenvalue calculation, she sees A a (possibly large) matrix, I a matrix of the same size, and λ a constant. With these distinctions in mind, as well as knowledge of how to multiply a scalar times either a matrix or a vector, she would step through the proof, noting that $\lambda\vec{v}$ (a scalar times a vector, which results in a vector) equals $(\lambda I)\vec{v}$ (a scalar times a matrix, yielding the matrix (λI), then times the vector $\vec{v}$, all of which produces a vector). Hence the correct chain of reasoning is

$$\begin{aligned} A\vec{v} = \lambda\vec{v} \quad &\Rightarrow \quad A\vec{v} - \lambda\vec{v} = 0 \\ &\Rightarrow \quad A\vec{v} - \lambda(I\vec{v}) = 0 \\ &\Rightarrow \quad A\vec{v} - (\lambda I)\vec{v} = 0 \\ &\Rightarrow \quad (A - \lambda I)\vec{v} = 0 \end{aligned}$$

In contrast, many novices see A only as a letter without paying attention to what that letter represents [12]. When Amanda writes down $A - \lambda$, a matrix minus a scalar, she's lost in a world where letters can just be manipulated without regard to what they represent. Amanda might be viewing I as the identity (which "leaves other elements alone"), leading to the conclusion that $\lambda = \lambda I$. This equation might look reasonable to a novice, but a single number can't possibly equal an entire matrix!

The other two students improve on Amanda's work slightly. While Boris realizes the need to justify the last step, only Cristina gets close to a justified explanation.

As we saw in Chapter Zero, the concept of *variable* confuses students at many levels and this is just the linear algebra version of that issue. Making sense of each symbol while doing complex calculations and proofs is not an easy task. To understand this particular calculation, one needs to understand the difference between 1 as an identity on the (field of) real numbers and I as an identity in the ring of $n \times n$ matrices. Stop for a moment and marvel at how you, as a math expert, view the following two equations as being very different in meaning, while a novice might only see typographical differences:

$$\lambda 1 = \lambda \qquad AI = A$$

Every Linear Algebra course involves hundreds of equations like this, and to keep the fog at bay students must come to a more mathematically mature understanding of letters as variables. As students progress through the undergraduate curriculum, letters stand for an increasing number of types of objects (numbers, functions, vectors, spaces, matrices, groups, operators, ...). Asking "What kind of mathematical object does that letter stand for?" or better yet getting students used to asking and answering that question, can help them emerge from that nebulous land of meaningless symbols.

Three Views of Mount Linear Independence

2: Answer three questions about linear independence:
I: Is the set of vectors below linearly independent?

$$\{< 4, -3, 1 >, < 3, 1, -1 >, < 2, -1, 3 >\}$$

Douglas:

$$\begin{vmatrix} 4 & -3 & 1 \\ 3 & 1 & -1 \\ 2 & -1 & 3 \end{vmatrix} = \begin{matrix} 4 & -3 & 1 & 4 & -3 \\ 3 & 1 & -1 & 3 & 1 \\ 2 & -1 & 3 & 2 & -1 \end{matrix} \quad \backslash\backslash\backslash - ///$$

$$= 12+6-3-(2+4-27)$$

$$= 15-(-21) = 36 \neq 0 \text{ Yes!}$$

II: Are nonzero, mutually orthogonal vectors linearly independent? Explain [6].

Douglas: No - if they are orthogonal the direction of each depends on the directon of the others.

III: If S is the span of linearly independent vectors $a_1 \ldots a_n$ in a vector space V and b is a vector in V but not in S, are the vectors $\{b, a_1, \ldots a_n\}$ linearly independent? Explain [6].

Douglas: Sometimes-depends on the exact form of b.

Like Mount Fuji, the concept of linear independence (LI) looks different from different perspectives. In Douglas's work, we see that some of his views appear crystal clear while others remain obscured by fog.

The first part of Problem 2 evokes a very arithmetic aspect of LI, and Douglas responds appropriately, doing a quick calculation of the determinant, via a process which (for him) involves recopying the first two columns on the right side of the matrix and subtracting the "up" diagonals from the "down" ones. Possibly relying on a long list of equivalent statements in a "The Following Are Equivalent"- type of theorem, he correctly concludes that a non-zero determinant implies the LI of the row vectors.

What is Douglas thinking when he checks the determinant? That's a much harder question. How students conceptualize the determinant varies widely depending on the course, professor, and text. For different people, the determinant of an $n \times n$ matrix, $\det(A)$, can mean:

- an inductively-defined formula involving expansion of minors
- a sum of products over all permutations in S_n, the symmetric group (a.k.a. the Leibniz formula)
- the expansion (or contraction) factor for (hyper-)volumes under the linear transformation A
- the signed (hyper)-volume of a parallelepiped
- the unique n-linear function which vanishes on matrices with two identical columns and gives 1 on the identity matrix
- some crazy computation that magically solves lots of linear algebra problems.

Given Douglas's work on problems II and III, we'd venture that his conception is closest to the last one.

Looking more closely at his answers, we see two important things. His answer to part II suggests that the everyday meaning of the mathematical terms "dependent" and "independent" can inhibit a full understanding. We've seen this *lexical ambiguity* issue before with other English language terms that appear (with different meanings) in mathematics (see Chapter Zero). For Douglas, each vector "depends" on the others in a colloquial sense. While it's true that, for a set of mutually orthogonal vectors, each one affects on the others, that doesn't make them linearly dependent in the mathematical sense.

Douglas's answers also illustrate the different types of reasoning mentioned above. Part I is clearly arithmetic, while the last two straddle the worlds of visual-geometric and structural.

It's interesting to contrast Douglas's work, easily dispatching with the first part while being stumped by the latter two, with an expert's view of these questions. Most mathematicians we know disdain the calculation required for the first question, not wanting to get their hands dirty. The last two questions, in the hands of the expert, are dispatched with ease, not with arithmetic calculations or references to definitions, but with geometrical reasoning based on a deep, very graphical understanding of concepts of span, linear combinations, and orthogonality.

Helping students build such understanding on top of their arithmetic reasoning should constitute an important goal of any Linear Algebra course. The first step in achieving that aim is realizing that this type of thinking doesn't spontaneously appear in most students' heads, and helping them practice it might help burn off the fog.

Reducing Row Reduction

3: Suppose $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 3 \\ 1 & -1 & 1 \end{bmatrix}$.

I. Row-reduce A.

Elida: [*work omitted to save space*]

$$\begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix}$$

II. What is the rank (the dimension of the row space) of the resulting matrix? Why?

Elida: 2 - two non-zero rows

III. What is the rank of the original matrix A?

Elida: ???

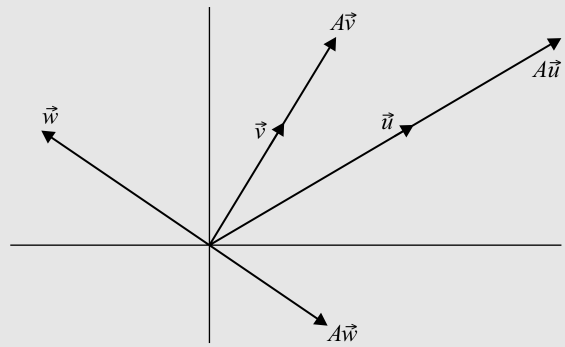
Some linear algebra students, including many who struggle, absolutely love row-reducing matrices. Generally, like Elida, they are very good at this mechanical process, sometimes also called Gaussian elimination. But do they understand the key ideas undergirding the process? Many of them, Elida included, haven't connected their procedural knowledge with their understanding of other key linear algebra concepts.

In Problem 3, Elida correctly reduces the matrix, and correctly determines that the resulting matrix, with its two non-zero rows, has rank 2. The equivalent question about the original matrix, however, stumps her. Even if she does understand what the row space is, she might not realize that row reduction doesn't change the row space. She may be able to **do** the scaler multiplication and vector addition of row-reduction, but she might not **understand** that any linear combination of the resulting matrix's rows is also a linear combination of the original matrix's rows.

Forty years ago, row-reduction formed an important and significant portion of most Linear Algebra classes. With all of the computational tools available today, most modern classes minimize (or even eliminate) hand computations of this type from the curriculum, focusing instead on the (much more important) conceptual ideas. No matter how much pencil and paper work you require, your students will be better off if you layer computational fluency over a strong foundation of conceptual understanding.

Graphical Eigenvalues

4: If A is a 2×2 matrix, explain why this picture is not possible.



Fausto: Since $A\vec{u}$, $A\vec{v}$, and $A\vec{w}$ are multiples of $\vec{u}$, $\vec{v}$, and $\vec{w}$ respectively, all three are eigenvectors. If A is a 2×2, it has a maximum of 2 independent eigenvectors.

Genevieve: You don't need that many vectors to span the plane

Hernan: It's got way too many vectors in it.

Iselle: No idea. (Adapted from [13].)

In Problem 4 we see an innovative question designed to probe students' graphical understanding of eigenvalues, a topic traditionally treated in a very computational manner. Fausto articulates an excellent answer to this question, tying together the graphical perspective given in the problem with a more structural understanding (as indicated by the "independent eigenvectors" wording). Though we wish he would have said more about the eigenvalues being different, his response suggests a fairly deep understanding of the material.

Fausto's answer, however, is far from the norm.

The students who gave the responses to Problem 4 had completed a fairly traditional linear algebra curriculum that emphasized arithmetic over visual-geometric or structural perspectives; thus, nowhere in the class did they discuss the concept that, for an eigenvector $\vec{z}$, $A\vec{z}$ points in the same direction (or the opposite direction) as $\vec{z}$. (Some computer animations beautifully illustrate this and allow students to graphically solve for eigenvectors—more on this below.) Without this graphical understanding of eigenvectors, only six of the 42 students interviewed were able to correctly argue why this picture was impossible.

The rest? Left in the fog, like Genevieve, Hernan, and Iselle, with their newfound conception of eigenvectors disconnected from any geometrical intuition. The most common answer? Iselle's "no idea."

Picture This

5: Let U be a subspace of vector space V, and let $\vec{b}$ be a vector in V but not in U. Is the set $U + \vec{b} = \{\vec{u} + \vec{b} | \vec{u} \in U\}$ a vector space? Explain your reasoning. [6]

Julio: Check the axioms:
Associativity.
$(x+b)+[(y+b)+(z+b)]$
$= \ldots$

Karina: You moved the whole thing over – it's missing 0

Lowell: They are all moved over – no way it's closed under + anymore

Here we see, in clear, bright light, the difference between knowledge that's grounded in geometric intuition (Karina and Lowell) and knowledge that isn't (Julio). It's likely that all three students would arrive at the correct conclusion: that U fails to satisfy all the requirements of a vector space. In particular, Karina is correct that since b is not the zero vector, U does not contain the 0 vector and Lowell is correct that U would fail the closure criterion (for addition). If Julio continued to check all of the axioms, he would likely come to the same conclusions, but the fact that he has decided to check all of them in order suggests that he doesn't yet own that definition in the way that Karina and Lowell appear to. Like experts, their conception of the vector space definition weaves together the axioms into a coherent and beautiful fabric.

As Fields Medal winner Terry Tao cautions his students, "Axioms are one way to think precisely, but they are not the only way, and they are certainly not always the best way" [15].

Helping students build visual-geometric reasoning takes time and effort. Using questions like this one to spur discussions can help students move from a mechanical view of axiom-checking to the more holistic view most experts use when answering such questions.

Talking about Dimensions

6: Suppose V is a 4-dimensional vector space. If G and W are subspaces of V, $\dim(U) = 2$, and $\dim(W) = 3$, is it possible for $dim(U \cap W) = 0$? Explain your reasoning.

Maria: No - there must be some overlapping between the two!

(Adpated from [6, p. 108].)

Problem 6 probes the visual-geometric and structural understanding of linear algebra and dimension, and Maria's answer clearly demonstrates an understanding of dimension firmly grounded in geometrical intuition, right? In fact, we would expect a similar answer from an expert, possibly with a more complete description of the dimensions of subspaces that contain the origin. So Maria's well on her way to the type of fog-busting understanding we seek, right?

Not so fast!

In the study that reported Maria's work, she was asked to explain what she meant [6]. In answering, she made it clear that she simply viewed the sets U, V, and W as having 2, 3 and 4 items each! Since the first two are subsets of the last, they must "overlap"—in a different sense of that word than you might have thought when doing this problem. If two of your four students are women and three of the four are math majors, the two sets have to "overlap"—you have at least one female math major. Instead of thinking about infinite planes of vectors, she's simply stating that the intersection of these two **finite sets** must be non-empty.

Rather than Maria's words serving as an example of the clear-headed thinking students are capable of, they serve as a cautionary tale that seemingly correct words can still mask serious conceptual problems. Much of an expert's intuition about linear algebra (and for that matter, much of advanced mathematics) is understood through an analogy. Unable to picture the true situation (in this case, four-dimensional space), we rely on lower-dimensional versions and describe the situation with words like "overlap." As with any analogy that connects two disparate objects, we should be careful to explicitly describe where those connections begin and where they end. In this case, an expert's idea of overlapping implicitly includes the fact the objects doing the overlapping are subspaces, not individual vectors.

Lost in (Abstract) Space

7: In the vector space of polynomials of one variable $F[x]$, is the set $\{3x^2 - 1, 2x, 7\}$ linearly independent? [6, p. 111].

Norbert: Make a linear combination:
$a(3x^2-1)+b(2x)+c(7)=0$
Hmm. How to solve for x?

Even students who have clear arithmetic and visual-geometrical understandings of linear algebra concepts can lose their way in the fog when it comes to structural understanding. In Problem 7, we see someone who might be in exactly that situation.

Norbert may have a robust graphical view of how vectors add, how the span of vectors creates a subspace embedded in the larger vector space, and how three vectors in $\mathbb{R}^n$ might be (or fail to be) linearly independent. However, when it comes to abstracting those ideas into a vector space of functions, he's lost in the fog again.

Here Norbert's issues appear to be related to ideas of variables, equality, and notation. He has correctly taken an arbitrary linear combination of the functions (with arbitrary coefficients a, b, and c), and written down the equation in question with a 0 on the right side. (The zero number? the zero vector? the zero matrix? That's not clear, but we'd put our money on him thinking of zero here as a number.) Having done so, his words suggest that he is now switching his view of the variables in question, seeing a, b, and c as the constants and x as the variable to be solved for! Instead, the question of linear independence rests on whether or not non-trivial a, b, and c exist so that the equation holds **for every** x.

As discussed in Chapter Zero, there are many meanings behind the simple symbol "$=$". When writing down an equation that checks the linear independence of three vectors in $\mathbb{R}^3$, Norbert might understand the equality as meaning "if this equation is true, what are the possible values of the constants a, b, and c—are there non-trivial ones?" However, in this more abstract setting, Norbert might interpret the equation as "for what values of x is this true?" Because the latter interpretation is more common with polynomials it's hard for students, including Norbert, to reinterpret the same symbol in a new light.

What You Can Do

In 1990, a group of college mathematics teachers (including both mathematicians and mathematics education researchers) organized the Linear Algebra Curriculum Study Group (LACSG) to "initiate substantial and sustained national interest in improving the undergraduate linear algebra curriculum" [3]. Their recommendations have provided a framework for much of the work done since then. The mathematics community produced curricula that responded to their recommendations and the mathematics education community worked to study the efficacy of their suggestions. Rather than rehashing their recommendations here, we focus on the ones reported to be most beneficial to linear algebra instructors, and on more recent work on getting students to reinvent the core linear algebra concepts.

Ground Linear Algebra in Geometrical Reasoning

The two authors of this volume had drastically different experiences as students of linear algebra. Dave's old-school course presented matrix multiplication and determinants in a formal way; Natasha's course used Banchoff's *Linear Algebra Through Geometry* text and grounded all topics in geometrical reasoning [1].

Evidence suggests the latter approach is more effective at keeping students out of the fog [11, 4], and many texts now use this method as a way to ground abstract linear algebra knowledge in students' more intuitive grasp of $\mathbb{R}^n$ (notably, [8], written by one of the members of the LACSG).

To give you a taste of this type of approach, here's a quick overview of matrix multiplication done in this manner. (Warning: If you've never thought this way, this paragraph might take you more than a few minutes to digest completely.)

> The product of a matrix A times a vector $\vec{x}$ can be viewed as a linear combination of the columns of A, where the coefficients come from the entries in $\vec{x}$. Then the first column of the matrix product AB can be thought of as a linear combination of the columns of A, with coefficients from the first column of B. A similar statement holds for each of the other columns in the product [3].

Concentrating instead on the rows, we can also view AB as being composed of row vectors, each of which is a linear combination of the rows of B, with coefficients given by the rows of A.

After weeks of thinking about matrices in this way, a successful student would, unlike Douglas in Problem 2, also see the determinant of a matrix as having meaning grounded in geometry. With multiplication by A viewed as a linear transformation, the absolute value of the determinant $|det(A)|$ would give the scale factor by which areas (or volumes, or hyper-volumes) are multiplied under the transformation; the sign would indicate whether or not the transformation preserved orientation.

Other topics can similarly be grounded in geometry. For instance, given any matrix A, most vectors $\vec{v}$ don't point in the same direction as $A\vec{v}$. The ones that do (or point in the opposite direction)? Those are the eigenvectors!

Of course you could arrive at these geometrical connections in two ways, either by developing the theoretical constructs first and then tying them to these geometric ideas, or by starting with geometry and later moving toward a more abstract understanding. The research we've seen comes down on the side of the latter path.

As for developing that geometrical intuition, research suggestions that technology can help, allowing us to engage students with things such as...

Dynamic Activities and Demos

As with differential equations, technology can greatly enhance student learning in Linear Algebra—especially when it comes to bridging the gap between graphical and arithmetic reasoning. Computers, or even graphing calculators, can also reduce the computational aspect of the course, instead saving time for the more conceptual challenges that bring in the fog.

Of course, this is far from a new idea. Gilbert Strang has been using *MATLAB* in his books since 1988 [14]. The list of linear algebra tech resources available today would fill a chapter by itself! Here we focus on what the research literature says about the effects of such tools on student learning.

If students are being asked to visualize vectors as a way of understanding linear algebra, they may not yet possess the tools to do that visualization correctly. We've seen similar issues before, with the student in the Derivatives chapter who could see the still pictures but couldn't visualize the movie. Dynamic geometry software (such as the many Java-based applets or the open source program Geogebra) appears to have benefits—more than static counterparts where the vectors and matrices are visually presented, but don't move dynamically [11, 9]. Showing students early in the semester the parallelogram that illustrates the sum of two vectors is one thing; having them change the two vectors to see how their movements affect the sum, while watching the numerical components of the vectors change dynamically, can help students connect the different styles of reasoning, resulting in knowledge that repels the fog.

More advanced tools allow students to graphically solve for eigenvectors, by depicting both the vector v and its transformed image Av and allowing the student to adjust $\vec{v}$ graphically until the two point in the same (or opposite) directions.

Whether these tools are used on-line, embedded in a graphical environment (like Geogebra), or developed by the students (in a more advanced computer algebra system like Mathematica), research suggests that they effectively help students bridge the gaps in their understanding between the computational aspects of linear algebra (e.g., solving $A - \lambda I = 0$) and the graphical aspects that are closer to the geometric intuition that drives much of the curriculum.

Asking Students for Explanations

In Maria's answer to Problem 6 (about subspaces overlapping), her words were close to perfect but her reasoning veered far off course. While her particular explanation is fairly unusual, it's extremely common for linear algebra instructors to not notice the fog their students are in. As one researcher put it, "the fact that the majority of the students passes the final examination 'helps' many teachers to remain unaware of their students' poor mathematical thinking" [16, p. 159].

Want to avoid joining the ranks of oblivious linear algebra teachers who don't notice that their classrooms have become socked in with fog?

Find out what your students are thinking! Ask them. Then ask them again in a different way.

If a particular computation isn't the part you want to test, then instead of asking them to do the dirty work, present the computation to them and then ask them about it. "Why does this computation show ..." gives your students a chance to demonstrate (or not) that they understand the meaning behind the symbols, and it gives you a chance to get inside their heads.

Of course, if the midterm exam is the first time your students are asked to justify their reasoning, the results are unlikely to be very impressive. Instead, get your students talking to each other early in the semester, discussing **why** the determinant gives you so much information, or **why** row operations don't change the linear independence of the row vectors.

Your expectation that students should **understand** all of the underlying ideas is the best way to keep the fog at bay. That expectation needs to be clear from the beginning, and needs to communicated through the questions you ask in class, on homework, on quizzes, and on exams. Then you may reach the end of the course with more of your students basking in the clear, sunny, brilliant light of linear algebra.

Inquiry-Oriented Linear Algebra (IOLA)

More recent work on the teaching and learning of linear algebra has been done by Megan Wawro and her team at Virginia Tech. The goal of the IOLA project is to develop activities that will prompt students to reinvent the core concepts of linear algebra for themselves. Their innovative sequence of problems has followed similar work done in differential equations (spearheaded by Chris Rasmussen), all of which has been guided by the Realistic Mathematics Education movement [5]. The carefully crafted sequence of problems is designed to facilitate an inquiry approach to the subject, with instructors paying close attention to students' thinking and responding to their reasoning to move the mathematical agenda forward. To give a sense of what this looks like, let's examine part of their first-day activity.

Magic Carpet Ride

> You are a young traveler with two special modes of transportation: a hover board and a magic carpet. The hover board's movement is restricted by the vector $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$, meaning that if you take the hover board for one hour, it would move along a diagonal path that would result in a displacement 3 miles east and 1 mile north of its starting location. The magic carpet has a similar restriction, but is restricted by the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.
>
> You are trying to visit Old Man Gauss, who lives in a cabin that is 107 miles east and 64 miles north of your home. Is it possible to use your two modes of transportation to get there? If so, how? If not, why is that the case? [17]

While some students approach this problem using a guess-and-check method, others figure out how to view the situation as a system of equations, and some groups invent vector notation complete with different notations for scalars, vectors, and linear combinations. An expertly guided class discussion highlights the similarities and differences among these approaches, reinforcing the power of the vector notation. This and follow-up activities lay the groundwork for an intuitive geometric understanding of the core ideas of linear independence, linear dependence, and span.

Midway through the semester, students still used the first-day activity to help them make sense of the definition of linear independence [19]. One student even contrasted the experience with a previous (apparently unsuccessful) semester of linear algebra: "Really, it goes way back to the first Magic Carpet problem. I think it's actually the first thing that enabled me to grasp it. Because I did take the class once before, and it was just definitions. I never would have thought of it as being, you have a set of vectors, can you get back to the same point? That never occurred. It's all abstract if you don't have that analogy."

The IOLA group has developed instructional sequences designed to get students to reinvent other key linear algebra concepts, including linear transformations, change of basis, and eigenvectors and eigenfunctions [19]. Their research provides evidence that with these materials, students do develop deep understanding of these key linear algebra ideas [18, 10, 7].

The IOLA materials are available at their website (`iola.math.vt.edu`).

A Good Place to Start

Work to ground your students' understanding of linear algebra concepts (vectors, matrices, determinants, eigenvalues) in their geometrical understanding of $\mathbb{R}^2$ and $\mathbb{R}^3$, which will slowly build into a geometrical understanding of $\mathbb{R}^n$.

Bibliography

[1] Thomas Banchoff. *Linear Algebra through Geometry*. Springer, 1992.

[2] David Carlson. Teaching linear algebra: Must the fog always roll in? *College Mathematics Journal*, pages 29–40, 1993.

[3] David Carlson, Charles R. Johnson, David C. Lay, and A. Duane Porter. The linear algebra curriculum study group recommendations for the first course in linear algebra. *College Mathematics Journal*, pages 41–46, 1993.

[4] Jean-Luc Dorier, Aline Robert, Jacqueline Robinet, and Marc Rogalski. On a research programme concerning the teaching and learning of linear algebra in the first-year of a french science university. *International Journal of Mathematical Education in Science and Technology*, 31(1):27–35, 2000.

[5] H. Freudenthal. *Revisiting Mathematics Education*. Kluwer Academic Publishers, 1991.

[6] Guershon Harel. The linear algebra curriculum study group recommendations: Moving beyond concept definition. In David Carlson, Charles R. Johnson, David Lay, Duane Porter, Ann Watkins, and William Watkins, editors, *Resources for Teaching Linear Algebra*, MAA Notes 42, pp. 107–126. Mathematical Association of America, 1997.

[7] Frances Henderson, Chris Rasmussen, Michelle Zandieh, Megan Wawro, and George Sweeney. Symbol sense in linear algebra: A start toward eigen theory. In *Proceedings of the 14th Annual Conference for Research in Undergraduate Mathematics Education*, 2010.

[8] David C. Lay. *Linear Algebra and its Applications*. Addison-Wesley, New York, third edition, 2006.

[9] David E. Meel and Thomas A. Hern. Tool building: Web-based linear algebra modules. *AMC*, 10:12, 2005.

[10] Chris Rasmussen, M. Trigueros, Michelle Zandieh, E. Possani Espinosa, Megan Wawro, and G. Sweeney. Building on students' current ways of reasoning to develop more formal or conventional ways of reasoning: The case of linear algebra. In *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, pp. 1577–1587, 2010.

[11] A. Sierpinska, T. Dreyfus, and J. Hillel. Evaluation of a teaching design in linear algebra: The case of linear transformations. *Recherches en Didactique des Mathématiques*, 19(1):7–40, 1999.

[12] Anna Sierpinska. On some aspects of students' thinking in linear algebra. In *On the Teaching of Linear Algebra*, pp. 209–246. Springer, 2000.

[13] Sepideh Stewart and Michael O.J. Thomas. Eigenvalues and eigenvectors: formal, symbolic and embodied thinking. In *Proceedings of the 10th Conference on Research in Undergraduate Mathematics Education*, 2007.

[14] Gilbert Strang. *Linear Algebra and its Applications*. Harcourt Brace Jovanovich, 1988.

[15] Terry Tao. Axioms of vector spaces, www.math.ucla.edu/ tao/resource/general/121.1.00s/vector˙axioms.html, (accessed April 8, 2014).

[16] Shlomo Vinner. Scenes from linear algebra classes. In David Carlson, Charles R. Johnson, David Lay, Duane Porter, Ann Watkins, and William Watkins, editors, *Resources for Teaching Linear Algebra*, MAA Notes 42, pp. 155–172. Mathematical Association of America, 1997.

[17] M. Wawro, M. Zandieh, C. Rasmussen, and C. Andrews-Larson, (2013). Inquiry oriented linear algebra: Course materials. Available at iola.math.vt.edu. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The carpet ride problem.

[18] Megan Wawro. Task design: Towards promoting a geometric conceptualization of linear transformation and change of basis. In *Twelfth Conference on Research in Undergraduate Mathematics Education, Raleigh, NC*, 2009.

[19] Megan Wawro, Chris Rasmussen, Michelle Zandieh, George Franklin Sweeney, and Christine Larson. An inquiry-oriented approach to span and linear independence: The case of the magic carpet ride sequence. *PRIMUS*, 22(8):577–599, 2012.

7

Differential Equations

Recent years have seen major changes in how mathematicians approach the teaching of Ordinary Differential Equations (ODEs). In an effort to improve students' conceptual understanding, increase the applicability of their learning, and respond to changes in the field itself, instructors have moved from courses focused on cookbook algorithms to modeling-intensive curricula; from pencil and paper exercises to computer-based assignments and activities; from quantitative approaches to qualitative ones; and from a nearly exclusive focus on closed-form solutions to (mostly linear) equations to more computational methods and non-linear equations (see [8] for a good overview). Publishers have supported these innovations, with numerous texts designed for such classes [4, 29, 5].

Research on student learning in these classes has also blossomed, the vast majority of it coming from a group of researchers led by Chris Rasmussen. In a decade-long sequence of studies, they have examined a host of issues, including

- how students' previous knowledge ties into learning ODEs [20]
- how students approach particular topics in ODEs [13, 12]
- how an inquiry-oriented approach changes the dynamics of the classroom [24, 27]
- comparisons of student performance in traditional and inquiry-oriented classes [16].

The take-home message from these studies, along with a handful of others, goes something like this: ODEs present students with significant challenges, but given the right approach, we can build on students' knowledge of calculus to improve their conceptual understanding without sacrificing any procedural competence [19].

No matter what type of class you offer, many of the issues students struggle with will be the same: What's an ODE? What's a solution to an ODE and how do I find it? How are vector fields, ODEs, solutions, and approximations all related? What exactly does existence and uniqueness mean? Some of their struggles are rooted in an incomplete understanding of what functions are. (For more on this, see the Functions section of Chapter Zero and the Derivatives chapter.) Other issues are more specific to differential equations. So what's going through students' heads as they struggle with these tough questions? Read on.

This Class is About . . .

1: What is a solution to a differential equation? (from [28].)

Andrea: You take the equation and, like, integrate it or do one of the other tricks we've learned, and get the answer.

Barry: It's what they started out with when they came up with the differential equation, and we're trying to figure out what it was

Chantal: It's something that fits the equation—you take all the derivatives and plug it into the equation and it works.

If you haven't already, take a minute to answer this question for yourself: What **is** a solution to an ODE?

One of the elements missing from all three student responses (at least explicitly) is that a solution to an ODE is a **function** or a **collection of functions**. It's hard to overstate the importance of this fact or how students' lack of clarity about it underpins many of their difficulties in this course. By the time students reach ODEs, they've had more than a decade of experience in searching for unknown **numbers** and **variables**. Understanding that in this course the unknown represents a **function** requires thinking on a new, more abstract level. Rasmussen calls this the "function-as-solution dilemma" and cites it as a major obstacle to student understanding [19]. Of course, students have already been exposed to situations where the solution is a function or collection of functions, although in our experience, students do not see the "$+\ C$" tagged on to every antiderivative problem as representing a whole family of functions (so if you teach that topic in calculus, emphasizing this point can pay dividends both then and once they reach ODEs). To understand how the dilemma about the nature of solutions shows up in differential equations, and in the student responses to Problem 1, let's take a closer look.

Andrea's response is typical of students in a traditional ODE course, learning one technique after another for a whole host of different types of ODEs. For such students, the word "solution" refers to the processes needed to answer a question, not the result itself. As you will see, this reliance on manipulation of symbols as a problem-solving technique is a recurring theme throughout this chapter. Andrea expresses her answer in such a vague way that it's hard to tell what she thinks of a "solution" to an ODE. In saying "get the answer," she probably refers not to finding a general form of all possible solutions, but simply to the end result of the process. What form that might take—a number, a variable, a function, or a family of functions—could be unclear to her. Also, this naive view of "solution" relies primarily on algebraic representations of functions, ignoring graphical or numerical representations. Such a singular focus on algebraic formulations is shared by many ODE students including Chantal [6], and possibly Barry.

When Barry refers to what "they started out with" he sounds like he is picturing some great math wizard pulling levers behind the curtain. This (possibly evil?) wizard somehow started with the function, figured out a differential equation satisfied by that function, and now Barry's job is to go backwards, figuring out what function the wizard started with. It's no wonder that such a student might take a cynical view of mathematics ("so many pointless exercises!") and math teachers ("all-knowing, vindictive torturers!"). Barry might be surprised by two important facts about ODEs. First, Newton's insight, that you can learn about unknown functions by studying the conditions their derivatives have to satisfy, was a major revelation to the mathematics community at the time [20]. Second, not only is his professor not the wizard, but most ODEs can't be solved with a closed form solution by anyone—a fact discussed with increasing frequency in ODE classes (and in the integration section in Calculus).

Chantal's answer comes closest to what we would hope for, with her emphasis on how a solution needs to satisfy the ODE in question. Still her imprecise language (what exactly does she mean by "plug it into the equation and it works"?) leaves us unsure if her thinking is broad enough to encompass a family of functions (she might say something like, "Weird, I still don't know what C is!"), or what she might think of a graphical representation of a solution. Finally, for many modern uses of ODEs, a "solution" actually means a numerical approximation of an exact solution. Chantal's answer gives us no hints about her thoughts on these types of solutions.

Compare the limitations of these student responses with your own answer and with the complexity of what mathematicians think of when dealing with solutions to ODEs. Experts' knowledge of solutions includes ideas of general and specific solutions, families of curves overlaying vector fields, graphical as well as algebraic representations, different methods of approximation, systems of equations, ODEs as special cases of PDEs, and uniqueness and existence. For linear ODEs, the list grows longer, including ideas of functions as vectors in a vector space with all the tools of linear algebra (linear independence, dimensions of solution spaces, etc.). This dizzying array of advanced mathematical topics gives some hints as to why students struggle so mightily to understand ODEs—and why instructors resort to a computational, cookbook-style curriculum.

How do students respond to a procedurally-focused course? Read on.

Separating variables . . .

2: Solve the following ODE, showing your work:

$$\frac{dy}{dt} = (y+1)t^2$$

Dean:

$$\frac{dy}{y+1} = t^2 dt \Rightarrow \ln y+1 = t^3/3 + C$$
$$\Rightarrow y+1 = Ce^{t^3/3}$$
$$\Rightarrow y = Ce^{t^3/3} - 1$$

Dean's work is representative of a recurring theme throughout this book: a correct answer doesn't imply a thorough understanding. Dean's excellent work would probably earn full (or nearly full) credit. Like many students, he leaves off the absolute values in the antiderivative of $\frac{1}{u}$ (which should be $\ln|u|$). To his credit, he correctly adopts mathematicians' convenient but confusing habit of changing the value of C from line to line. (This convention draws frustrated responses from many students: "Hey, if C is a constant, how can it keep changing?")

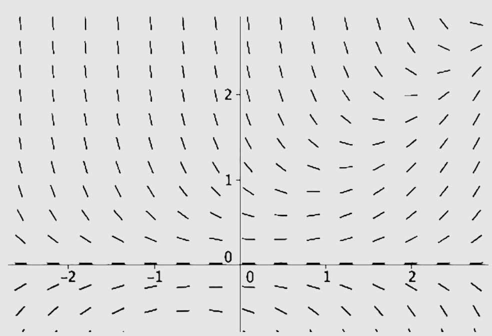
Despite his excellent application of the **method** of Separation of Variables, he might not understand the **concepts** that are carefully hidden beneath the powerful symbolic expressions he writes. In fact, it's possible that he, like the students above, might not see that his solution results in a **function**! He might simply be following a well-learned algorithm [21], with the last step being to isolate the **variable** y, not thinking of it as the function $y(t)$. (For a longer discussion of issues surrounding variables and functions, see Chapter Zero.)

This reliance on symbol manipulation might help explain his very first line, which actually has little mathematical meaning. We know what $\frac{dy}{dt}$ is, but what are dy and dt separately? (Many texts don't go into a detailed discussion of differentials, though physicists make frequent use of this notational shorthand.) A more rigorous starting point would be to integrate both sides of the given equation in the variable t, using the substitution rule (a.k.a., the chain rule) to see that $\int \frac{dy}{dt}\frac{1}{y+1}\,dt = \int \frac{1}{y+1}dy = \ln|y(t)+1| + C$. Leibniz's differentials, though enormously handy, obscure some of the more conceptually challenging ideas, making it difficult to discern if Dean understands the ideas or if he has merely learned how to manipulate the symbols.

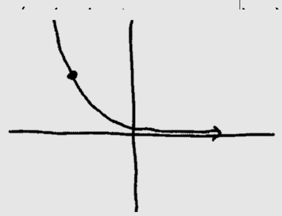
Misreading Slope and Vector Fields

3: Here is the slope field for the ODE $\frac{dy}{dt} = y(t-y)$: Sketch the solution that passes through the point $(-2, 2)$. Discuss the behavior of this solution as $t \to \infty$.

(from [19, p. 81])



Erin: As $t \to \infty$, the curve gets sucked toward the axis, so $y \to 0$.



A course in differential equations is likely to be the first place students delve deeply into slope fields. In the early part of the course, most students fairly easily begin to adapt their knowledge of slopes and derivatives to understand slope fields and solutions of ODEs [2]. For instance, when presented with four different slope fields and four ODEs and asked to match them appropriately, students will generally be able to fluidly translate between the algebraic form of the ODEs and the graphs, looking at places where the slopes are zero, negative, positive, and constant to correctly complete the task.

Here, however, we see a more subtle question that reveals where flawed student thinking regarding slope fields might arise. The beginning of Erin's solution looks good, as the function slopes downward through the third quadrant. Her solution appears to cross the y-axis with a small, positive intercept, and a small, negative slope. This presents a problem for her: should the graph of the solution curve upward and go to infinity or hug the axis and approach 0? Having seen a few problems where solutions tend asymptotically to the equilibrium solution (like e^{-x} approaching 0), she might be overgeneralizing here. Envisioning that the equilibrium solution somehow attracts all nearby solutions, she draws her solution as approaching 0, not an unusual choice. (Half of the students interviewed about Problem 3 made similar mistakes [19].)

According to Rasmussen's analysis, students making this type of error have an intuitive sense that solutions should be stable; for them, the word "stable" means that solutions that approach the equilibrium solution $y(t) = 0$ should remain in the vicinity of that horizontal line. This false idea was surprisingly strong, even in the face of contradictory evidence. One student continued to hold onto it despite being convinced that when $t = 10$, the solution had a positive y value **and** positive slope (which she estimated from the ODE itself). Another eventually convinced himself that the solution tended toward infinity, but still admitted, "I don't understand why it would want to go away from that line [meaning $y = 0$]" ([19, p. 82]). This student seemed to imagine that the solution has a mind of its own, some strange strong functional yearning to approach $y = 0$ asymptotically.

As noted in the section of Chapter Zero about graphs, when students look at visual representations of mathematics, they don't always see what we want them to see.

Equilibrium Solutions

4: Find all equilibrium solutions to the following differential equations, showing your work. (Adapted from [28].)

a) $\frac{dy}{dt} = y - 3$

Felix:

$\frac{dy}{dt} = 0$ so

$y = 3 \leftarrow$ only equilibrium point

b) $\frac{dy}{dt} = y - t$

Felix:

$\frac{dy}{dt} = 0 \Rightarrow y = t$ is the only ~~point~~ line

While equilibrium solutions play a key role in the study of ODEs, several studies revealed consistent misconceptions that result in work like Felix's. To look for equilibrium (constant) solutions, one assumes that $y(t) = c$ is a solution for some constant c and tests to find which constants work. Felix makes an equivalent assumption (that $\frac{dy}{dt} = 0$), but fails to also substitute a constant for y. Although he does the same thing in both problems, his method is only revealed as problematic in the second. His calculation actually produces an interesting result, namely the set of points where $\frac{dy}{dt} = 0$. He misses the fact that this set is sometimes not the graph of a solution (the equilibrium solutions are a subset of his set) and mistakenly claims that $y = t$ is an equilibrium solution. This should raise a big red flag: $y = t$ can't be an equilibrium solution because it's not a constant function! Not seeing this, he simply states the result, claiming an equilibrium solution where there is none. As we have seen before, Felix's reasoning seems to follow a template of steps that worked on similar problems, without understanding the limitations of the method.

Looking closer at Felix's work, his choice of words gives us more evidence that he doesn't fully understand equilibrium solutions. His answer to the first question, that the correct line $y = 3$ is an "equilibrium point" shows that he might be viewing y not as a function of t (so that $y = 3$ is a constant function, sometimes written as $y \equiv 3$ to make the distinction clearer), but instead as a variable taking on the single value of 3. For the second question, he realizes that $y = t$ isn't a point but rather a line. We might hope that this realization would prompt him to correct his wording to the first question; he doesn't, indicating that he might still be struggling with the function-as-solution dilemma [28]. In part, errors like this are a product of our ambiguous notation; we do nothing to distinguish between $y = 3$, meaning that y is the number 3, and $y = 3$, meaning y is a function whose value is always 3. Students might be helped by using $y(t) = 3$ for a function situation of this sort.

Of course, it also doesn't help that when working on a phase line, an equilibrium solution **is** correctly represented as a single point! Speaking of problems with phase lines...

Interpreting Graphs

5: a) Find all equilibrium solutions to the given ODE which models a population of insects over time. For each equilibrium solution, determine if it is stable or unstable. [19, 25]

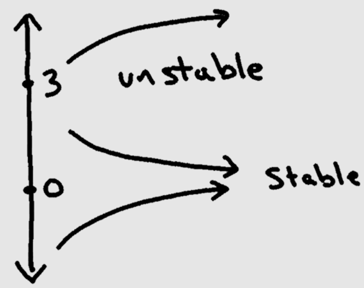
$$\frac{dN}{dt} = -4N(1 - N/3)$$

Gabielle: $\frac{dN}{dt} = 0$ @ 0 and 3

$N'(-1) = -4(-1)(1+\frac{1}{3}) > 0$

$N'(1) = -4(1)(1-\frac{1}{3}) < 0$

$N'(4) = -4(4)(1-\frac{4}{3}) > 0$



b) What is the long-term limiting population for the initial value $N(0) = 1$? Show your work.

Gabielle: Start @ (0,1) ?

$N = \int \frac{dN}{dt} = \ldots$

If I could integrate it, I would find the solution & send $t \to \infty$.

Gabielle's work shows so much promise! She starts out by correctly naming the equilibrium solutions. She then tests three points (one on either side of each equilibrium point solution) and sketches a phase line (along with a few representative solutions) to illustrate the results. Although we can't be sure, it looks like she uses this graph to correctly classify both equilibria.

And that's where she mysteriously veers off course. Given such a beautifully drawn graph, why can't she just read it to figure out that if $N(0) = 1$, $N(t)$ converges to 0 as $t \to \infty$? She practically has the solution drawn on her graph! In fact, she's already calculated the slope of such a solution at $t = 0$. What could be going through her head that would keep her from seeing the obvious?

Rasmussen interviewed four students who had similar responses to Gabielle's to shed light on what they were thinking [19]. These students viewed this type of graph as a tool to determine stability, and didn't think that the curves they sketched represented how the population changed over time. One student even noted that the equilibrium solutions could be placed in **any order** on the vertical line without affecting his analysis. The relative position of the equilibria on the N-axis only matters if you are thinking in terms of the solutions (or the related vector field), not if you are simply following a sequence of steps to determine stability. Their responses remind us again that what we see in a graph and what our students see are not always the same thing.

If you've taught math at any level, you've probably seen students follow an algorithmic procedure hoping that getting the right answer will be enough. This problem shows that algorithms with non-algebraic representations (in this case, graphical) are also susceptible to the same sort of issues.

Approximate solutions

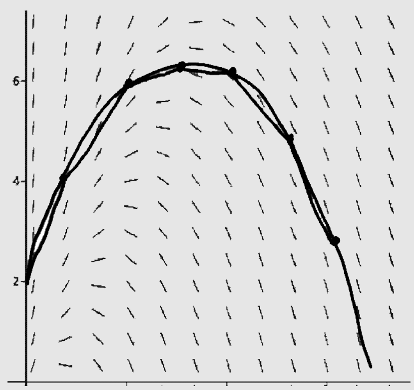
6. Consider the given vector field for the first-order differential equation

$$dy/dt = (2y - 5x + 3)/(x + 1).$$

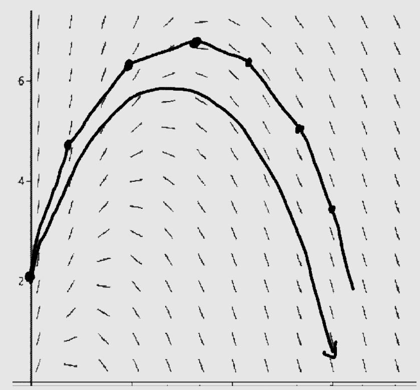
On the vector field, sketch a graph of the exact solution that satisfies the initial condition $y(0) = 2$.

Starting with $y(0) = 2$, sketch the approximation solution using Euler's method with steps of $\Delta t = 0.5$. ([19, p. 25])

Humberto:



Ingrid:



The computational simplicity of Euler's method belies its underlying conceptual challenges, as Problem 6 illustrates. Humberto's graph typifies many students' initial ideas about Euler approximations [2]. Connecting points along an exact solution, Humberto views the approximation as coming from **just that one** exact solution, as opposed to viewing each segment of the approximation as tangent to a different solution curve. This distinction shows up in the recursive formula that generates these approximations, with the derivative being evaluated at the previous point:

$$y_{n+1} = y_n + \Delta t \frac{dy}{dt}(t_n, y_n).$$

For Humberto, the approximation is just the connect-the-dots version of the exact solution. This idea suggests that the approximation might actually be dependent on knowing the exact solution, so that if you didn't know the exact solution you wouldn't be able to picture (or calculate) the approximation. Of course, it's not clear why you would want an approximation if you knew the exact solution, unless it was just to appease the evil wizard!

Humberto's work also shows that he hasn't made the appropriate connections between this graphical representation and the algebraic form shown above. We derive Euler's formula by assuming that the approximation and the exact solution have the same slope at the beginning of each time step, clearly not the case in Humberto's graph, where a small angle separates each segment from the tangent to the curve.

Ingrid's work shows more sophistication, as she apparently realizes that the approximate solution doesn't keep meeting the exact solution as it does for Humberto. However, if you look closely, each segment of the approximation has the same slope as her smooth solution does at the same t-value. So her third segment does not match the slope at its left endpoint as it should; instead, it shares the slope of the smooth solution at its left x-value. In other words, Ingrid is still using just that single exact solution to construct the approximation, instead of (correctly) seeing the approximation as tangent to a **different** exact solution at each point. For Ingrid, this allows her approximation to not stray too far away from the exact solution, something that many students think should be true [19].

A complete understanding of Euler approximations relies critically on two facts: that the solutions to differential equations come in infinite families and that the approximation is tangent to (probably a different) one of those curves at each successive point. As we've seen above, many students are still struggling with the idea that a "solution" could actually be a function, much less an (uncountably) infinite family. Picturing a collection of exact solutions, each tangent to the Euler approximation at one point, is a fairly complicated task. Faced with this difficult challenge, some students resort to just learning how to do the approximation.

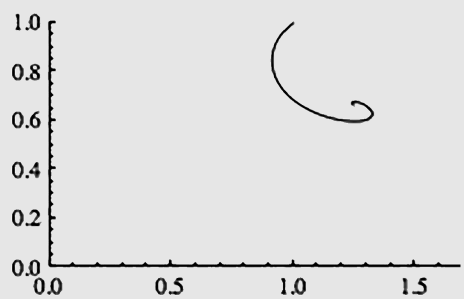
It's worth noting that in this problem we get an interesting picture of what's going on inside these students' heads because the question asked for sketches of both the exact and approximate solutions. In fact, if we had asked for just one or the other, both students might have produced perfectly acceptable graphs! As it is, both Humbrto and Ingrid's work shows some common, flawed thinking about solutions of ODEs.

Da Phase Plane! Da Phase Plane!

7: Suppose that we are modeling the populations in a predator-prey system, with $x(t)$ representing the population of prey and $y(t)$ the population of predators

$$\frac{dx}{dt} = 2x\left(1 - \frac{x}{2.5}\right) 1.5\,xy \qquad \frac{dy}{dt} = -y + 0.8\,xy.$$

The phase plane for this system is pictured here, along with one solution curve that starts at $P = (1, 1)$.



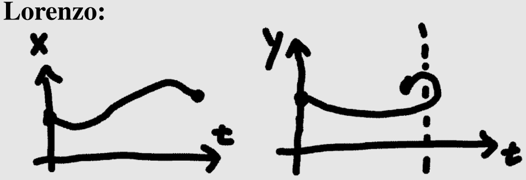
What do you think the related graphs of $x(t)$ and $y(t)$ look like for this solution? What does this mean in terms of the behavior of both populations? (From [**?**, p. 71].)

Jerry:

x	y
.9	.8
1	1
1	.6
1.3	.6

Karen: I can't remember how to solve this system, but if I could, I'd solve for both x and y and then graph them.

Lorenzo:



Not a function—fails vertical line test!

Even when students understand and can work flexibly with a slope field that describes a single ODE, phase space diagrams that describe systems of two equations give them fits. Here we see three students struggling to decipher a given phase portrait, looking to uncover graphs of the two individual populations, $x(t)$ and $y(t)$.

In fact, the phrasing of the question as a search for graphs of $x(t)$ and $y(t)$ subtly hides something that might not be apparent to Jerry: that in the phase plane, *time* is the unseen independent variable! Jerry's not-uncommon response to this task is to locate a handful of points along the solution curve and create a table of those values [25]. Not understanding that both x and y are functions of **time**, or possibly confused by the fact that the phase portrait doesn't give enough information to deduce **exactly** how x and y vary over time, he is unable to complete the task and gives up.

While we frequently simplify situations by ignoring variables (usually the time variable), this practice can be very confusing to students. If y is a function of t, how can you all of a sudden write it as just a function of x? Although they have run into this issue before, probably while dealing with parametric equations in Vector Calculus, not all students left that course with a robust understanding. A Differential Equations course provides another opportunity for them to broaden their conception of functions.

Like many college math students, Karen tends to favor working with an algebraic representation over the given graphical representation. While a preference for one representation over another is not necessarily harmful, in this case it keeps Karen from seeing that a graphical approach would be more successful. (In this case, as with many ODEs, an analytical solution is impossible.) This sort of over reliance on analytical methods may have been built up over years of math courses slanted in that direction. Research indicates that this attitude is both common and persists even after an ODE course laden with graphical and numerical methods [6].

Lorenzo appears to be the only one of the three who comes close to an appropriate solution; his graphs show many correct characteristics. One common way to approach such tasks is to pay particular attention to the points in the phase diagram with vertical or horizontal tangents, matching those points up with maxima and minima on the two solution versus time graphs [25]. Lorenzo appears to have done exactly that, with the horizontal tangents sketched in to verify that they do indeed match up with the correct places on the phase diagram. Only in the last section of Lorenzo's $y(t)$ graph does he go wrong, looping the graph back on itself in time. Back in the Derivative chapter, we saw that many students have a strong and unconscious tendency to draw a graph of the derivative that resembles the given graph. Lorenzo falls prey to this same issue while drawing $y(t)$, creating a graph which fails the vertical line test and is thus not a function.

Lorenzo's graphs have two other features worth noting. Both of his graphs end in points, presumably representing the equilibrium solution which is, in this case, an attractive point or sink. Although it is hard to glean this from the phase diagram, the solution asymptotically approaches this point without ever reaching it. By drawing solid dots, Lorenzo incorrectly indicates that both populations **reach** their limit in finite time. He may be confusing the population "stopping right there" with time somehow stopping. This error is reminiscent of Jerry's issues in understanding that the time variable is the hidden variable in the phase diagram and is ultimately another example of the function-as-solution dilemma.

Finally, the way Lorenzo positioned his two graphs makes it harder to tie both graphs back to the time variable. Lined up vertically, it is clearer why a single t would simultaneously give an x-value and a y-value, one drawn from each graph.

What You Can Do

OK, you get the point by now: Teaching students to *solve* differential equations might not be so bad, but if you want them to *understand* what's going on, you (and they) are going to have to work a whole lot harder. Now some faculty in client disciplines might still be happy if your students can just crank through problems without understanding the underlying ideas, but increasingly that's not going to fly. What's going to help them—even force them—to successfully tackle the challenging concepts packed tightly into those terse ODE problems? Here's what's worked for others.

What's the Big Idea?

In his book, *What the Best College Teachers Do*, Bain [3] compares and contrasts the teaching styles of professors who are universally regarded as excellent teachers. Among the patterns he reports is that these college teachers often structure their classes around a single, compelling question, something that the students naturally care about. The curriculum unfolds throughout the semester as an answer to that question. In contrast, less successful teachers have no such overarching theme, tending instead to provide the answers to questions that **no one has raised** ([3, p. 107]).

The examples of student thinking above don't directly address this issue of framing a course around a single question. However, a compelling theme can spark enough intellectual curiosity to overcome students' dearth of internal motivation. So, what's the single, compelling line that might provide a framework to understand the entire semester of Differential Equations? You get to choose for yourself and your students, but our money is on this paraphrase of the opening of a popular Differential Equations textbook [4]:

> This course is about how to predict the future.

Contrast this compelling, motivational, grounding statement, one that you could refer back to throughout the semester, with the implicit theme of too many math courses: "This course is about getting through Chapter 6." Nothing gets students' heart racing on the first day like the thought that they might get through Chapter 6, whatever that is!

After framing the course in this compelling way, every topic can take its rightful place as one of the many pictures in the frame. Second order linear ODEs are no longer just another unmotivated subtopic, they become tools to understand damped springs, kids on swings, and eventually electron orbits. With a clearer, more compelling goal in mind, students will be more motivated to work hard and be better able to organize their thoughts into a coherent understanding of the subject.

Focus on Concepts and Build on What They Already Know

If you want students to build a deep conceptual understanding of ODEs, then not surprisingly, your focus should be on the concepts. You can have your cake and eat it too. Research on conceptual understanding versus procedural fluency makes it clear that you can teach the former without sacrificing the latter and that too intense a focus on procedures initially can actually interfere with conceptual understanding [11].

While a conceptual focus can take many different forms, it certainly involves asking students questions that compel them to dig deeper than the standard "Classify this ODE and solve it." More conceptually focused questions force many students out of their comfort zone; as several of the questions above demonstrate, students tend to retreat to the confines of algorithmic methods whenever possible. Depending on when you learned ODEs (and who taught you), this might force you into uncharted territory as well!

What do conceptually-focused questions look like? Figure 7.1 shows the problem Rasmussen gives his students on the very first day of class [18].

A couple of points about this problem are worth noting. Even though the topic, systems of linear equations, usually comes much later in a typical course, the wording is appropriate for any student with a background in calculus, so students even on the first day of class can dive in. It also refers to a situation that students might easily imagine, using their background knowledge of ecosystems to help them make sense of these systems of ODEs. If the lesson is structured so students can work on this problem and discuss the ideas, that might serve as an excellent review of key ideas of calculus, without boring them with a review lecture. (See below for information about especially productive

> Which system of rate of change equations below describes a situation where the two species compete and which system describes cooperative species? Explain your reasoning.
>
> (A) $\dfrac{dx}{dt} = -5x + 2xy$, $\dfrac{dy}{dt} = -4y + 3xy$
>
> (B) $\dfrac{dx}{dt} = 4x - 2xy$, $\dfrac{dy}{dt} = 2y - xy$

Figure 7.1. First Day Problem

> One way to model the growth of fish in a pond is with the differential equation $\frac{dP}{dt} = kP$, with time measured in years. Use this differential equation with a growth parameter $k = 1$ to approximate the number of fish in the pond for the next several years if there are initially (a) 200 fish, (b) 400 fish, and (c) 0 fish. Record your results in tabular and graphical forms.

Figure 7.2. Euler's Method Problem

features of such activities and discussions.) What's missing in the question is also notable: There is no request to *solve* anything in the traditional sense, just to understand the model well enough to say something about it. The underlying pedagogical goal is to get students used to thinking about the meaning of ODEs, not just the algorithms needed to solve them.

Another example of a conceptually-focused question that allows students to use their previous knowledge to build a new concept comes from Rasmussen's unit on Euler's Method. Instead of the traditional exposition, where a teacher might derive the formula and present a few examples, Rasmussen asks students (who have yet to explicitly solve any ODEs) to work on questions like the one in Figure 7.2 (from [17, p. 63]).

Leveraging the powers of student collaboration, with minimal coaxing, and with help from each other, students can invent an approximation method for themselves, in part because the subtle wording of the question nudges them in that direction. These hints include asking for a prediction for "the next several years" instead of for a longer term and requesting an answer in a tabular form (as opposed to an algebraic form). Building on their work, later questions guide students to generalize their method, eventually leading to the formula that a traditional class would have started with.

In a comparison of students who went through this guided reinvention with those from a traditional course, the former group outperformed the latter when asked to calculate the first two steps of an Euler approximation [16]. Although there were many other differences between the two classes, researchers speculate that the process of reconstructing the approximation method was key in producing this performance gain.

Both of these problems illustrate how a carefully structured task, together with engagement in productive discussion, can help your students use their previous knowledge to build a solid foundation of conceptual understanding of ODEs.

Engaging Students in Productive Discussion

When you give students interesting problems that address the underlying concepts, you open up the possibility of having productive class discussions, suggested by many researchers as a way to improve student learning [23].

After students have had an opportunity to discuss a problem in small groups, one might call the class together and begin to have groups report on their progress. Carefully guiding the discussion by asking students to explain and justify their group's thinking, refine their conjectures, respond to and challenge each other's ideas, and compare and contrast different methods, a skillful facilitator can move the discussion forward toward the desired goal. The particular questions one asks, as well as the way an instructor focuses the discussion, can serve to bring out the important conceptual ideas and ultimately help students build knowledge of ODEs that is grounded in their previous knowledge of rates of change [16, 15]. This approach to discussions can prove "highly successful in both fostering students' participation in mathematical argumentation and their acquisition of important concepts and methods in differential equations" [27, p. 278].

While researchers generally agree on the potential benefits of engaging students in guided argumentation, several authors caution teachers on the challenge faced by instructors who try to adopt these methods in ODEs. One study points out that having students (as opposed to the instructor) doing the bulk of the talking, and asking students to explain and justify their reasoning involves a radical change from the social norms of the traditional mathematics classroom [27]. Students may take some time to be comfortable with classroom roles that initially seem unfamiliar or even out of place in a mathematics classroom. However, by selecting engaging tasks (see below for some suggestions) and conveying to students that you value discussion of ideas and strategies, you can help them take on these more active roles in your classroom.

Other studies followed experienced professors through the process of implementing these types of class discussions [26, 22]. They concluded that one instructor's ability to use these discussions to further his instructional goals was at times hampered by his inability to predict how students might respond to the tasks and how their responses could be channeled in productive ways. This work demonstrates that even teachers with extremely strong knowledge of ODEs need to build other types of knowledge in order to productively lead discussions. (One goal of this book is to help you build exactly these types of knowledge!)

Connecting Multiple Representations

Calculus reformers in the 1990s coined the term "Rule of Three" to refer to three different mathematical representations: algebraic, graphical, and numerical [7]. Later, verbal representation rounded out what was called the Rule of Four. Moving among representations poses significant challenges for students. A key part of a robust, nuanced understanding of differential equations is to be able to use and transition fluently among multiple representations. Research from other mathematical contexts indicates that these skills have the potential to enhance student learning [10]. Here we present a few ideas about how to help students develop these abilities.

Differential Equations courses have historically incorporated all four of these representations, though as the problems above suggest, students typically retreat to their comfort zone: algebra. Reformers have focused on getting students to make connections among the different ways of portraying ODEs, leading to an avalanche of new types of problems. Here we present just a handful of such problems that don't specifically rely on technology; a few that do will be discussed below. Others sources of innovative problems include various newer texts (see above for a list) and a couple of special journal issues dedicated to teaching ODEs ([1, 8, 14]).

A common type of question that gives students the opportunity to build connections between algebraic forms of ODEs and slope fields sounds deceptively simple. Six slope fields and six related ODEs (in algebraic form) are given, and students are asked to match them appropriately, giving reasons for their choices [2]. Variations of this include giving more equations than graphs or vice versa (so that process-of-elimination is less effective); replacing either the slope fields or the ODEs themselves with verbal descriptions of what's being modeled; and doing the same exercise with systems of equations and the corresponding phase portraits.

Questions that force students to reconcile graphical and algebraic representations needn't be complicated or technical. Again, given students' preference for algebraic representations, questions like the one in Figure 7.3 can foster a deeper understanding of different classes of ODEs.

> Is it possible to tell from a slope field whether or not an ODE is autonomous? Homogeneous? Linear? Why or why not? Can a solution to an autonomous ODE oscillate? Why or why not?

Figure 7.3. Visual Classification Problem

For all of these types of problems (including Problem 7 above), a good group or class discussion with students explaining, justifying, challenging, and defending each other's answers can help solidify the connections they make among the various representations.

When To Use Technology?

As you might imagine, a wide range of technological devices have been used in ODE courses. These can range from hand-held calculators and simple spreadsheets for use in making numerical calculations, to computer algebra systems (CAS), to ODE solving packages that now come with Maple, Mathematica, and MatLab. Your choice of what technology to use probably rests in large part on what is available for you to use. While there is still much to be examined via research on how the use of technology affects student learning, here we discuss a few studies that give us a window into what successful uses of technology look like.

In a class where technology is fully integrated into the course, students' understanding of the graphical representations used throughout ODEs can be improved [6]. Moreover, when graphical technology has been combined with the other innovations mentioned above into an inquiry-based, technology-enhanced course, students' performance on conceptual questions rises significantly without lowering their computational fluency [16].

Another study looked at four classes, two that included demonstrations using a CAS and two without [9]. A common exam given at the end of the semester uncovered no significant differences between the two groups in their ability to algebraically solve ODEs. Although few students in the sections using a CAS opted to use it outside of class, their attitudes toward the use of computers did change significantly. These results are hardly surprising since few people argue for the use of CASs as ways of improving computational fluency.

What is clear from this research is that technology is not, as some have claimed, a holy grail, nor is it the death of students' computational ability. Instead, future research is needed to better understand which content is better communicated to students through the use of which technology.

A Good Place to Start

Start the course with a conceptually-focused activity, such as the one above about cooperative versus competitive species, putting the focus on concepts not procedures. This also helps students come to understand that solutions to ODEs are **functions** (not numbers).

Bibliography

[1] *College Mathematics Journal* — special issue on differential equations, Nov. 1994.

[2] M. Artigue. Functions from an algebraic and graphic point of view: cognitive difficulties and teaching practices. *The concept of function: some aspects of epistemology and pedagogy, MAA Notes*, 25:109–132, 1992.

[3] K. Bain. *What the best college teachers do*. Harvard University Press, 2004.

[4] P. Blanchard, R. Devaney, and G. Hall. *Differential Equations*. Brooks/Cole, 4th edition, 2011.

[5] Robert L Borelli and Courtney S Coleman. Differential equations: a modeling approach, 2004.

[6] S. Habre. Exploring students' strategies to solve ordinary differential equations in a reformed setting. *The Journal of Mathematical Behavior*, 18(4):455–472, 2000.

[7] D. Hughes-Hallett, A.M. Gleason, and D.E. Flath. *Calculus*. Wiley, 1994.

[8] M.J. Kallaher. *Revolutions in Differential Equations: Exploring ODEs with modern technology*. Number 50 in Notes. Mathematical Association of America, 1999.

[9] T.J. Klein. *A comparative study on the effectiveness of differential equations instruction with and without a computer algebra system*. PhD thesis, Peabody College for Teachers of Vanderbilt University Nashville, TN, USA, 1993.

[10] S. Monk. Students' understanding of a function given by a physical model. *The concept of function: Aspects of epistemology and pedagogy, Mathematical Association of America, Washington, DC*, pages 175–194, 1992.

[11] Dolores D Pesek and David Kirshner. Interference of instrumental instruction in subsequent relational learning. *Journal for Research in Mathematics Education*, pages 524–540, 2000.

[12] C. Rasmussen and H. Blumenfeld. Reinventing solutions to systems of linear differential equations: A case of emergent models involving analytic expressions. *The Journal of Mathematical Behavior*, 26(3):195–210, 2007.

[13] C. Rasmussen and M. Keynes. Lines of eigenvectors and solutions to systems of linear differential equations. *Primus*, 13(4):308–320, 2003.

[14] C. Rasmussen and O. N. Kwon. An inquiry-oriented approach to differential equations. *The Journal of Mathematical Behavior*, 26, 2007.

[15] C. Rasmussen and O.N. Kwon. An inquiry-oriented approach to undergraduate mathematics. *The Journal of Mathematical Behavior*, 26(3):189–194, 2007.

[16] C. Rasmussen, O.N. Kwon, K. Allen, K. Marrongelle, and M. Burtch. Capitalizing on advances in mathematics and K-12 mathematics education in undergraduate mathematics: An inquiry-oriented approach to differential equations. *Asia Pacific Education Review*, 7(1):85–93, 2006.

[17] C. Rasmussen, M. Zandieh, K. King, and A. Teppo. Advancing mathematical activity: A practice-oriented view of advanced mathematical thinking. *Mathematical thinking and learning*, 7(1):51–73, 2005.

[18] C. Rasmussen, K. A. Keene, J. Dunmyre, and N. Fortune, (2018). Inquiry oriented differential equations: Course materials. Available at iode.wordpress.ncsu.edu. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

[19] C.L. Rasmussen. New directions in differential equations: A framework for interpreting students' understandings and difficulties. *The Journal of Mathematical Behavior*, 20(1):55–87, 2001.

[20] C.L. Rasmussen and K.D. King. Locating starting points in differential equations: A realistic mathematics education approach. *International Journal of Mathematical Education in Science and Technology*, 31(2):161–172, 2000.

[21] C.L Rasmussen and K. Whitehead. Learning and teaching ordinary differential equations, January 2003.

[22] N.M. Speer and J.F. Wagner. Knowledge needed by a teacher to provide analytic scaffolding during undergraduate mathematics classroom discussions. *Journal for Research in Mathematics Education*, 40(5):530–565, 2009.

[23] E. K. Stein, M. K., Engle, R. A., Smith, M. S., & Hughes. Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, 10(4):313–340, 2008.

[24] M. Stephan and C. Rasmussen. Classroom mathematical practices in differential equations. *The Journal of Mathematical Behavior*, 21(4):459–490, 2002.

[25] M. Trigueros. Students' conceptions of solution curves and equilibrium in systems of differential equations. In *Proceedings of the 22nd annual meeting of the North American chapter of the international group for the psychology of mathematics education*, pages 93–97, 2000.

[26] J.F. Wagner, N.M. Speer, and B. Rossa. Beyond mathematical content knowledge: A mathematician's knowledge needed for teaching an inquiry-oriented differential equations course. *The Journal of Mathematical Behavior*, 26(3):247–266, 2007.

[27] E. Yackel, C. Rasmussen, and K. King. Social and sociomathematical norms in an advanced undergraduate mathematics course. *The Journal of Mathematical Behavior*, 19(3):275–287, 2000.

[28] M. Zandieh and M. McDonald. Student understanding of equilibrium solution in differential equations. In *Proceedings of the 21st Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, pages 253–258, 1999.

[29] Dennis G. Zill. *A first course in differential equations with modeling applications*. Cengage Learning, 2012.

8

Analysis

> "I'm on another planet when it comes to Analysis. It seems just completely surreal to me. It sparks in a lot of people in the group ... a lot of people. I don't think there is anybody who understands it. And a lot of people are getting very frustrated, with it. I just want to throw books around the room and ... get up and leave." (an Analysis student, cited in [23, p. 7])

No doubt about it—students think analysis is hard. Reasons for this are easy to guess. Real Analysis students are asked to reason about mathematical objects that have complex definitions, use those definitions to produce rigorous proofs, and translate back and forth between deceptively simple pictures and intimidating notation. The cognitive loads demanded of Analysis students rise considerably beyond those required in earlier courses. As one student opined, "It's Calculus on steroids!"

Historically, analysis made a big leap forward when Cauchy, Weierstrass, and Reimann improved its rigorous foundation with their powerful mix of quantifiers and arbitrarily small, positive numbers [19]. The power of these new techniques belied their theoretical difficulty, as students of analysis have learned ever since.

The reasons students struggle with these concepts abound. Some students start this course without a firm understanding of the real line (though some would argue that the point of an Analysis course is to understand the real line in a deep way). Also, foundational topics like functions, set theory, and proof-writing play an important role. While we discuss some aspects of these issues in this chapter, you may also want to check out the sections on various calculus topics (especially Sequences and Series), and Proofs before proceeding. Here we focus on the challenges students face that are more specific to analysis.

Researchers looking into student thinking in analysis and related topics have drawn several distinctions that will be discussed throughout this chapter. An extensive set of studies in the UK indicates that students come into the course with a strong tendency toward either visual thinking or non-visual thinking [2, 3, 4, 5]. Students who prefer visual thinking refer repeatedly to mental pictures and draw diagrams when asked about definitions and proofs. Other students prefer to work directly with the formal definitions and rarely introduce diagrams unless prompted. Students in both groups can be extremely successful at producing rigorous proofs; on the other hand, many students in both groups struggle to do so.

A second distinction made in the literature involves how students think about and use definitions. Tall and Vinner distinguish between what a student thinks about a concept's strict mathematical definition (her "concept definition") and all of the mental images and ideas related to that concept (her "concept image") [25]. For instance, a student's concept definition for continuity might involve a (possibly jumbled) expression of εs and δs, while the ideas of "being able to draw the graph without picking up your pen" and/or the image of the graph being "pretty smooth" might be, accurate or not, part of their concept image for continuity. Researchers have repeatedly shown that students' concept images can conflict with their concept definitions, that only sometimes are students aware of this conflict, and that students frequently draw on their (sometimes incorrect) concept images to reason about and prove statements [11, 15].

With that background information, here's a sampling of what you'll see from your students when you teach Analysis, and what they're thinking.

Start From the Beginning

1. Define what it means to say "b is an upper bound of a set $S \subseteq \mathbb{R}$" and explain what the definition means in everyday language. (From [8, p. 3].)

Alessia:

$\forall s \in S$ $s \leq b$ Means b is bigger than everything in S.

2. Define what it means to say "b is the least upper bound (LUB) of S" and explain what the definition means in everyday language.

Alessia:

$\forall k \in \mathbb{R}$ s.t. $s \leq k$ and $b \leq k$

It's an Upper Bound – and the least of all of them

Given the responses to Problems 1 and 2, what do you think Alessia understands? How appropriate is her concept definition for upper bound and least upper bound (a.k.a., supremum or sup)? What are the appropriate and inappropriate aspects of her concept image for these two fundamental ideas about the real numbers?

While we have no way of answering these questions for certain, we can look for clues in Alessia's responses. Her correct formal definition of upper bound might conflict with her informal definition, depending on what she means by "bigger than." If she sees it as a strict inequality, then a follow-up question such as "Is 1 an upper bound of the set $(0, 1]$?" might help her see the conflict. In addition, her phrase "everything in S" is sometimes interpreted as (and sometimes meant as) "everything in S except the point in question" which might be more appropriately phrased as "everything *else* in S." Notice that these two subtle alternate interpretations might cancel each other out, so that if she is thinking of "bigger than" as a strict inequality and "everything" as "everything else," then she would answer the question about the upper bound of $(0, 1]$ correctly!

When it comes to the LUB, her thinking is clearly more muddled. Her formal definition leaves out important parts (notably, that b is an upper bound of S) and mixes up others. Her seemingly spot-on informal answer leaves room for a range of interpretations. It could be that in her confusion, she has chosen to parrot something she heard in class or read in the text, without really knowing what she was saying. Alternatively, her concept image for LUB might be relatively appropriate, with her limited ability to translate those thoughts into a quantified statement holding her back. In either case, she has work to do before her conception of LUB approaches the point where she could fluently and flexibly work through a complicated proof involving such objects.

As an interesting thought experiment, go back to the original problem and put yourself in the position of the person who wrote it. What would Alessia's answer have looked like without the second part of the problem ("explain what the definition means in everyday language")? This illustrates how you as a teacher can phrase problems to give you more insight into your students' thinking.

Worth How Many Words?

3. Define

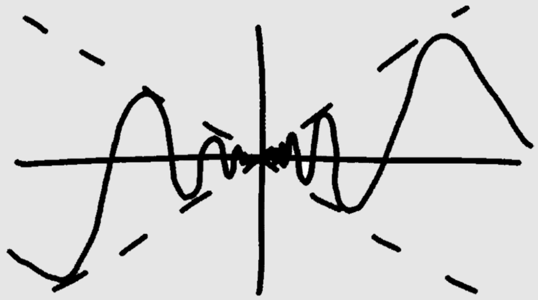
$$g(x) = \begin{cases} x\sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Prove that g is continuous at $x = 0$.

Bruce: Given $\varepsilon > 0$, let $\delta = \varepsilon$ so that when $|x - 0| = |x| < \delta$, since $|\sin| \leq 1$, we have $|g(x) - g(0)| = |g(x)| < \varepsilon$. Thus g is continuous at $x = 0$.

Catherine:

g is trapped between
x and −x so g has to → 0.



In Problem 3 we see the stark differences between a visual and non-visual approach to analysis. Clearly both students have a fairly good grasp of the key parts of continuity, right?

Or do they? Take a closer look at Bruce's work. If we remove the bound on the sine, what remains is hardly more than the skeleton of a continuity proof. One of the more popular coping strategies for students writing proofs is to simply memorize the outline of a proof and fill in the details that they see as needed [22]. It's possible that Bruce is doing exactly that (throwing in the bound on sine as a lucky guess), or he might have simply forgotten (or not had time) to copy the line $|g(x)| = |x \cdot \sin(1/x)| \leq |x| \leq \delta = \varepsilon$ from his unseen scratch work.

As for Catherine's work, the details of the graph indicate that she probably has a fairly good grasp of the behavior of this functions, although she doesn't see that g, as the product of two odd functions, is actually even. While students may blindly memorize some graphs, details like the spacing of the peaks and the bounding lines $y = \pm x$ make this unlikely in her case. Her work, however, leaves open many questions about what ideas of continuity she is relying on in forming her answer. She might be relying on as little as a concept image of continuity as "being able to draw the graph without picking up your pen" with little understanding of the formal definition. With her strong reliance on visual representations, how would she even begin to think about $x \cdot \chi_{\mathbb{Q}}(x)$ or the Cantor function? And what would she do with a continuous but nowhere differentiable function, whose visualization taxes experts?

Each of these students could probably benefit from the insights of the other. Although detailed research has yet to confirm analysis instructors' hunch, students who tend toward either visual or non-visual methods will probably come to a deeper understanding if they work to understand other approaches. (See below for tips on how one might lead them in that direction.)

How Definitions are Used in Mathematics

4. Is the function $f(x) = |x|$ continuous at $x = 0$?

Dylan: Let's see. I remember my high school Calc teacher explaining continuity as "the limit at the point equals the actual value at the point." I know the real definition now:

$$\forall \varepsilon > 0\ \exists \delta > 0\ s.t.\ |x - a| < \delta\ \rightarrow |f(x) - f(a)| < \varepsilon.$$

So let's look at $|x|$. Hmm. I think cusps, or there were a whole bunch of things that were not continuous. And, I think this is one of them. Although, it looks pretty continuous. I'm pretty sure . . . I feel like this is not continuous and my high school teacher's definition isn't cutting it—it's giving me a different answer, so I'm looking at the, at the real one. But that, but I know cusps, and sharp peaks are not continuous, but from the definition, if we're saying that the two one-sided limits equals $f(a)$, then that would be continuous. . . . But it's not, I just know it's not continuous.

In Problem 4 we see a conflict between Dylan's concept definition (his formal definition) and his concept image (all the thoughts surrounding this topic). His concept definition for continuity appears to have two logically equivalent parts (the ε-δ definition and the "limit = value at the point" idea), though Dylan may be unaware of their logical equivalence.

On the other hand, his concept image appears to incorrectly include the idea that a function is discontinuous at a cusp or corner. Ignoring the answer he gets when he explicitly references the definition, Dylan chooses to rely instead on his concept image and concludes that the absolute value function is discontinuous.

This example illustrates two typical phenomena. First, Dylan's struggle to come to terms with the subtle differences between his formal and intuitive mathematical reasoning echoes students' struggles in many mathematical subjects. Second, Dylan demonstrates a misunderstanding of the very nature of definition in mathematics. Each point is worth a closer look.

From an early age, students of mathematics develop naive impressions of the concepts they are learning, only to have those ideas rejected at a later date. Kids are expert noticers of patterns and typically believe that addition and multiplication "make numbers bigger" whereas subtraction and division "make numbers smaller." Negative numbers and fractions bring about cognitive conflict, forcing kids to question these patterns. The history of mathematics is also riddled with such conflicts, from the Pythagoreans' reaction to irrational numbers to the initial rejection of Cantor's pioneering work on infinity.

Analysis students bring with them tons of ideas about the real line, functions, continuity, and limits. Some of these ideas are naive and incorrect; others hold up under further scrutiny. Dylan's struggle illustrates students' fight to understand which are which.

The second point, regarding students' understanding of the nature of mathematical definitions, rears its well-camouflaged head in many different places. Mathematical definitions differ from those used in everyday language in that they are *stipulated* by the mathematical community, not *extracted* from common experience (see [16] for an extended discussion of this distinction). Children extract the meaning of "cat" after exposure to repeated examples; they do not distinguish cats from other animals based on the definition handed down by the Society for Feline Standards. Mathematicians view definitions as stipulated; given a definition, an object either does or does not satisfy it. Research suggests that for students, it's the leap from definitions that they construct through experience to stipulated mathematical definitions actually asks quite a bit of them [3].

In Dylan's response above, we see hints that he has not fully made this transition to thinking of definitions as stipulated. Even though he knows the stipulated definition that the mathematics community has come to agree upon, he puts less weight on that then the other parts of his concept image. He has extracted an incorrect definition of continuous function based on his prior experience (or memorized categories), and rejects arguments based on the mathematical definition. If he viewed mathematical definitions as stipulated, he would likely put more weight on the formal definition.

The importance of definitions in Analysis can hardly be overstated. Many instructors have responded to this by including activities geared toward student understanding of the key definitions in the subject. These include gateway exams where students are asked to recall definitions and apply them in elementary situations. However, while such activities might help students improve their understanding of definitions, they may do little to address the challenges of coming to understand the very nature of those definitions.

Working From a Prototype

5. Consider a sequence $\{a_n\}$. Which of the following is true? Justify your answer.

(a) $\{a_n\}$ is bounded $\Rightarrow$ $\{a_n\}$ is convergent

(b) $\{a_n\}$ is convergent $\Rightarrow$ $\{a_n\}$ is bounded

(c) $\{a_n\}$ is bounded $\Longleftrightarrow$ $\{a_n\}$ is convergent

(d) None of the above. (From [3, p. 3].)

Edna: Well if it converges, you get closer and closer ... I think it's (b). [while drawing a monotone decreasing, convergent sequence] ... It's convergent, yes, so if it's convergent it's always ... or ... [draws an increasing convergent sequence] it could be the other way around, going up this way. Yeah, the answer is (b). (Paraphrased from [3, p. 3].)

Here we see an example of a student who is completely ignoring mathematical definitions. Instead, she pictures a prototypical sequence (or two) and uses these to draw conclusions about all convergent sequences. Her pictures might even suggest that any convergent sequence is bounded between the first term and the eventual limit, an easily falsifiable claim. If she had appealed to the definition, she might not have implicitly assumed that all sequences are monotone.

Edna is far from alone in using prototypical examples to reason about mathematical statements, especially among students early in an Analysis course. In fact, students in almost every level of mathematics find examples incredibly compelling, frequently more so than deductive proofs (see the Proofs chapter for more on this).

Before we dismiss this type of thinking, we should note that experts in mathematics use examples in a wide variety of productive ways. While few make the naive mistake of relying entirely on examples to prove a general statement, we frequently use examples to shed light on a problem. Alcock has identified three separate ways in which mathematicians use examples [1]. After interviewing experienced college instructors, she reported that while working to understand a statement (for instance, a new definition), they claim to construct and examine examples. While working to generate a deductive argument, they also turn to prototypical examples with the goal of generalizing a proof that worked in one case to the wider situation. Finally, the experts described using examples while checking the validity of arguments, sort of a reality check to make sure that the logical connection between one step and the next is at least plausible.

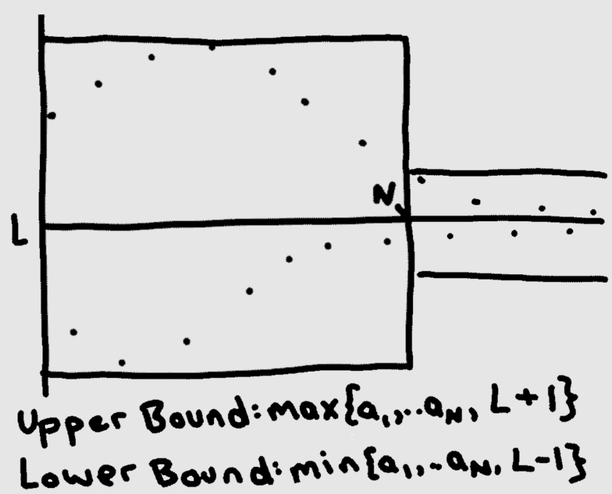
What undoubtedly distinguishes experts from students like Edna in their use of examples is that experts have a wider set of examples at their disposal, understand the limitations of those examples, and realize the potential danger of reasoning from them without appealing to definitions and deductive reasoning [7, 26].

OK, So Prove It!

6: Prove: $\{a_n\}$ is convergent $\Rightarrow$ $\{a_n\}$ is bounded.

Fletcher: OK, we use the definition of convergence here ... so eventually you'll find an ε such that ... for all n bigger than N, $|a_n - a| < \varepsilon$. So you've got it bounded between $a + \varepsilon$ and $a - \varepsilon$—for all $n > N$. And for all n smaller than big N, you know that a_n has a maximum and a minimum because it's finite.

Gillian:



(Adapted from [3].)

In Problem 6 we see two radically different, but both promising, approaches to this proof. While neither is complete, we might imagine both students going on, after a bit of work, to write a completely rigorous proof of this statement.

Fletcher's quintessentially non-visual approach works straight from the definition, although his exact wording suggests some confusion. Though he might have just misspoken, his wording ("eventually you'll find an ε") might indicate a reversal of the quantifiers. Still the general outline of his work captures the heart of the argument: use the convergence to bound the tail and the finiteness of the remaining terms to complete the argument. A simple question, such as "How are you going to find ε and N?" might help him to tighten up his argument.

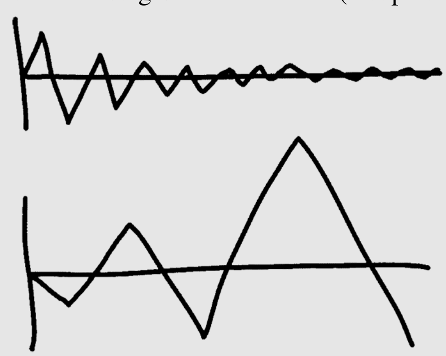
Gillian's quintessentially visual approach shows exactly how powerful a well-done diagram can be. All the elements of the proof are there, including the ε bound (in her case, with $\varepsilon = 1$) on the later terms and the bound on the first N_0 terms. While one can easily imagine a correct proof being based on this picture, it does leave some room for possible errors. Does Gillian understand the order of events represented in this picture, or the logical importance of that order? In her notation, L must be chosen first (from the definition of convergence), followed by N_0 (dependent on her choice of $\varepsilon = 1$), followed by the upper bound and lower bound (which depend on N_0, and thus her choice of $\varepsilon = 1$). Still, Gillian's work showcases how students who tend toward visual methods might leverage their pictures to create powerful mathematical arguments.

Analysis experts have the ability to understand both Fletcher and Gillian's approaches, a goal that students might be more likely to approach (or even actually reach) if they have opportunities to see and analyze alternative approaches.

The Power Series

7: When does $\sum_{n=1}^{\infty} \frac{(-x)^n}{n}$ converge? Justify your answer.

Hadi: If x is one or smaller, it will converge.
If x is bigger than 1—diverge. (Adapted from [4, p. 11].)



While Hadi's conclusions in Problem 7 (ignoring negative values of x) are basically correct, a glance at his graphs leaves one wondering how he came to his conclusions—maybe he just made a lucky guess. While Hadi's work is certainly far from ideal, we can learn something about students' visual approaches to a topic like power series.

In the research Hadi's work is based on, the authors include interview excerpts with a student who produced sketches like these [4]. The student carefully explained how he took two particular values for x (in this case, it might be $x = 0.5$ and 2) and looked at what happened to the sequence of partial sums. These drawings represented his qualitative analysis, with the horizontal axis representing the term in the sequence of partial sums and the vertical axis giving the value of that sequence.

Whether or not Hadi could translate this type of thinking into a correct proof (possibly using a ratio test, with a different analysis at $x = \pm 1$) is unknown. However, understanding the correct thinking behind such a picture might give you clues as to how to help a student build on the correct aspects of their thinking and move toward a more complete understanding.

Compactness

8: What does it mean for a set S in $\mathbb{R}^2$ to be compact?

Ita: It's compact if it can be covered by a collection of open sets, so that the union of these finitely many open sets contains S.

Jack: S is compact if for each collection of open sets which covers S, there exists a finite subcollection of open sets which will again cover the set S. (Adapted from [13].)

9: What does it mean for a set S in $\mathbb{R}^2$ to **not** be compact?

Ita: When you cannot, in any case, cover it with finitely many open sets.

Jack: It would not be compact if for each collection of open sets, infinite collection, that covers it, there is no finite subcollection which will also cover it. So you sometimes really do need the whole infinite collection. (Adapted from [13].)

Compactness is one of the more advanced (and widely useful) concepts included in a typical Analysis course. Its definition is recognized by mathematicians as being different in character from other definitions. As Dudley wrote, "Although it perhaps has less immediate intuitive flavor and appeal than most definitions, it has proved quite successful mathematically" [14, p. 34]. Historically, this definition was preceded by what we would now call sequential compactness (where every sequence in a set has a subsequence that converges in that set), which is logically equivalent in $\mathbb{R}^n$.

While the sequentially compact definition (or even a "closed and bounded" definition) would suffice for most topics covered in Analysis, the wider applicability of the covering definition leads most authors and teachers to use this more complicated version.

In Problem 8, we see two students struggling with this less-than-intuitive definition. When defining compactness, Jack gives a complete answer while Ita's confused answer doesn't have all the pieces, and the ones she does have don't fit together. Her definition would actually be satisfied by any bounded set in $\mathbb{R}^n$.

When asked about the negation of compactness in Problem 9, both students struggle. Ita gives an incorrect answer, though it is the correct negation to her first answer. Jack fails to make the switch from a universal quantifier to an existential one in his initial statement. This common mistake, negating the proposition without negating or switching the quantifiers occurs both in and out of the mathematics world. In this case, however, we have some indication that despite his incorrect words, he might have the correct idea. His last sentence ("So you sometimes really do need the whole infinite collection") indicates a possible awareness that only one example of an infinite cover without a finite subcover would suffice.

What opportunities do these students need so that they might gain a better understanding of compactness? Dubinsky and Lewin, after years of working with students like this, have decomposed students' understanding of compactness into three levels, all relying on several foundational topics [13]. At the initial level, students like Ita work to combine ideas of set operations, quantification, and negation. Students at level two (like Jack) have successfully put together these basic ideas but are still working to coordinate the definition and its negation, a necessary step to a complete understanding of compactness. Achieving level three requires rolling all of these ideas into a single mental object, with the ability to test the definition against specific sets and to use the definition in writing proofs. After laying out these levels, Dubinsky and Lewin advocate helping students transition through these levels, gradually developing the appropriate mental structures.

The Proof of the Pudding is in the Details

10: Prove that any sequence contained in [0, 1] has a convergent subsequence.

Kate: Take a sequence $\{a_n\} \in [0, 1]$. Since a_n has an infinite number of points, either $[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$ has an infinite number of points in a_n; label that interval as I_1. Choose a_{n_1} to be one of the points in I_1. Split I_1 into two closed intervals each of length $1/4$ and let a_{n_2} be one of the points of a_n in I_2, the half containing an infinite number of points in a_n. Continue this process, dividing the intervals in half each time, choosing a_{n_k} in I_k (the half with infinitely many points) where $|I_k| = \frac{1}{2^k}$. Then $a_{n_k} \to \cap I_k$, with the intersection being non-empty by the Nested Interval Theorem.

There's a lot to like in Kate's sketch of the proof of the Bolzano-Weierstrass Theorem. The details, however, lack precision, and might get her in trouble later. Mistakes in real analysis proofs run the gamut from rather obvious to incredibly subtle. If you haven't already, go back and find all of the errors (some of them rather subtle) in the answer to Problem 10 before reading on.

Starting from the top:

- $\{a_n\} \subset [0,1]$, not $\{a_n\} \in [0,1]$.
- There's notational confusion between a_n representing the entire sequence or just a term in the sequence.
- $\{a_n\}$ has an infinite number of terms, not necessarily an infinite number of points (for instance, a constant sequence has infinitely many terms but only one point.)
- While the "or" in "$[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$" might be inclusive, the next line ("label **that** interval") indicates exclusivity.
- You must choose $n_2 > n_1$ or the resulting object isn't a subsequence (and similarly, $n_{k+1} > n_k$). This is a key difference between a limit point of the set of numbers, in which order doesn't matter, and the limit of a (sub) sequence, in which order is key.
- $\cap I_k$ is a set, not a number, so it cannot be the limit of any sequence.
- It remains to prove that $\cap I_k$ is a set containing exactly one point.

Each one of these errors comes up repeatedly in Analysis. Some of these might be simple mistakes; others indicate more serious underlying conceptual issues. Students sometimes get caught up in the larger ideas, sometimes brushing the details under the rug. Many professors actually prefer this to the "missing the forest for the trees" option. These are, however, serious difficulties and students need opportunities to over comes them.

There are several ways of helping students steer clear of, or at least quickly move beyond, these types of mistakes. First, you can familiarize yourself with these mistakes so that you'll recognize them in your students' work. You could also warn them ahead of time that these are typical pitfalls. They'll make mistakes anyway, but you'll be able to refer them back to what you said. Finally, you could regularly have them analyze incorrect proofs (like the one above, or possibly ones they've written) to help them build the ability (and habit) of searching for the logical errors. Several texts include such proof evaluation exercises, supported by research [24, 6], where students are presented with a (sometimes false) theorem and (sometimes flawed) proof and asked to analyze both.

What You Can Do

The above research suggests many different things about the way we might approach an Analysis course. Based on this literature, college math instructors have instituted changes ranging from completely restructuring a course to including new types of assignments.

At the University of Warwick, mathematicians partnered with the Mathematics Education Research Centre to revamp their Real Analysis course. For one year they ran both the new and old versions, allowing them to study the differences between two cohorts of students [2].

In the new classes, each with about 30 students, a teacher and two peer tutors formed students into groups of three to six. In two-hour classes twice a week, the students worked through sets of questions from a textbook [9], designed to "develop rationales for the main definitions, construct the central arguments that lie behind the main theorems, and allow the students to use those theorems in subsequent arguments" [2, p. 101]. Students also compiled their answers into a portfolio that was submitted for credit. In addition to these workshops, they also attended a one-hour lecture each week, during which the main ideas of that week's questions were summarized.

The Warwick team was then able to compare the performance of these students with those from a more traditional, lecture-based course on common exams and in the subsequent semester of Analysis. On both measures, the students in the new course outperformed their counterparts by approximately 12 percentage points. The students were not randomly placed in the two groups, limiting the conclusions we can draw from this study, but it mirrors work in other courses where student-centered approaches outperform lecture-based methods [18].

While students performed very well on assessments, researchers argued that the course accomplished another goal: helping students move toward more mathematically mature thinking. For instance, students in the new course were more likely to correctly view mathematical definitions as stipulated (as opposed to extracted).

As for more specific techniques you can use to address the issues discussed above, other researchers have used the insights from their work to suggest classroom techniques.

The problems in this chapter indicate that many of students' challenges in Analysis are centered on the use of definitions. On this point, researchers have several specific suggestions designed to help students better understand definitions and correctly use them in proofs. Several authors suggest that when presented with a new definition or theorem, many students need guidance in how to make sense of the new concept. Put in the same situation, experienced mathematicians naturally begin playing around, finding examples and non-examples, and seeing why a theorem might fail if each of the hypotheses are (in turn) omitted. Dahlberg and Housman suggested that students should be explicitly asked to do the same, early in the process "requiring them to generate their own examples or have them verify and work with instances of a concept before providing them with examples and commentary" [10, p. 297].

Another way of helping students understand definitions is to have them compare the standard definition with other possible ones. Fendel and Resek's bridge course text does this for the concept of continuity, offering four or five tempting alternatives, all of which fall short [17]. Asking students to find examples that differentiate between any two possible definitions lets them get their hands dirty with the end result so that they understand the actual definition much better.

To help students deal better with definitions, Edwards suggests that students would benefit from explicit discussions of the differences between concept definitions and concept images [16]. After such a discussion, students might become more aware of the roots of their thinking and whether they are relying on rigorous definitions or more fallible concept images (as in the discussion of Problem 3 above). A similar strategy, also advocated by Edwards, is to explicitly discuss the differences between extracted and stipulated definitions. In both cases, the goal is to help students understand that there is a difference between formal and informal thinking, as well as to give them a vocabulary for discussing each others' work.

In terms of how students use definitions in Analysis, a simple observation might go a long way to helping student write better proofs. Early in an Analysis course, many, if not most, questions come in one of two flavors:

- This one thing has property B (e.g., Show that $\frac{n}{n+1}$ converges).
- Everything that has property A also has property B (e.g., Prove that every convergent sequence is bounded).

In these proofs, definitions are used in one of two ways. First, if we know that an object has property A, we can use the definition of that property in a proof. Less obvious to most students is how we use a definition in proving that an object has property B. The fact that the definition of B can dictate the logical structure of the proof is something that many students struggle with [16]. A statement that uses the same definition in both ways (for instance, "If $\{a_n\}$ and $\{b_n\}$ both converge, then $\{a_n + b_n\}$ converges") can be used to explicitly show students both uses.

Finally, many of the examples above focus on the differences between visual and non-visual approaches. While research indicates that students tend to take one approach or the other, all students could benefit from understanding multiple approaches to the subject. Research in calculus and K-12 mathematics shows that the best students flexibly translate among different representations, including visual and algebraic ones [20, 21]. In related fields such as physics, fluency in moving among graphs, equations and other representations is key [12]. Building such fluency takes years, but several classroom activities can help.

Simply presenting two different approaches is unlikely to get students to make the connections between them that you might see yourself. Instead, you could ask students who have taken different approaches to discuss their thinking, analyzing the parallels and differences for others in the class. Or if students put visual and non-visual approaches on the board, you could help students explicitly draw the connections between the two. Being extremely explicit in showing, for instance, how the ε in the formal proof of continuity corresponds to the (usually thin) neighborhood in the range in the picture can help students begin to more flexibly move between two representations. Another more direct approach is to explicitly ask students for multiple representations (e.g., Give a formal proof and illustrate your proof with a picture). This tactic urges students to avoid the path of least resistance (and least learning?) which might involve figuring out one way to approach a topic and ignoring all others.

A Good Place to Start

Help your students understand the importance of definitions in analysis. Many students extract their own view of "convergent" as the property shared by all of the convergent sequences they've ever seen. Writing an analysis proof requires understanding the formal (stipulated), ε–N definition.

Bibliography

[1] L. Alcock. Uses of example objects in proving. *Proceedings of the 28th Conference of the International*, 2:17–24, 2004.

[2] L. Alcock and A. Simpson. The Warwick Analysis Project: Practice And Theory. *Mathematics Education Research Centre, University Of Warwick, Coventry, UK*, 2000.

[3] L. Alcock and A. Simpson. Two components in learning to reason using definitions. *Proceedings of the 2nd International Conference on the Teaching of Mathematics (at the Undergraduate Level)*, 2002.

[4] L. Alcock and A. Simpson. Convergence of sequences and series: Interactions between visual reasoning and the learner's beliefs about their own role. *Educational Studies in Mathematics*, 57(1):1–32, 2004.

[5] L. Alcock and A. Simpson. Convergence of Sequences and Series 2: Interactions Between Nonvisual Reasoning and the Learner's Beliefs about their own Role. *Educational Studies in Mathematics*, 58(1):77–100, 2005.

[6] L. Alcock and K. Weber. Proof validation in real analysis: Inferring and checking warrants. *Journal of Mathematical Behavior*, 24(2):125–134, 2005.

[7] Lara Alcock and Matthew Inglis. Doctoral students' use of examples in evaluating and proving conjectures. *Educational Studies in Mathematics*, 69(2):111–129, 2008.

[8] L. Bills and D. Tall. Operable Definitions in Advanced Mathematics: The case of the Least Upper Bound. *Proceedings of PME*, 22(2):104–111, 1998.

[9] R.P. Burn. *Numbers and Functions: Steps Into Analysis*. Cambridge University Press, 1992.

[10] R.P. Dahlberg and D.L. Housman. Facilitating Learning Events Through Example Generation. *Educational Studies in Mathematics*, 33(3):283–299, 1997.

[11] R.B. Davis and S. Vinner. The Notion of Limit: Some Seemingly Unavoidable Misconception Stages. *Journal of Mathematical Behavior*, 5(3):281–303, 1986.

[12] A. DiSessa and B. Sherin. Meta-representation: An introduction. *The Journal of Mathematical Behavior*, 19(4):385–398, 2000.

[13] E. Dubinsky and P. Lewin. Reflective abstraction and mathematics education: The genetic decomposition of induction and compactness. *Journal of Mathematical Behavior*, 5(1):55–92, 1986.

[14] R.M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.

[15] B. Edwards. An undergraduate student's understanding and use of mathematical definitions in real analysis. *Proceedings of the 19th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, 1:17–22, 1997.

[16] B.S. Edwards and M.B. Ward. Surprises from Mathematics Education Research: Student (Mis) use of Mathematical Definitions. *The American Mathematical Monthly*, 111:411–424, 2004.

[17] D.M. Fendel and D. Resek. *Foundations of Higher Mathematics: Exploration and Proof*. Addison-Wesley, 1990.

[18] M. P. S. Freeman, S.L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M.P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

[19] J.V. Grabiner. Who Gave You the Epsilon? Cauchy and the Origins of Rigorous Calculus. *The American Mathematical Monthly*, 90(3):185–194, 1983.

[20] G. Leinhardt, O. Zaslavsky, and Mary Kay Stein. Functions , Graphs , and Graphing : Tasks , Learning , and Teaching. 60(1):1–64, 1990.

[21] S. Monk. Representation in school mathematics: Learning to graph and graphing to learn. In J. Kilpatrick, D. Schifter, and G. Martin, editors, *A Research Companion to the Principles and Standards for School Mathematics*, pp. 250–262. NCTM, Reston, VA, 2003.

[22] R.C. Moore. Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3):249–266, 1994.

[23] M.M.F. Pinto. Students' Understanding of Real Analysis. *Unpublished PhD Thesis, Warwick University*, 1998.

[24] A. Selden and J. Selden. The Role of Logic in the Validation of Mathematical Proofs. *The DIMACS Symposium on Teaching Logic and Reasoning, Rutgers University*, pp. 25–26, 1996.

[25] D. Tall and S. Vinner. Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2):151–169, 1981.

[26] A. Watson and J. Mason. *Mathematics as a constructive activity: The role of learner generated examples*. Erlbaum, 2005.

9

Abstract Algebra

In the late 1990s, a prominent mathematician made a challenge to the nascent undergraduate mathematics education community: “Show me one piece of math ed research that would help mathematicians be better teachers.” Georgia State Professor Ed Dubinsky, whose research focus had shifted from functional analysis to mathematics education, answered by pointing to a series of three papers that had recently been published. “Read these papers,” he claimed, “and you’ll be a better teacher.”

All three papers explored aspects of student thinking in one area: abstract algebra. (Curious? All three appeared in a special issue of the *Journal of Mathematical Behavior*: [5, 2, 1].)

What is it about binary operations, groups, cosets, quotient groups, rings, and fields that leaves students confused and frustrated, instead of amazed to learn the language of symmetries and the underpinnings of most of their K-12 mathematics? And what was in those papers that made Dubinsky, a research mathematician and experienced teacher, believe they could help you teach better?

To answer the second question, keep reading. The short answer to why students find abstract algebra so hard: it’s the “abstract” part that gets them.

Students who are successful (as you probably were) learn about the group axioms in an abstract sense, but still see how they are tied to the more familiar number systems of integers, rationals, and real numbers. They see how group elements acting on a normal subgroup create a partition, the pieces of which can be treated as the *elements* of a different group, the quotient group. And all of their knowledge rests on the firm foundation of things they learned years ago about odd and even numbers, clock arithmetic, and geometrical symmetries.

Other students learn sets of axioms and theorems that use these foreign objects called groups and rings, none of which they’ve ever seen before at all. The abstractness overwhelms them, and rather than grounding their new knowledge in bedrock, it’s tenuously surfing on a foundation of sand.

The fundamental concepts of abstract algebra require a solid understanding of prerequisite topics, deep thinking, and the ability to put complicated concepts together in new ways. And decades of research have now added many reams to the three papers Dubinsky suggested (mostly on groups, fewer on rings and fields), shedding light on what students are thinking when they struggle in Abstract Algebra.

The Sudoku Rule

1: In a group, why must each element x appear at most once in each row of the group table?

Andres:

Every group we've seen does this.

Blanca: If it appears twice $ab = ac$
- we can divide out the a's: so $b = c$.

Carlos: If it appears twice, $ab = ac$.
Multiplying by a^{-1} on the left: $b = c$.
So it only really appeared once.

As in the number grid game Sudoku, each element appears exactly once in each row (and each column) of the group's Cayley table. While students can be lost in the abstractness of (frequently unmotivated) group axioms, they tend to notice this property quite quickly.

Andres doesn't seem to have moved beyond this initial pattern recognition, simply noting that the Sudoku property holds in all the examples of groups he's familiar with.

Blanca makes better progress in two ways, correctly writing a symbolic form of the hypothesis (that $ab = ac$) and understanding that logical justification is necessary. However, like many group theory students, she abandons the abstract thinking needed to understand an arbitrary group with an arbitrary operation, and instead falls back on a more comfortable operation: multiplication and its inverse, division (as indicated by her "divide out" language). Being attentive to the fact that there is an operation is a step in the right direction, but while this cancellation law is correct in this case, her words suggest that she doesn't understand why it holds for all groups.

This strategy of reducing abstraction from something uncomfortable and new (in this case, an abstract binary operation) to something old and familiar (multiplication) happens throughout the course [9, 10] and even much earlier in students' mathematical learning [12]. This simply reflects a basic reality of learning: when learning something new, we seek to tie the new concept into some old framework that we already understand. Eventually, that new concept (in Blanca's case, that we can talk about an abstract binary operation without referring to a specific one) becomes second nature, something to which we can tie some new concept (e.g., subgroups). That can be a long process, one that Blanca hasn't yet finished. Her work does make us wonder: if she had been thinking the same thing, but said "cancel the *a*s" would we have seen her conceptual error? If the problem had given a new symbol to the operation (as in $a\#b$) would she have fallen back on multiplicative language, or been completely flummoxed?

Carlos finds the correct way to justify Blanca's "dividing out," operating by the inverse of a on the left. Most mathematicians would give his work full credit, and take a few points off of Blanca's work. But hang on a second. There are ways that Carlos's answer could be improved. Fully justifying the computation requires not just the use of the inverse axiom, but also the associativity axiom (as in $a^{-1}(ab) = (a^{-1}a)b = eb = b$). So while Blanca errs by giving the name "divide" to the inverse, Carlos errs when he does not explicitly justify his response by citing associativity. Why do we grade Blanca's mistake more harshly than Carlos's? You might have your own answer, but for us, Blanca's work indicates that she is still inappropriately reducing abstraction, whereas Carlos isn't. Because working at an abstract level is key to abstract algebra, and because we see Carlos's response as more theoretically advanced in these ways, we think Carlos is further along on the path to understanding the ideas.

Finally, one other positive aspect of these students' work is worth mentioning. Many students fall prey to the mistaken idea (see also Problem 5 in the Proof Writing chapter):

$$\text{Different Names} \Rightarrow \text{Different Objects} \quad [20].$$

After all, the symbols b and c look different so they must refer to different things! Understanding that two variables, with different names, might refer to the same thing is crucial in proving that any function is one-to-one ($f(x) = f(y) \Rightarrow x = y$) and it appears that both Blanca and Carlos have this idea down.

What's in a Subgroup?

2: Is $\mathbb{Z}_3$ a subgroup of $\mathbb{Z}_6$? Explain your reasoning.

Dolores:

Yes. It's a subset and a group itself.

Enrique:

Yes - by Lagrange's Theorem: $3 | 6$

Felicia:

Yes: $\{0, 2, 4\}$

(Adapted from [8].)

Let's just make sure we're on the same page here, a different page from all three of these students. The correct answer to Problem 2 is an emphatic "NO!" The operation in $\mathbb{Z}_6$ is addition *mod 6*, and that's not the same operation as in $\mathbb{Z}_3$. If it's not the same operation, you don't have anything to check—one can't be a subgroup of the other! Beyond that, even the elements are different. When we write $0 \in \mathbb{Z}_3$, the 0 represents a coset in the quotient group ($\mathbb{Z}/3\mathbb{Z}$, where $0 = [0] = \{\cdots -3, 0, 3, 6, 9, \dots\}$). The "0" is simply a different beast from the coset $0 = [0] = \{\cdots -6, 0, 6, 12, \dots\} \in \mathbb{Z}_6$, even though we frequently use the same symbol (and name) to represent both.

It's likely that Dolores is thinking of $\mathbb{Z}_3$ as the set $\{0, 1, 2\}$, and $\mathbb{Z}_6$ as the set $\{0, 1, 2, 3, 4, 5\}$ without paying attention to what the digits stand for or the operation on either set. This problem of thinking of groups more like sets of numbers, without reference to the operation, happens all the time, to both students and teachers. A simple statement like "The group $\mathbb{Z}_3$ has three elements: 0, 1, and 2" may in fact mean very different things to experts and novices. In fact, mathematicians regularly refer to groups without specifying the operation, unless the operation isn't clear from the context. Experts always have the operation in the backs of their minds, pulling it out when necessary. Novices, like Dolores, frequently haven't come to understand this yet and don't always realize when greater precision is needed [8].

Enrique's work has shades of Delores' thinking ($\mathbb{Z}_3 \subseteq \mathbb{Z}_6$), but also cites one of the most important theorems in the course—but he gets it backwards. A correct statement of Lagrange's theorem puts restrictions on the possible orders of subgroups, but doesn't imply the existence of such subgroups. Students incorrectly interpret the theorem as going both ways, concluding some pseudo-converse version of the theorem:

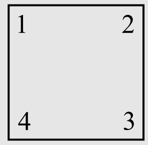
> If the order of the subset divides the
> order of the group, it's a subgroup. [11]

This thinking isn't unusual. In one study, over 17% of students responded like Enrique to this exact problem [11]. In fact, logical errors related to Lagrange's theorem pop up fairly frequently. Many students incorrectly reason that there can't be a subgroup of the symmetric group S_5 which is isomorphic to S_4 because "four doesn't divide 5," ignoring the fact that the subscript doesn't describe the order of the group [11].

Finally, Felicia's response most closely matches how experts talk about these two groups: "There's a copy of $\mathbb{Z}_3$ inside $\mathbb{Z}_6$." If pushed, an expert could of course explain that $\mathbb{Z}_6$ *does have* a subgroup which is *isomorphic* to $\mathbb{Z}_3$. Still the distinction between "being a subgroup" and "being isomorphic to a subgroup" is an important one that Felicia may still be working to understand. Helping your students to make that distinction explicit will clarify the meanings of these tricky mathematical objects and help you know whether or not they really do understand.

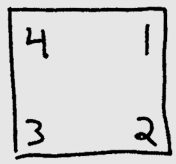
Some Things Stay the Same, . . .

3: The symmetries of a square form the dihedral group D_4, which can also be viewed as a subgroup of the symmetric group S_4 (the group of permutations of a four-element set). One of the elements of D_4 is R, a 90° clockwise rotation. What is the geometrical result of R applied to the square below?



Guillermo:

Hilda:



What is the result of R applied to the original square, given as a permutation in S_4?

Ignacio:

$1 \to 2, 2 \to 3, 3 \to 4, 4 \to 1$

or $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$

Jimena:

$1 \to 4, 2 \to 1, 3 \to 2, 4 \to 3$

(Adapted from [25].)

Once you've worked with symmetry groups for a few years, the visualization needed to perform the transformations that are the elements of these groups becomes second nature. For students new to the topic, this facility takes some time to develop. The students' answers to Problem 3 show some of what makes these ideas difficult.

The core idea of symmetry groups involves a potentially confusing convention. A symmetry might be defined as the set of rigid motions that brings a figure back exactly to where it begins. But to visualize such a motion, as with the square in Problem 3, we label the figure. Are the labels part of the figure? Then *no* rigid motion works (other than the trivial one). So the labels are separate from the figure. But then why do the labels move as the figure is rotated? And do the labels turn with the figure as in Guillermo's work, or do they magically travel with the figure but remain upright, as in Hilda's depiction? Somehow we want the corners of the object to simultaneously be indistinguishable (so non-trivial symmetries exist) but distinguishable (so we can see and talk about the different symmetries). This convention may be especially confusing to students whose experiences with transformations may come primarily from their secondary school geometry courses. In those courses, typically when a transformation is performed on something like triangle ABC, the resulting triangle is referred to with a new name, such as A′B′C′.

But the confusion doesn't stop with the labels, as the second part of Problem 3 suggests. Since any symmetry of the square permutes the four vertices, there should be a natural homomorphism of D_4 into S_4. While this map may feel like a identity map, essentially embedding one group inside a larger one, the notations we use for the symmetric groups look very different from the the notation we use in D_4.

Both Ignacio and Jimena have reasonable choices for the image of the rotation R under this map. Ignacio might be thinking of the numbers in his representation of S_4 as referring to the positions of the four corners, positions that don't turn when the figure is rotated. With this frame of reference, the vertex occupying position 1 (the northwest corner, NW) moves to position 2 (NE), and so on.

In contrast, Jimena reads the labels off the square after the transformation is applied. The labels start in position $\{1, 2, 3, 4\}$ (reading clockwise from the NW corner) and end up in position $\{4, 1, 2, 3\}$. In some sense, Jimena is rotating her frame of reference as the object rotates, where Ignacio keeps his frame of reference fixed [25].

Which is the correct interpretation, that is, which is part of the homomorphism from D_4 to S_4? It's Ignacio's choice that works. Unfortunately, many students (8 of 10 in one study) find Jimena's more natural [25].

To see this, imagine composing the rotation above with a flip over the vertical axis (F). This flip would switch 1 with 2, and switch 3 with 4, so it would be represented (for both Ignacio and Jimena) as $F = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$. But now if we compose the two (reading left to right), using Jimena's notation, we would have

$$F * R = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} * \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}.$$

This clearly doesn't match the geometrical version, where a flip followed by a 90° clockwise rotation would end like this:

$$\boxed{\begin{matrix} 1 & & 2 \\ & & \\ 4 & & 3 \end{matrix}} \xrightarrow{F} \boxed{\begin{matrix} 2 & & 1 \\ & & \\ 3 & & 4 \end{matrix}} \xrightarrow{R} \boxed{\begin{matrix} 3 & & 2 \\ & & \\ 4 & & 1 \end{matrix}} \quad .$$

In fact, Jimena's choice does embed D_4 in A_4, just not in the way that corresponds to the geometrical interpretation [25]. The take-home message? Working with symmetries is complicated and students need to understand both the ideas and the necessity for being explicit and precise when working with those ideas. Although as instructors we may explain, for example, what we mean when we label a square, what stays fixed, and what moves as we introduce symmetry groups, in the end it is the students who need to be able to communicate in these ways about the ideas they are learning.

Am I Normal?

4: Explain how you would determine if a subgroup K of a group S is normal.

Kevin: You check: is multiplying two elements the same in the reverse order.

Linda: if $s \in S$ and $k \in K$, see if you always get $sk = ks$.

Marty: if the set sK is always the same as Ks.

(Adapted from [1].)

The definition of a subgroup $K \subseteq S$ being normal is surprisingly easy to state:

$$\forall s \in S, sK = Ks.$$

Understanding what this actually means—and why we care—is far from easy. All three answers to Problem 4 demonstrate some inkling about the definition, but none of these students gets it exactly right.

Kevin's answer is vague and difficult to decipher. Are both of the elements from the subgroup K? Is one of them from the subgroup and the other from the larger group? He's not too far from defining commutativity, or even the center of the group, or saying that K is a subset of the center, but without more attention to quantifiers and where the elements live, it's difficult to say for sure what he really means.

But his vague language does give us some insights into the jumble of algebra thoughts inside his head. He might be trying to recall the details of the definition, which he has tried to store in its symbolic form. What he remembers of the definition appears to be disconnected from any intuitive sense of what "normal subgroup" might mean. Without any intuition to fall back on, he seems to be trying to recall the definition, and now the more vague he is, the more likely a teacher will be to brush over the lack of precision, “Yeah, that’s basically right!” giving him the benefit of the doubt.

Linda’s work adds a bit of precision, with a clear statement of where the elements s and k live, and at least some understanding of the universal quantification (though for her, “you always get” may mean quantification over $s \in S$ or $k \in K$ or both). In fact, with a few minor changes, she’d have a perfectly good definition of K being the center of the group S which is not an uncommon mistake among students first learning about normal subgroups [1].

Finally, Marty’s work comes closest to a correct definition, with only the $\forall s \in S$ missing. However, that quantifier is key; if you assume he means $\exists s \in S$, then every subgroup satisfies his definition (because for the identity e, $eK = Ke$ trivially). This may just an oversight or his conception of normality might need some further development; only a followup question might help us know.

A question these students might have even more trouble answering would be this: why do we care if a subgroup is normal? Understanding that the normality condition is exactly what we need in order for the cosets themselves to have a group structure is a worthy (and reachable) goal for an Abstract Algebra course.

Je m’appelle Coset(te)

5: If G is the group $\mathbb{Z}_{18}$ and H is the normal subgroup generated by 3 (i.e., $H = < 3 >$), then the cosets form the quotient group G/H. In that group, what is the sum of the two cosets, $(1 + H) + (2 + H)$? Explain.

Nora: 3+2H – just add the parts

Olaf: 1+2+ H+H = 3+H
(H is a subgroup so H+H = H)

Patricia: 4∈1+H, 8∈2+H so you get
12+H = 0+H (12≡0 mod 3)

Rick: (1+H)+(2+H) = {1,4,7,10,13,16}
+ {2,5,8,11,14,17
= {3,6,9,12,15,18,21,24,27,30,33}
= 0+H.

(Adapted from [8].)

Defining normal subgroups is the key to defining quotient groups. But even students who have a firm grasp of subgroups and normality can struggle to understand what it means to create a group out of the cosets. In the work on Problem 5 we see four different approaches, of varying degrees of sophistication, to how a group operation extends to an operation on the cosets.

Nora’s work is the most naive, and the most problematic. It looks like she has left behind the world where symbols have meaning and is simply manipulating them: she’s doing high school algebra on abstract algebra objects. This is a common response to being lost in the many layers of abstraction, and might require a return to earlier concepts to make sure her knowledge is better grounded.

Olaf’s work shows more evidence of understanding the underlying concepts, with an acknowledgement that H is itself a group. Still, there is nothing to suggest that he’s dealt with the complexity that both $(1 + H)$ and $(0 + H)$ are themselves sets. While H clearly represents a subgroup for him, it’s less clear that the cosets are well understood.

Patricia and Rick show two different ways to deal with the complexities of working with cosets both as subsets of one group (the original group) and as elements of another (the quotient group), both achieving partial success. Patricia starts by choosing representatives different from those given in the problem, showing more sophisticated thinking than Olaf. Then she adds the elements using the addition in the original group, finding the coset to which the result belongs. Given that the subgroup is normal, this is a perfectly good way to add cosets, since normality is exactly the condition that ensures that the result will not depend on your choice of elements. Does Patricia understand that subtlety, the need for the operation to be well-defined? There's no way to tell from her work.

Finally Rick's work shows a different strategy. Instead of boiling each coset down to a representative element, he leaves the cosets intact, completing the addition by listing all of the possible sums of two elements. After doing so, he repackages this set as $(0 + H)$. Unlike Patricia's strategy, the use of normality is more visible here. If the subgroup had not been normal, the result of the coset addition might not have been a single coset, a subtlety Rick may or may not have noticed.

If your goal is to help students develop an understanding of how the operation in the quotient group is related to the operation in the original group, one group of researchers has shown that, with the right prompts, students will naturally reinvent the concept of the quotient group using something akin to Rick's method [16]. In the compact notation of cosets (like aH), the question of normality is hidden as a non-obvious question of the operation being well-defined. When written out in terms of a coset's elements, the normality question is more obvious to students. (More on this "Guided Reinvention" method below.)

Can I Get a Quotient In Here?

6: Consider the group G given by the following operation table:

$*$	a	b	c	d	e	f
a	a	b	c	d	e	f
b	b	d	f	a	c	e
c	c	e	a	f	b	d
d	d	a	e	b	f	c
e	e	f	d	c	a	b
f	f	c	b	e	d	a

Could one use the subgroup $H = \{a, c\}$ to construct a quotient group? Explain [15].

Sandra:

$aH = \{a,c\}$ $Ha = \{a,c\}$

$bH = \{b,f\}$ $Hb = \{b,e\}$

Not Normal

Terry:

Elements:

$\{a,c\}$, $b\{a,c\} = \{b,f\}$, $d\{a,c\} = \{d,e\}$

	$\{a,c\}$	$\{b,f\}$	$\{d,e\}$
$\{a,c\}$	$\{a,c\}$	$\{d,e\}$	$\{d,e,f,b\}$
$\{b,f\}$			
$\{d,e\}$			

No! Not closed!

In Problem 6 we see two radically different approaches, both successful, to a non-standard problem about quotient groups. Sandra approaches the problem from the perspective of the definition of normal subgroup, checking to see whether or not the left and right cosets are the same. Finding two that are different, she draws the correct conclusion. While it appears that she knows that normality is required to form a quotient group, it's less clear whether she knows why this is the case.

In contrast, Terry finds three of the cosets that partition the set and then attempts to make a group table for them. He runs into trouble when one operation produces a set that is not in the partition and correctly concludes that no such operation on that particular partition is possible.

Students in traditional Abstract Algebra courses are unlikely to make an attempt like Terry's, in which he clearly demonstrates a deep understanding of how cosets can form a group themselves [15]. However, this material can be taught in such a way that students approach problems in such a sense-making manner. Intrigued? Stay tuned for more.

Missing the Forest for the Trees

7: Are the additive groups $\mathbb{Q}$ and $\mathbb{Z}$ isomorphic? Explain [23].

Vivian: Well $\mathbb{Z}$ is smaller ... wait ... they are both countable. What was the cool isomorphism Cantor constructed?

Waldo: They're both countable ... but $\mathbb{Z}$ is cyclic—it's generated by 1, but $\mathbb{Q}$ isn't cyclic.

When researchers interviewed students and asked them whether two groups were isomorphic, an interesting pattern appeared [23]. In response to tasks such as ones in Problem 7, the undergraduate math majors largely reasoned as Vivian did. They first checked the cardinality of the two sets; if they had the same cardinality, they looked for an isomorphism. In stark contrast, the graduate students paid attention to the group structures, as Waldo does.

Groups can have many properties; they can be abelian (commutative), cyclic, simple, free—the Wikipedia page lists more than 60 such characteristics (including "metanilpotent" !?!) [24]. Graduate students rely on a handful of these when comparing groups; undergraduates tend to rely on just one: cardinality [23].

This difference indicates a larger issue seen when undergraduate students were asked to write abstract algebra proofs. They largely lacked the big-picture understanding of groups needed to put together a coherent proof. Weber described this important knowledge as "strategic knowledge" [21] including:

- Knowledge of common proof techniques
- Knowledge of important theorems and their typical use
- Knowledge of when and how to use straightforward "syntactic" strategies

While attempting to prove statements about abstract algebra, undergraduates sometimes got tripped up because they couldn't recall a particular definition or theorem. But even when they had all of the requisite knowledge on hand, their lack of strategic knowledge frequently sank their attempts.

What You Can Do

Yes, the content in Abstract Algebra challenges the abstract reasoning abilities of students everywhere, but do not despair! The same researchers who identified the problematic reasoning discussed above also studied ways to improve student thinking about groups, cosets, quotient groups, and the like. You've seen some of these suggestions in other chapters. In one study, Belgian students studying group theory in, appropriately, small groups, achieved better understanding and retention, at the expense of a slower pace [6], a result you might expect if you've been reading other chapters. Here we focus more on the subject-specific research and what it suggests for teaching abstract algebra.

Examples and Counterexamples First, Then Definitions

We all know that working with examples of groups (and non-groups) is absolutely vital to really understanding the material. But should you work with the examples first, or the definition? And does working with counterexamples (e.g., sets with operations that fail one or more of the group axioms) help or hinder understanding?

One of the earliest studies of student learning in abstract algebra sought to answer these questions by having students watch different introductions to groups [17]. The videos came in four flavors in a tidy 2×2 grid: definitions first or afterwards; counterexamples included or not. The results? Learning is best when you do both examples and counterexamples, and do them before presenting the axioms.

But before you go write that perfect lecture, the research provides two alternatives for getting students to construct the mental structures needed to understand abstract algebra: guided reinvention and programming. Indications are that both produce a more grounded understanding than passive lectures.

Guided Reinvention

The Teaching Abstract Algebra for Understanding (TAAFU) project sought a way to leverage students' prior mathematics understanding to help them make sense of the many difficult concepts in the course. In a series of studies beginning in 2002, Sean Larsen led a team of researchers to build a curriculum through which students would naturally develop the key concepts of group, isomorphism, and finally quotient group [13, 14, 16]. The resulting resources now include instructional support materials [18] and have been implemented by other mathematicians [15].

So, what exactly is *guided reinvention*?

In this process, students work collaboratively to explore carefully crafted questions and generate their own ways of understanding and approaching algebraic topics. The first few activities involve exploring the symmetries of an equilateral triangle, naming them, and understanding what it means to compose them. With hints and suggestions, students settle on a definition of a group, with all the important axioms coming from their observations of the triangle's symmetries. Proper instructional suggestions are key, especially when it comes to notation. For instance, encouraging them to adopt a scheme that includes a single rotation (say, $R = 120°$ clockwise rotation) and a single reflection (say, $F =$ flip over vertical axis) allows them to more simply name the remaining symmetries (e.g., RR for a 240° rotation; FR for a rotation followed by a flip). Such a naming system naturally leads students to consider the composition of symmetries, as well as the associative property (so that, for instance, $F(FR) = (FF)R = R$). This simple setting is enough to get students to conjecture and prove a number of results, including:

- The cancellation law (needed in Problem 1 above).
- An appropriate definition of a subgroup.
- Any non-empty, closed, finite subset of a group must be a subgroup.

Through subsequent activities, the concept of the isomorphism is reinvented and defined, and the students find two subsets of D_4 that mimic the familiar concept of evens and odds, seeing that these subsets behave as a group themselves. Generalizing to other subsets, and using colors to distinguish the various sets (as in Figure 9.1), students eventually construct the condition we call normality, exactly the right condition under which the cosets themselves form the quotient group.

While this process of giving students carefully crafted exercises that will lead them to construct the key group theory concepts is long, studies suggest that the resulting understanding is significantly deeper. For instance, when asked to justify whether or not a particular subgroup could be used to construct a quotient group (like Problem 6) about 30% more students from TAAFU classes answered correctly when compared with those from a traditional class [15]. And while less than 10% of the students in a traditional class attempted a strategy of writing out a group table whose elements were cosets (as Terry did in Problem 6), more than 30% of the TAAFU students did [15].

	$\{\mathbf{I}, \mathbf{R}^2\}$	$\{\mathbf{R}, \mathbf{R}^3\}$	$\{\mathbf{F}, \mathbf{FR}^2\}$	$\{\mathbf{FR}, \mathbf{FR}^3\}$
$\{\mathbf{I}, \mathbf{R}^2\}$	$\{\mathbf{I}, \mathbf{R}^2\}$	$\{\mathbf{R}, \mathbf{R}^3\}$	$\{\mathbf{F}, \mathbf{FR}^2\}$	$\{\mathbf{FR}, \mathbf{FR}^3\}$
$\{\mathbf{R}, \mathbf{R}^3\}$	$\{\mathbf{R}, \mathbf{R}^3\}$	$\{\mathbf{I}, \mathbf{R}^2\}$	$\{\mathbf{FR}, \mathbf{FR}^3\}$	$\{\mathbf{F}, \mathbf{FR}^2\}$
$\{\mathbf{F}, \mathbf{FR}^2\}$	$\{\mathbf{F}, \mathbf{FR}^2\}$	$\{\mathbf{FR}, \mathbf{FR}^3\}$	$\{\mathbf{I}, \mathbf{R}^2\}$	$\{\mathbf{R}, \mathbf{R}^3\}$
$\{\mathbf{FR}, \mathbf{FR}^3\}$	$\{\mathbf{FR}, \mathbf{FR}^3\}$	$\{\mathbf{F}, \mathbf{FR}^2\}$	$\{\mathbf{R}, \mathbf{R}^3\}$	$\{\mathbf{I}, \mathbf{R}^2\}$

Figure 9.1. Colored Cosets of the Dihedral Group D_4

0	(0 1)(2 3)	(0 3)(1 2)	(0 2)(1 3)	(1 3 2)	(0 1 2)	(0 3 1)	(0 2 3)	(1 2 3)	(0 1 3)	(0 3 2)	(0 2 1)
(0 1)(2 3)	0	(0 2)(1 3)	(0 3)(1 2)	(0 1 2)	(1 3 2)	(0 2 3)	(0 3 1)	(0 1 3)	(1 2 3)	(0 2 1)	(0 3 2)
(0 3)(1 2)	(0 2)(1 3)	0	(0 1)(2 3)	(0 3 1)	(0 2 3)	(1 3 2)	(0 1 2)	(0 3 2)	(0 2 1)	(1 2 3)	(0 1 3)
(0 2)(1 3)	(0 3)(1 2)	(0 1)(2 3)	0	(0 2 3)	(0 3 1)	(0 1 2)	(1 3 2)	(0 2 1)	(0 3 2)	(0 1 3)	(1 2 3)
(1 3 2)	(0 3 1)	(0 2 3)	(0 1 2)	(1 2 3)	(0 3 2)	(0 2 1)	(0 1 3)	0	(0 3)(1 2)	(0 2)(1 3)	(0 1)(2 3)
(0 1 2)	(0 2 3)	(0 3 1)	(1 3 2)	(0 1 3)	(0 2 1)	(0 3 2)	(1 2 3)	(0 1)(2 3)	(0 2)(1 3)	(0 3)(1 2)	0
(0 3 1)	(1 3 2)	(0 1 2)	(0 2 3)	(0 3 2)	(1 2 3)	(0 1 3)	(0 2 1)	(0 3)(1 2)	0	(0 1)(2 3)	(0 2)(1 3)
(0 2 3)	(0 1 2)	(1 3 2)	(0 3 1)	(0 2 1)	(0 1 3)	(1 2 3)	(0 3 2)	(0 2)(1 3)	(0 1)(2 3)	0	(0 3)(1 2)
(1 2 3)	(0 2 1)	(0 1 3)	(0 3 2)	0	(0 2)(1 3)	(0 1)(2 3)	(0 3)(1 2)	(1 3 2)	(0 2 3)	(0 1 2)	(0 3 1)
(0 1 3)	(0 3 2)	(1 2 3)	(0 2 1)	(0 1)(2 3)	(0 3)(1 2)	0	(0 2)(1 3)	(0 1 2)	(0 3 1)	(1 3 2)	(0 2 3)
(0 3 2)	(0 1 3)	(0 2 1)	(1 2 3)	(0 3)(1 2)	(0 1)(2 3)	(0 2)(1 3)	0	(0 3 1)	(0 1 2)	(0 2 3)	(1 3 2)
(0 2 1)	(1 2 3)	(0 3 2)	(0 1 3)	(0 2)(1 3)	0	(0 3)(1 2)	(0 1)(2 3)	(0 2 3)	(1 3 2)	(0 3 1)	(0 1 2)

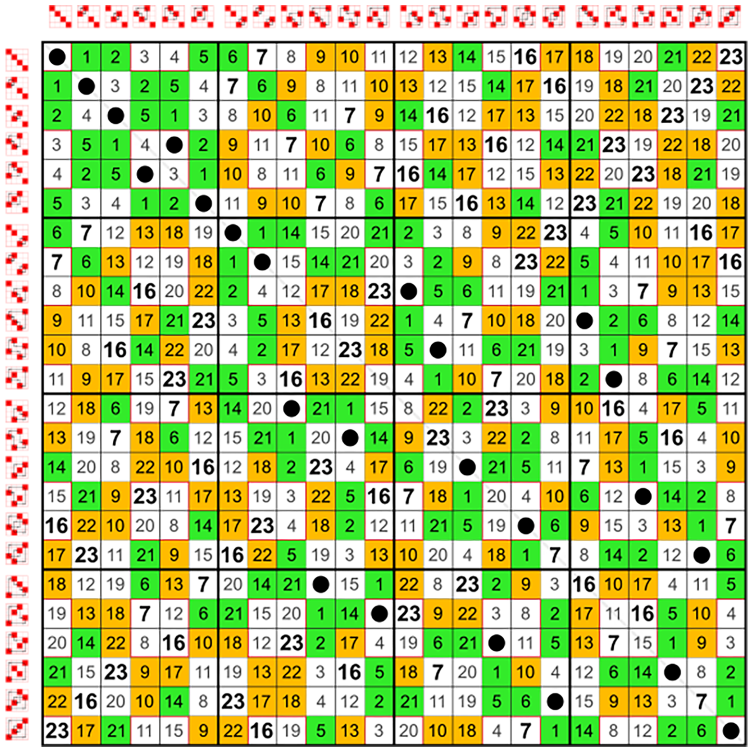
Figure 9.2. Group table for the alternating group A_4, ordered and colored to highlight the $\mathbb{Z}_3$ quotient group structure for the normal subgroup $\{(0), (01)(23), (03)(12), (02)(13)\}$.

Visual Representations

Although there is little research on the effectiveness of using visual representations of groups, research from other topics suggests this can be a powerful way to improve students' understanding, especially their intuition about the structure of groups. The many different visualization tools available for the topics in this course include:

- Cosets in group tables (using colors, as in Figure 9.2).
- Group tables with element properties highlighted (as for the symmetric group S_4 in figure 9.3).
- Homomorphisms as maps between group tables.

These representations can be powerful. For instance, a normal subgroup partitioning a group and forming a quotient group is equivalent to coloring the elements in the larger group's table with each coset taking on a different color so that the group table, *now viewed only as a table of colors*, forms a group table itself. Such a representation allows one to see both the structure of the original group and the structure of the quotient group simultaneously. These visualizations are numerous enough that Nathan Carter compiled many into a book called *Visual Group Theory* which leverages visual



●	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	●	3	2	5	4	7	6	9	8	11	10	13	12	15	14	17	16	19	18	21	20	23	22
2	4	●	5	1	3	8	10	6	11	7	9	14	16	12	17	13	15	20	22	18	23	19	21
3	5	1	4	●	2	9	11	7	10	6	8	15	17	13	16	12	14	21	23	19	22	18	20
4	2	5	●	3	1	10	8	11	6	9	7	16	14	17	12	15	13	22	20	23	18	21	19
5	3	4	1	2	●	11	9	10	7	8	6	17	15	16	13	14	12	23	21	22	19	20	18
6	7	12	13	18	19	●	1	14	15	20	21	2	3	8	9	22	23	4	5	10	11	16	17
7	6	13	12	19	18	1	●	15	14	21	20	3	2	9	8	23	22	5	4	11	10	17	16
8	10	14	16	20	22	2	4	12	17	18	23	●	5	6	11	19	21	1	3	7	9	13	15
9	11	15	17	21	23	3	5	13	16	19	22	1	4	7	10	18	20	●	2	6	8	12	14
10	8	16	14	22	20	4	2	17	12	23	18	5	●	11	6	21	19	3	1	9	7	15	13
11	9	17	15	23	21	5	3	16	13	22	19	4	1	10	7	20	18	2	●	8	6	14	12
12	18	6	19	7	13	14	20	●	21	1	15	8	22	2	23	3	9	10	16	4	17	5	11
13	19	7	18	6	12	15	21	1	20	●	14	9	23	3	22	2	8	11	17	5	16	4	10
14	20	8	22	10	16	12	18	2	23	4	17	6	19	●	21	5	11	7	13	1	15	3	9
15	21	9	23	11	17	13	19	3	22	5	16	7	18	1	20	4	10	6	12	●	14	2	8
16	22	10	20	8	14	17	23	4	18	2	12	11	21	5	19	●	6	9	15	3	13	1	7
17	23	11	21	9	15	16	22	5	19	3	13	10	20	4	18	1	7	8	14	2	12	●	6
18	12	19	6	13	7	20	14	21	●	15	1	22	8	23	2	9	3	16	10	17	4	11	5
19	13	18	7	12	6	21	15	20	1	14	●	23	9	22	3	8	2	17	11	16	5	10	4
20	14	22	8	16	10	18	12	23	2	17	4	19	6	21	●	11	5	13	7	15	1	9	3
21	15	23	9	17	11	19	13	22	3	16	5	18	7	20	1	10	4	12	6	14	●	8	2
22	16	20	10	14	8	23	17	18	4	12	2	21	11	19	5	6	●	15	9	13	3	7	1
23	17	21	11	15	9	22	16	19	5	13	3	20	10	18	4	7	1	14	8	12	2	6	●

[19]

Figure 9.3. Group table for the symmetric group S_4, where dots are the identity and the odd permutations are colored (transpositions in green, 4-cycles in orange). Can you figure out why some values are in boldface?

representations to help make sense of algebraic structures [4]. The book's accompanying software, Group Explorer is available free online [3].

Programming

If you've read this book from the beginning, you're already familiar with Ed Dubinsky's affinity for getting students to construct new mathematical knowledge by programming the ideas in some rudimentary computer language (for him, typically ISETL). His first project in this vein wasn't in any of the areas of calculus that you've read about—it was abstract algebra. And by his own admission, it was with group theory that programming seemed to help the most [7].

The way the topics of group theory fit together almost begs for a programming approach. Taking a group table as input, the processes of checking for an identity, inverses, and associativity follow an algorithmic structure perfect for a computer. Putting a number of subroutines together might create a program that would find all subgroups of a group and classify each as normal or not. By asking the students not just to have the computer check the many cases, but to program the computer to do that checking (for instance, carefully programming either an existential or universal quantifier), Dubinsky and his colleagues hypothesized that they would internalize the structures of the mathematical objects.

Although they worked with a small number of students, their efforts seemed promising, with students struggling but eventually learning the material. For instance, 65% of students in the programming-intensive class scored nearly perfectly on an interview question about cosets [1]. When it comes to normality, a topic that leaves many students in a fog, only 24% of the programming students did not show any understanding [1]. The process of programming a computer to check the various group-theoretic properties appears to have potential advantages over a traditional lecture-based class.

Teaching Strategic Knowledge

Finally, Keith Weber wanted to know if the strategic knowledge the undergraduates were missing (see Problem 7) was teachable. Is it possible to help students write better proofs by giving them guidance in how one chooses which

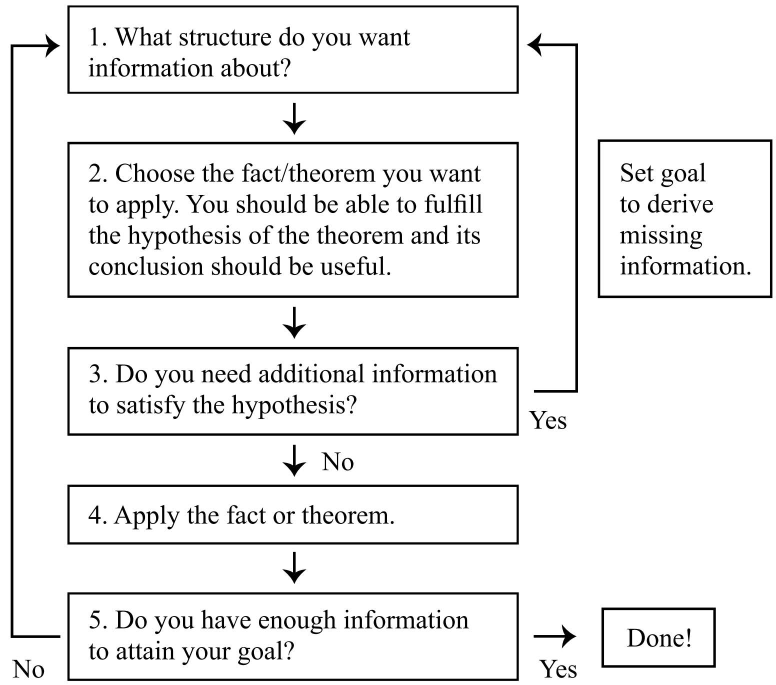


Figure 9.4. Weber's proof procedure for proofs about homomorphisms.

theorems to apply and exactly how one goes about using a theorem? The affirmative answer, at least for some abstract algebra proofs, comes in the form of a study where he taught struggling students a proof-writing heuristic shown in figure 9.4 [22].

While the results were impressive (six students produced valid proofs in 41 of their 48 attempts), Weber cautions that some students may have been able to produce a correct proof without having much understanding of the underlying mathematics. This still lends credence to the idea that helping students with the meta-thinking of proof writing, for instance, exploring what questions they might ask themselves when stuck, might be a promising way to encourage better proof writing.

Now that you have heard these research-based insights into student thinking, we hope that you agree with the sentiment expressed in the chapter's opening paragraph: research on the teaching and learning of abstract algebra can indeed help instructors improve their teaching and, as a result, improve their students' understanding of these interesting and important ideas.

A Good Place to Start

Make sure your students are leveraging their previous knowledge by tying these beautiful mathematical concepts to concrete examples, like constructing the symmetry group on a triangle, early and often.

Bibliography

[1] Mark Asiala, Ed Dubinsky, David M Mathews, Steven Morics, and Asuman Oktaç. Development of students' understanding of cosets, normality, and quotient groups. *The Journal of Mathematical Behavior*, 16(3):241–309, 1997.

[2] Anne Brown, David J DeVries, Ed Dubinsky, and Karen Thomas. Learning binary operations, groups, and subgroups. *The Journal of Mathematical Behavior*, 16(3):187–239, 1997.

[3] *Group Explorer 2* download: groupexplorer.sourceforge.net/. [Online; accessed 9-June-2014]. *Group Explorer 3* (web app) is available here: nathancarter.github.io/group-explorer/index.html

[4] Nathan Carter. *Visual Group Theory*. MAA, 2009.

[5] Julie M Clark, David J DeVries, Clare Hemenway, Denny St John, Georgia Tolias, and Roozbeh Vakil. An investigation of students' understanding of abstract algebra (binary operations, groups and subgroups) and the use of abstract structures to build other structures (through cosets, normality and quotient groups). *The Journal of Mathematical Behavior*, 16(3):181–185, 1997.

[6] Ivan Cnop and Francine Grandsard. Teaching abstract algebra concepts using small group instruction. *International Journal of Mathematical Education in Science and Technology*, 29(6):843–850, 1998.

[7] Ed Dubinsky. personal communication.

[8] Ed Dubinsky, Jennie Dautermann, Uri Leron, and Rina Zazkis. On learning fundamental concepts of group theory. *Educational studies in Mathematics*, 27(3):267–305, 1994.

[9] Orit Hazzan. Reducing abstraction level when learning abstract algebra concepts. *Educational Studies in Mathematics*, 40(1):71–90, 1999.

[10] Orit Hazzan. How students attempt to reduce abstraction in the learning of mathematics and in the learning of computer science. *Computer Science Education*, 13(2):95–122, 2003.

[11] Orit Hazzan and Uri Leron. Students' use and misuse of mathematical theorems: The case of lagrange's theorem. *For the Learning of Mathematics*, 16:23–26, 1996.

[12] Orit Hazzan and Rina Zazkis. Reducing abstraction: The case of school mathematics. *Educational Studies in Mathematics*, 58(1):101–119, 2005.

[13] Sean Larsen. *Supporting the guided reinvention of the concepts of group and isomorphism: A developmental research project*. PhD thesis, Arizona State University, 2004.

[14] Sean Larsen. Reinventing the concepts of group and isomorphism: The case of jessica and sandra. *The Journal of Mathematical Behavior*, 28(2):119–137, 2009.

[15] Sean Larsen, Estrella Johnson, and Joanna Bartlo. Designing and scaling up an innovation in abstract algebra. *The Journal of Mathematical Behavior*, 32(4):693–711, 2013.

[16] Sean Larsen and Elise Lockwood. A local instructional theory for the guided reinvention of the quotient group concept. *The Journal of Mathematical Behavior*, 32(4):726–742, 2013.

[17] Richard A Lesh. The influence of two types of advanced organizers on an instructional unit about finite groups. *Journal for Research in Mathematics Education*, 7(2):87–91, 1976.

[18] Elise Lockwood, Estrella Johnson, and Sean Larsen. Developing instructor support materials for an inquiry-oriented curriculum. *The Journal of Mathematical Behavior*, 32(4):776–790, 2013.

[19] mate2code. Symmetric group s_4: Cayley table — wikimedia commons, 2010. bit.ly/39uRWQt [Online; accessed 19-May-2014].

[20] Annie Selden and John Selden. Errors and misconceptions in college level theorem proving. In *Proceedings of the second international seminar on misconceptions and educational strategies in science and mathematics*, volume 3, pages 457–470. ERIC, 1987.

[21] Keith Weber. Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, 48(1):101–119, 2001.

[22] Keith Weber. Investigating and teaching the processes used to construct proofs. *Research in collegiate mathematics education, VI*, pages 197–232, 2006.

[23] Keith Weber and Lara Alcock. Semantic and syntactic proof productions. *Educational Studies in Mathematics*, 56(2-3):209–234, 2004.

[24] Wikipedia. Category: Properties of groups — Wikipedia, the free encyclopedia, 2014. en.wikipedia.org/wiki/Category:Properties_of_groups [Online; accessed 19-May-2014].

[25] Rina Zazkis and E. Dubinsky. Dihedral groups: a tale of two interpretations. *Research in collegiate mathematics education*, 2:61–82, 1996.