Title

How ‘Natural’ are the Natural Numbers? Inconsistencies Between Formal Axioms and Undergraduates’ Conceptualizations of Number

Author

Relaford-Doyle, Josephine

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

How 'Natural' are the Natural Numbers?
Inconsistencies Between Formal Axioms and Undergraduates' Conceptualizations of Number

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Cognitive Science

by

Josephine Relaford-Doyle

Committee in charge:

Professor Rafael Núñez, Chair
Professor David Barner
Professor Andrea Chiba
Professor David Kirsh
Professor Jeffrey Rabin

2022

The dissertation of Josephine Relaford-Doyle is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

For my Dad.

EPIGRAPH

“The paradox is now fully established that the utmost abstractions are the true weapons with which to control our thought of concrete fact.”

Alfred North Whitehead, 1933

TABLE OF CONTENTS

Dissertation Approval Page……………..……………………………….…….. iii

Dedication…………...…………………………………………………...…… iv

Epigraph………………………………………………………………...…... v

Table of Contents…..………………………………………………..…... vi

List of Figures………..…………………………………………...……….. vii

List of Tables ………………………………………………………………... viii

Acknowledgements……………………..…………………...………..……. ix

Vita……………………………………………………………………...... x

Abstract of the Dissertation …………………………………………………... xi

Introduction……….…………………………………………………………… 1

Chapter 1: Characterizing students' conceptual difficulties with mathematical induction using visual proofs ………………… 11

Chapter 2: Investigating Contexts in Which Undergraduates Demonstrate Reasoning by Informal Mathematical Induction……….…………... 45

Chapter 3: Are There Infinitely Many Natural Numbers? Differential Prominence of the Successor Principle in Math- and Non-Math Majors' Justifications.......................... 71

Chapter 4: Formal Consistency of Undergraduates Number-to-Line Mappings for Large-Magnitude Scales………… 98

Conclusion ……………………………………………………………….. 125

LIST OF FIGURES

Figure 1.1: Visual proofs of theorems that could be formally proven using mathematical induction. ……………………..19

Figure 1.2: Using a visual proof to demonstrate the necessity of a theorem……. 20

Figure 1.3: Participant workspace……………………………………………….25

Figure 1.4: MI-Trained participants overwhelmingly preferred pattern-based explanations, while the majority of MI-Untrained participants provided case-based explanations. ……………29

Figure 1.5: In their tutorial videos some MI-Untrained participants produced images that violated the essential row-column structure of the original figure ……………..30

Figure 1.6: All participants generalized the theorem to nearby cases (a). However, MI-Untrained participants were less likely to doubt that large-magnitude counterexamples to the statement are possible (b). ……32

Figure 2.1: Mean bet amounts by item…………………………………………...58

Figure 2.2: Percent of participants on each item who bet a smaller amount on the far prediction target than on the close prediction target.…………59

Figure 2.3: Percent of participants who bet the maximal amount of $100, by item and target condition. ………………..60

Figure 3.1: Cards arranged haphazardly in front of participant…………………80

Figure 3.2: Cards rearranged in order of magnitude……………………………..82

Figure 3.3: Initial justifications provided by participants when asked to prove that there are infinitely many natural numbers.……………….85

Figure 3.4: Proportion of participants referring to the Successor Principle during the interview. ………..………86

Figure 4.1: Mean participant responses and lines-of-best-fit by Condition…….107

Figure 4.2: Summary of Tukey post-hoc tests for differences in slopes by condition. …………………………109

Figure 4.3: Frequency distributions of slopes of individual participants' least-squares regression lines.…………………..110

Figure 4.4: Mean responses by condition………………………………………112

LIST OF TABLES

Table 1.1: Scoring criteria and sample responses for rating………………….. 27
participants' resistance to the possibility of large-magnitude counterexamples.

Table 2.1*:* Evidence sets and predictions for each of the six……..…………. 50
mathematical items, and one example non-mathematical item.

Table 4.1: Fit of linear and logarithmic models for all conditions……….…….108

ACKNOWLEDGEMENTS

I am deeply grateful to Professor Rafael Núñez, whose guidance was invaluable to the development of these ideas, and whose openness, support, and kindness made the work joyful.

Many thanks to the members of the Committee for meeting me with enthusiasm and enriching this project through their perspectives, questions, and expertise.

This dissertation was supported by a National Academy of Education/Spencer Foundation Dissertation Fellowship. I am grateful to the NAEd/Spencer Foundation for their support of this project and the mentorship I received through the program.

Portions of the Introduction and Chapter 3 are reprints of selected material as it appears in the chapter *Beyond Peano: Looking Into the Unnaturalness of Natural Numbers* in *Naturalizing Logico-Mathematical Knowledge* 2018. Relaford-Doyle, Josephine; Núñez, Rafael, Routledge, 2018. The dissertation author was the primary author of this chapter.

Chapter 1, in full, is a reprint of the material as it appears in *The International Journal of Research in Undergraduate Mathematics Education*, 2021, Relaford-Doyle, Josephine; Núñez, Rafael, 2021. The dissertation author was the primary investigator and author of this paper.

VITA

2007	Bachelor of Arts, University of California, Berkeley
2009	Master of Arts, Claremont Graduate University
2008-2013	Teacher, Chaffey Joint Union High School District
2014-2019	Teaching Assistant/Instructor of Record, University of California San Diego
2022	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Relaford-Doyle, J., & Núñez, R. (2018). Beyond Peano: Looking into the unnaturalness of natural numbers. In *Naturalizing Logico-Mathematical Knowledge* (pp. 234-251). Routledge.

Relaford-Doyle, J., & Núñez, R. (2019). Can a picture prove a theorem? Using empirical methods to investigate visual proofs by induction. *Advances in Experimental Philosophy of Logic and Mathematics*, (pp. 95-121). Bloomsbury.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science?. *Nature human behaviour*, *3*(8), 782-791.

Relaford-Doyle, J., & Núñez, R. (2021). Characterizing students' conceptual difficulties with mathematical induction using visual proofs. *International Journal of Research in Undergraduate Mathematics Education*, *7*(1), 1-20.

FIELDS OF STUDY

Major Field: Cognitive Science

Cognitive Science of Mathematics
Professor Rafael Núñez

ABSTRACT OF THE DISSERTATION

How 'Natural' are the Natural Numbers?
Inconsistencies Between Formal Axioms and Undergraduates' Conceptualizations of Number

by

Josephine Relaford-Doyle

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2022

Professor Rafael Núñez, Chair

It is widely assumed within developmental psychology that spontaneously-arising conceptualizations of natural number – those that develop without explicit mathematics instruction – match the formal characterization of natural number given in the Dedekind-Peano Axioms (e.g., Carey, 2004; Leslie et al., 2008; Rips et al., 2008). Specifically, developmental researchers assume that fully-developed conceptualizations of natural number are characterized by knowledge of a starting value 'one' and understanding of the successor principle: that for any natural number *n*, the next natural number is given by *n*+1. This formally-consistent knowledge

is thought to emerge spontaneously as children learn how to count, without the need for formal instruction in the axioms of natural number. In developmental psychology, the assumption that spontaneously-arising conceptualizations of natural number match the formal characterization has been taken as unproblematic, and has not been subject to empirical investigation.

In this dissertation I seek to provide a more rigorous, thorough, and empirically-grounded characterization of spontaneously-arising natural number concepts. What do we know about natural number, without being explicitly taught? And to what extent are these conceptualizations consistent with the formal mathematical definition given in the Dedekind-Peano Axioms? In the four studies that comprise the dissertation I use an open-ended problem-solving task, a computer-based judgment task, a number-line estimation task, and semi-structured interviews to explore undergraduates' conceptualizations of natural number. I find evidence that undergraduates who have not received explicit instruction in the axioms of natural number often recruit conceptualizations that are at odds with the formal characterization given in the Dedekind-Peano Axioms, and that many students demonstrate formally-consistent reasoning in only the simplest and most foundational mathematical contexts. Taken together, these results call into question the assumption that formally-consistent notions of natural number arise spontaneously from mastery of counting. Based on these findings, I make recommendations for future research into number concept development and discuss practical implications for classroom instruction for concepts that build on the natural number system (for instance, mathematical induction), as well as mathematics teacher development.

INTRODUCTION

Motivation and Theoretical Approach

How people come to develop mathematical knowledge has long been a central question in philosophy, psychology, and the cognitive sciences. Mathematics presents a challenge to those interested in human concepts: unlike the objects of daily life, the objects of formal mathematics are idealized and abstract – while we can think about numbers, lines, and sets, we never directly experience or interact with these entities in our daily lives. How, then, do we come to have such precise knowledge of these objects? This question has driven an extensive body of research, and carries important practical implications for mathematics education and teacher development.

In order to understand how this question has been approached, it is critical to acknowledge the cultural context, and specifically the presence of widespread, popularized, and highly influential notions about the nature of mathematics itself. Lakoff and Núñez (2000) identified a "mythology" of mathematics, a set of deeply-held cultural beliefs that permeate how mathematics is viewed in modern society. According to this mythology, mathematics is *real* and has an objective existence independent of human minds; it is built into the physical universe, providing structure to the natural world and therefore our experiences within it. These notions are reflected and reinforced in popular culture, including popular science; take, for instance, a recent article on the website LiveScience "What is mathematics?"[1] which asserts, "Math is all around us, in everything we do. It is the building block for everything in our daily lives." One also frequently encounters these ideas in educational materials, often in an attempt to make mathematics relevant and interesting for students (e.g., "Think You Don't Need Math? Think

[1] Hom, E. J., & Gordon, J. (2021, November 11). *What is Mathematics?* LiveScience. Retrieved January 12, 2022, from https://www.livescience.com/38936-mathematics.html

Again…..It's All Around You!"[2]) Crucially, these notions have also permeated scientific research on mathematical cognition. In influential works with titles like *The Math Gene* (Devlin, 2000), *The Mathematical Brain* (Butterworth, 1999), *The Number Sense* (Dehaene, 1997), and *A Brain for Numbers* (Nieder, 2019), researchers have argued that humans are "hard-wired" for mathematics, primed and ready to access the mathematics that is built into our everyday experiences in the natural and physical world.

What this and much other existing research into mathematical cognition fails to take into account are the ways in which formal mathematical objects are fundamentally *different from*, and in fact *impossible in*, the physical world that we inhabit. Spurred by the assumption that mathematics is a building block of our experience in the natural world, researchers over-use labels like "mathematical" and "numerical", applying these terms to abilities and behaviors that may not actually meet the criteria for these types of knowledge (Núñez, 2017). In this way, the mythology described above can act as a roadblock to developing a full and accurate understanding of how we come to develop the abstract, idealized concepts that constitute mathematics.

In this dissertation, I examine undergraduates' reasoning on a variety of tasks involving natural number in order to subject to rigorous empirical research the idea that mathematical knowledge - knowledge that is consistent with formal mathematical characterizations - arises spontaneously from everyday experience. Rather than assuming mathematics to be pre-existing, "built in" to the physical world, and accessible through everyday life, I approach mathematics as

---

[2] Hofstaedter, M. (2020, September 21). *Think you don't need math? think again.....it's all around you!* InspireMyKids. Retrieved January 12, 2022, from https://inspiremykids.com/think-you-dont-need-math-think-again-its-all-around-you/

a highly peculiar human conceptual system which has emerged as the result of particular cultural needs, which are often internal to mathematics itself (Lakoff & Núñez, 2000; Wilder, 1981). This stance acknowledges the ways in which the objects of mathematics differ from the objects of everyday life, and thus allows for a more thorough exploration of how we come to acquire knowledge that is truly mathematical in nature. I take as a case study one of the most basic and foundational of mathematical objects: the natural numbers {1, 2, 3…}. While previous research has explored differences between everyday concepts and formal mathematics in high-level domains (Marghetis & Núñez, 2013, Núñez, 2005), it has gone largely unconsidered the ways in which students' conceptualizations of even the simplest mathematical objects may differ from their formal characterizations. In fact, a large and productive research program in developmental psychology has taken it as *given* that formally-consistent knowledge of the natural number system emerges spontaneously from everyday counting practices. This assumption has been regarded as unproblematic, and has shaped the way studies are designed and results are interpreted in studies of number concept development. But, is this assumption warranted? Are the numbers that we use in counting the same as the formal set of natural numbers?

Counting Numbers and Natural Numbers

The case of natural number is a particularly rich domain to explore the relationship between everyday experience and mathematical knowledge. The numbers 1, 2, 3, … are ubiquitous in our daily lives from the time we are children. Speakers of English (or any other language with a numeral system) generally learn how to count by around age five; this happens spontaneously with language exposure and does not require any formal mathematical education. In developmental psychology, however, a child who has learned to count is assumed to have done more than simply master a particular cultural practice; they are said to have developed concepts of *natural number* (a label pulled directly from formal mathematics). Furthermore, in all major lines of research it is assumed that these spontaneously-arising concepts of natural number – those that develop without formal classroom instruction – match the formal mathematical definition given in the Dedekind-Peano axioms (e.g. Carey, 2004; Leslie, Gelman, & Gallistel, 2008; Rips, Bloomfield, & Asmuth, 2008). Specifically, it is assumed that concepts of natural number are characterized by a unique starting value 'one', and knowledge of the *successor principle*: for any natural number n, the next natural number is given by n+1 (and named by the next item in the count list). This principle is taken to be the key to "mature" conceptualizations of natural numbers, allowing children to develop exact representations of large numbers (Sarnecka & Carey, 2008), eventually understanding "the logic of natural number" (Cheung, Rubenson, & Barner, 2017) and recognizing the natural numbers as an unbounded and countably infinite set (Rips, Bloomfield, & Asmuth, 2008). The assumption that our spontaneously-arising concept matches the mathematical formalism is taken as unproblematic, with some psychologists even crediting Dedekind with giving researchers "a firm idea about the constituents of the natural number concept" (Rips, Bloomfield, & Asmuth, 2008, p. 640).

The successor principle itself is simple: for any natural number n, the next natural number is given by n + 1. However, as in the axioms, the successor principle carries important and deep implications about natural numbers, including,

I. The natural numbers are unbounded, since one can always be added to produce the next number.

II. Every natural number has a unique successor, and adding one is the <u>only</u> way to generate the next natural number.

III. The same relation holds between any natural number and its successor (e.g., the pairs 3 & 4, 8 & 9; 9,381,763 & 9,381,764; and 1023 & 1023 + 1 are all governed by the same +1 relation).

IV. No natural number is more ‘natural’ than any other.

These properties distinguish the natural numbers from the counting numbers. The natural numbers extend infinitely, while the set of nameable counting numbers—the numbers a person could actually recite in a count sequence—is finite. In our experiences with counting numbers, not all numbers are alike. The smallest counting numbers ‘one’ through ‘nine’—the prototypical counting numbers (PCNs) — are frequently encountered and highly familiar (Núñez & Marghetis, 2014). The PCNs can be written using single-digit numerals, and in English the associated number words are monomorphemic. Larger counting numbers (like, say, 197 or 5,823) are encountered far less frequently, require composite notation, and are lexically more complex. In the domain of natural number, however, these differences cease to be relevant. Any number that can be reached by successive additions of one, no matter how large, is a natural number; there are no ‘degrees’ of naturalness.

In addition to these differences, it is important to note that the counting numbers and the formal set of natural numbers came into existence thousands of years apart, in order to serve vastly different cultural needs. Counting practices began to emerge in various parts of the world about 6,000 years ago. In some regions, counting developed concurrently with the rise of cities, fulfilling a new cultural need to keep track of increasingly complex trade and commerce. In other parts of the world, counting practices developed outside of agriculture and complex societies; however, common to all cases is that counting emerged from a cultural need for exact quantification to serve practical purposes including record-keeping, economic activity, and measurement. The formal characterization of the natural numbers, on the other hand, did not exist until the 19th century. Unlike counting numbers, the natural numbers were brought into being in order to serve specific goals that were internal to the mathematical community of the time (Wilder, 1981) - an elite, exclusive group composed almost entirely of educated European men. After the discovery of various counterintuitive results, mathematicians of the 19th century held a common goal: to establish rigorous logical foundations for all of mathematics. Developing a formalization of the natural numbers was needed in order to establish a firm logical foundation for arithmetic and ensure that formal number theory was consistent and complete. Mathematicians and philosophers including Gottlob Frege, Georg Cantor, Charles Sanders Peirce, Giuseppe Peano, and Richard Dedekind dedicated themselves to this task, and after being published in 1889 the Dedekind-Peano axioms were adopted as the formal characterization of the natural numbers within the mathematical community. Thus, while it may be tempting to see the axioms of natural numbers as simple or obvious, it is important to remember that these axioms were created with great effort and as the result of a specific cultural need within 19th century mathematics (and in fact, didn't exist until hundreds of years *after* the invention of calculus!).

In sum, the counting numbers and formal set of natural numbers have different properties and came into existence in order to serve completely different cultural needs. With this analysis in place, the assumption that formally-consistent knowledge of the natural number system emerges spontaneously from mastery of counting is not only surprising, but in fact quite controversial.

Limitations of Existing Research and Contribution of Dissertation

The assumption that spontaneously-arising concepts of natural number match the formal mathematical definition assumption has severely limited existing empirical work in key ways.

First, researchers have been interested primarily in when (not whether) formally-consistent conceptualizations emerge; as a result, empirical studies have focused almost exclusively on very young children. Moreover, studies often test children's understanding of the successor principle within the PCNs, and thus examine only the smallest, most familiar numbers, generally no greater than nine (e.g., Sarnecka & Carey, 2008). It is assumed that children's knowledge of the successor principle eventually extends to "all possible numbers", allowing adults to grasp "the logic of natural number", but these conclusions are drawn from studies that never test children older than six or numbers larger than 100 (e.g. Cheung, Rubenson, & Barner, 2017). Studies rarely investigate children's understanding of high-level implications of the successor principle (for instance, that the natural numbers extend infinitely), and those that do rely on interview protocols that include leading questions and ignore responses suggesting alternative conceptualizations (see e.g. the protocol used in Cheung, Rubenson, & Barner, 2017 and Hartnett & Gelman, 1998). Finally, in formal mathematics the Dedekind-Peano axioms allow for *mathematical induction*, a powerful proof method that can be used to demonstrate that a given property is *necessarily* true for all natural numbers. Researchers have neglected to check whether our everyday conceptualization of natural number – assumed to mirror the axioms – allows for a comparably powerful form of reasoning.

In this dissertation I use a variety of empirical methods to more thoroughly and rigorously characterize conceptualizations of natural number in university undergraduates. In addressing the limitations outlined above, this work seeks to illuminate the ways in which the fully-developed natural number concept – the target of existing developmental accounts – may in fact be substantially different from the formal characterization. The dissertation addresses two key questions through four empirical studies:

(1) To what extent are spontaneously-arising concepts consistent with "the logic of natural number"? Do these concepts include knowledge of higher-level implications of the successor principle like I-IV (above)?

(2) How robust are spontaneously-arising natural number concepts? Can everyday conceptualizations be leveraged to perform feats of reasoning comparable to mathematical induction?

Chapters 1 and 2 address question (2), examining contexts in which university undergraduates demonstrate reasoning by an informal version of mathematical induction. In Chapter 1, I use an open-ended problem solving task and semi-structured interview to assess whether undergraduates recognize the necessity of a simple mathematical theorem when provided with visual evidence. Chapter 2 uses a web-based study to assess undergraduates' reasoning in a wider variety of mathematical contexts, with the intention of determining when, if ever, students demonstrate reasoning by informal mathematical induction. In Chapters 3 and 4 I turn to question (1), assessing to what extent undergraduates recognize and make use of the successor principle on tasks involving natural number. Chapter (3) uses an open-ended activity and semi-structured interview to explore how undergraduates justify their knowledge that the natural numbers are infinite, and specifically how likely they are to refer to the successor principle in these justifications. In Chapter 4 I make use of a well-known paradigm, the bounded number-line task, to examine undergraduates' number-to-space mappings for large-magnitude scales and assess to what degree these mappings are consistent with a number system universally governed by the successor principle. Finally, in the Conclusion I summarize findings and discuss implications for research, mathematics education, and mathematics teacher development, as well as how this project bears on the “mythology” of mathematics as real, objective, and built into our everyday lives.

A portion of the Introduction is a reprint of selected material as it appears in the chapter *Beyond Peano: Looking Into the Unnaturalness of Natural Numbers* in *Naturalizing Logico-Mathematical Knowledge* 2018. Relaford-Doyle, Josephine; Núñez, Rafael, Routledge, 2018. The dissertation author was the primary author of this chapter.

References

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59-68.

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive psychology*, *92*, 22-36.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.

Devlin, K. J. (2000). *The math gene: How mathematical thinking evolved and why numbers are like gossip*. New York: Basic Books.

Hartnett, P., & Gelman, R. (1998). Early understandings of numbers: Paths or barriers to the construction of new understandings? *Learning and instruction, 8*(4), 341-374.

Hofstaedter, M. (2020, September 21). *Think you don't need math? think again.....it's all around you!* InspireMyKids. Retrieved January 12, 2022, from https://inspiremykids.com/think-you-dont-need-math-think-again-its-all-around-you/

Hom, E. J., & Gordon, J. (2021, November 11). *What is Mathematics?* LiveScience. Retrieved January 12, 2022, from https://www.livescience.com/38936-mathematics.html

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from.* New York: Basic Books.

Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in cognitive sciences, 12*(6), 213-218.

Marghetis, T., & Núñez, R. (2013). The motion behind the symbols: A vital role for dynamism in the conceptualization of limits and continuity in expert mathematics. *Topics in cognitive science*, *5*(2), 299-316.

Núñez, R. (2005). Creating mathematical infinities: Metaphor, blending, and the beauty of transfinite cardinals. *Journal of Pragmatics*, *37*(10), 1717-1741.

Núñez, R. (2017). Is there really an evolved capacity for number? *Trends in cognitive sciences, 21*(6), 409-424.

Núñez, R., & Marghetis, T. (2015). Cognitive linguistics and the concept(s) of number. In R.C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 377-401). New York: Oxford University Press.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and brain sciences, 31*(6), 623-642.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*(3), 662-674.

Wilder, R.L. (1981). *Mathematics as a cultural system.* Oxford: Pergamon.

CHAPTER 1

Characterizing students' conceptual difficulties with mathematical induction using visual proofs

Introduction

Studies have repeatedly documented students' considerable difficulties in learning mathematical induction. In particular, various studies have revealed that students who master the procedure of mathematical induction (that is, students who can perform the base case and inductive step) often lack conceptual understanding of these steps, why they are necessary, and why they allow for the conclusion that a theorem is *necessarily* true for all natural numbers (Baker, 1996; Harel, 2001; Lowenthal & Eisenberg, 1992; Woodall, 1981). This finding – that students have *procedural* but not *conceptual* understanding of mathematical induction – seems to be at odds with developmental psychologists' claims that children as young as five can engage in an informal version of mathematical induction before they receive any training in the formal procedure (Baroody, 2005; Smith, 2003).

This apparent inconsistency may be due to limitations in the existing research in mathematical induction, both formal and informal. Formal mathematical induction consists of a specific algebraic procedure, and studies of students' understanding of the formal method have a hard time disentangling procedural from conceptual obstacles. Additionally, the small number of studies claiming to find evidence of untrained children's ability to engage in informal mathematical induction haven't convincingly demonstrated that children understand the necessity of their conclusions, and thus may in fact bear on standard inductive reasoning.

In this study we use a new method to assess both trained students' understanding of mathematical induction, as well as untrained students' capacity to engage in an informal version of mathematical induction. We use a 'visual proof by induction' – an image designed to demonstrate a theorem that could be formally proven using mathematical induction. The image is simple and free of algebraic notation, thus allowing us to explore (1) whether students who *are* familiar with formal mathematical induction can transfer their knowledge to a new, non-algebraic representational system (where the algebraic procedure of mathematical induction no longer applies), and (2) whether students who are *not* familiar with the formal method spontaneously use the image to recognize the necessity of the theorem it represents, and if not, what conceptual obstacles they encounter.

A Note on Terminology

As there is some overlap in the words used to describe the various forms of proof and reasoning relevant to this article, a brief clarification of terms is necessary.

Formal mathematical induction is a formal mathematical proof technique that can be used to demonstrate that a particular property holds for all natural numbers $n = 1, 2, 3, \ldots$. A formal proof by mathematical induction has two steps: first, in the base case, it is shown that the property holds for some initial value, typically $n = 1$. Then, in the inductive step, it is shown that if the property holds for some arbitrary value k, then it must also hold for its successor $k + 1$. By the Axiom of Induction (one of the Dedekind-Peano axioms of natural number), it follows that the property is true of all (infinitely many) natural numbers.

Researchers have noted that formal mathematical induction has both a *procedural* component and a *conceptual* component (Baker, 1996; Harel, 2001; Lowenthal & Eisenberg, 1992; Woodall, 1981). In this paper, *procedural knowledge of formal mathematical induction* refers to an ability to perform or comprehend a proof by formal mathematical induction; that is, a student with procedural understanding could successfully carry out the base case and the inductive step (which, for our purposes in this paper, would entail performing the correct algebraic manipulations). *Conceptual knowledge of formal mathematical induction* refers to a deeper comprehension of the proof technique, including understanding why the base case is necessary, the role of the inductive step, and why together those two steps allow for the conclusion that the property holds for all natural numbers. Unlike procedural understanding, conceptual understanding of formal mathematical induction isn't linked to any particular procedure; it wouldn't rely on algebraic manipulations of k and $k + 1$, but rather it would entail a more general understanding that the proof technique involves establishing the truth of a base case and an invariance of the relationship between successive instances. In principle, this type of conceptual understanding could be transferred to different, non-algebraic representational systems.

Despite its name, formal mathematical induction is actually a form of deductive reasoning; based in the axioms of natural number, it demonstrates the necessity of the result for all natural numbers, such that counterexamples are impossible. This is very different than everyday *inductive reasoning*, which involves generalizing a rule from a finite (and usually quite limited) number of observed cases (for instance, one might see a robin, a pigeon, and a dove and conclude that all birds can fly). The resulting *inductive generalization* is distinct from a formal proof by mathematical induction in that it remains flexible to the possibility of counterexamples (in our example, penguins are still birds despite the fact that they can't fly). Inductive reasoning is commonplace in everyday life; formal mathematical induction is a formal technique that requires explicit instruction in mathematics to master.

Finally, throughout this paper we will refer to a distinct form of reasoning, *informal mathematical induction*. Informal mathematical induction refers to a type of reasoning in the domain of natural number in which the reasoner generalizes a rule based on observed cases (similar to inductive reasoning), *and* recognizes the necessity of the result such that counterexamples in the natural numbers are recognized as impossible (as in formal mathematical induction). This form of reasoning is called *informal* because it does not require that the reasoner see or perform a formal proof of the result.

Background

Formal mathematical induction is a notoriously difficult method for students of all levels, including pre-service teachers, to learn (Avital & Libeskind, 1978; Ernest, 1984; Fischbein & Engel, 1989; Movshovitz-Hadar, 1993; Stylianides et al., 2007). In particular, studies have repeatedly shown that students may develop procedural fluency while still lacking conceptual understanding of the proof method; in other words, they may be able to successfully carry out the base case and inductive step, but fail to understand the meaning of these steps or why they are necessary (Baker, 1996; Harel, 2001; Lowenthal & Eisenberg, 1992; Woodall, 1981).

Formal mathematical induction always consists of a specific procedure, and so when a student encounters difficulty it can be hard to assess whether it is procedural or conceptual in nature. For instance, Baker (1996) analyzed videos of advanced secondary and undergraduate students writing and analyzing formal proofs by induction, and characterized procedural and conceptual knowledge as follows: “Procedural knowledge was demonstrated by recognizing a missing base case, recognizing correctly argued proofs, and identifying the elements of a proof by mathematical induction. Conceptual knowledge was demonstrated by identifying the need for multiple base cases and conceptually describing mathematical induction” (pg. 7). The line between procedural and conceptual is blurry here; for instance, someone may recognize a missing base case *either* because they know that it’s Step 1 of the formal mathematical induction *procedure*, or because they understand that without the base case the inductive step cannot actually demonstrate the truth of the theorem for all natural numbers. Labeling that knowledge as “procedural” obscures the potentially rich conceptual understandings that may be at play.

The difficulty of disentangling procedural from conceptual knowledge is apparent in how Baker classifies particular students’ performance on his task. He presents two examples of student-generated descriptions of formal mathematical induction (pg. 14):

> Student 1: “First you show that the statement is true for the first number P_1. Then you assume it is true for any number k and show that you can get to the next number P_{k+1}.”
>
> Student 2: “1. Prove base case. 2. Prove that for any arbitrary starting point, if that point gives a true value then the next consecutive point also gives a true value.”

Baker classifies Student 1 as demonstrating procedural understanding, and Student 2 as having shown conceptual understanding. Is this distinction valid? Both students have correctly described how to perform a proof by mathematical induction; while Student 2 may have used more sophisticated language, their response is not qualitatively different than that of Student 1. But this raises the question: how can one demonstrate purely conceptual understanding of formal mathematical induction, when the method itself *is* a procedure?

Existing studies of students' difficulties learning formal mathematical induction ask participants to read, produce, or provide explanations of formal proofs. Thus, these studies have by necessity focused on students who have received at least some training in the formal method of mathematical induction. Another line of research that has received less attention has examined untrained children's ability to engage in an informal version of mathematical induction. An operational definition of informal mathematical induction is given by Smith (2003), who characterizes it a type of reasoning which entails observing a base equality or inequality, assessing universality about number, and gauging necessity about number. In other words, in informal mathematical induction the reasoner (a) observes one or more particular cases of the theorem, (b) generalizes the theorem to all natural numbers, and (c) recognizes that the theorem is *necessarily* true of all numbers.

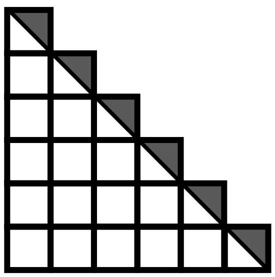
Crucially, (c) recognition of mathematical necessity distinguishes informal mathematical induction from everyday inductive reasoning. In inductive reasoning, the reasoner generalizes a rule based on observed cases; for instance, after multiple encounters with red apples a child might come to believe that all apples are red. Importantly, however, this generalization remains flexible to the existence of counterexamples; it is possible that there are apples that are *not* red, and should the child come across one they would update their rule appropriately. Informal mathematical induction, on the other hand, involves recognizing the necessity of the result, such that counterexamples are impossible; the generalization truly applies to *all* possible cases, without exception. At the same time, informal mathematical induction is distinct from formal mathematical induction in that the reasoner need not provide an explicit justification or proof of the necessity of the result.

Some developmental psychologists have claimed that children as young as 5-7 years old can engage in informal mathematical induction, which is somewhat surprising given the significant *conceptual* difficulties that older students face when learning the formal method. This is also a surprising claim in light of various studies that have repeatedly shown that children don't develop an understanding of logical necessity until age 8-11 (e.g., Miller et al., 2000; Morris & Sloutsky, 2001). The evidence supporting the existence of informal mathematical induction in young children is quite limited. Smith (2003) claims to have found that children as young as 5-7 years old can reason by informal mathematical induction. In his study, children were presented with two containers, either both empty or one empty and the other containing one item. First, each child was asked to add one item at a time to each of the containers; this was then repeated. After observing the results of a few iterative additions, the child was asked various questions about the results of hypothetical additions to the boxes (for instance, "If you add any number here and the same number to that, would there be the same in each or more in one than the other?"). Smith found that the majority of children responded correctly; they *generalized* that adding the same number to two equals gives the same result. Next Smith assessed recognition of necessity by asking, "Does there have to be same number in each, or not?" Fewer than half of the children answered this question correctly, providing only very limited evidence that children recognized the necessity of their generalization. Moreover, as Baroody (2005) notes, simply answering such a yes-or-no question correctly doesn't necessarily indicate that the child is convinced of the necessity of the outcome, so even the correct responses don't provide compelling evidence for recognition of necessity. Thus, Smith's study doesn't actually provide solid evidence that understanding of necessity – a defining feature of informal mathematical induction – is present in children. Instead, the study bears more on inductive reasoning in the domain of natural numbers (Rips et al., 2008).

Baroody (2005; Baroody et al., 2013) makes a case for informal mathematical induction with the story of a kindergartener named Nikki. Baroody asks Nikki what the largest number is, and the girl responds, “A million.” He then asks her what number comes after a million, and after a moment’s thought she replies, “A million and one.” He asks what number comes after a million and one, and the girl responds, “A million and two.” He asks the same question again, and the girl answers, “There is no largest number.” Baroody says that Nikki’s reasoning demonstrates a primitive, informal version of mathematical induction; indeed, that it is informal mathematical induction that allows Nikki to “comprehend infinity.” These are compelling claims, but this story provides no evidence that Nikki understands the necessity of her answer. Her response – while consistent with informal mathematical induction – is also consistent with everyday inductive reasoning.

Visual Proofs by Induction

In summary, previous studies of students’ understanding of formal mathematical induction are limited in two key ways. First, in studies using formal proofs the line between *procedural* and *conceptual* difficulties can be hard to define, making it unclear where the students’ difficulties may be originating. Second, studies of informal mathematical induction have failed to conclusively demonstrate that children recognize the necessity of the result, and thus do not make a clear distinction between informal mathematical induction and everyday inductive reasoning. In this study we attempt to address both of these issues by examining undergraduates' understanding of a *visual proof by induction* – an image that can be used to demonstrate that a particular theorem is necessarily true, but which does not explicitly refer to any specific algebraic procedure. Two examples of such visual representations are given in Figure 1.1.

(a) 

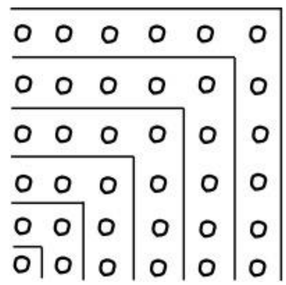
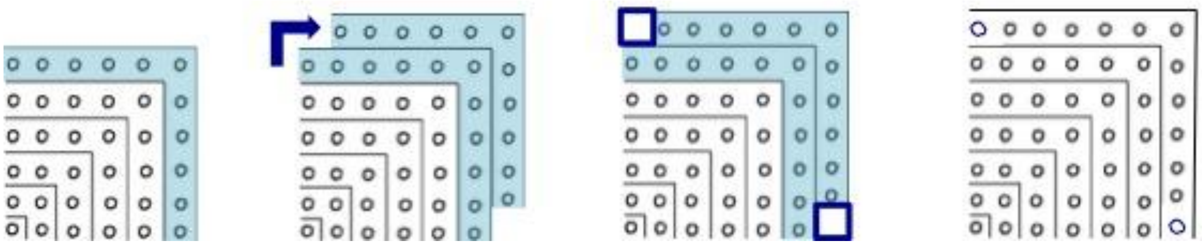
(b) 

Figure 1.1: Visual proofs of theorems that could be formally proven using mathematical induction. (a): $1 + 2 + 3 + \ldots + n = (n^2 + n)/2$. (b): $1 + 3 + 5 + \ldots + (2n – 1) = n^2$. Figures adapted from Brown (1997).

These images display a finite number of particular cases of the theorems they are intended to demonstrate, and so it could be that they function similarly to a set of examples presented numerically: they may serve as the basis for a standard inductive generalization, allowing the viewer to conclude that the theorems are *probably* true for all natural numbers. However, these images contain additional structure that is *not* present in numerical examples, and which could be exploited to show that there can be no counterexamples to the theorem in the natural numbers. To demonstrate this, consider the image in Figure 1.1(b), which shows that the sum of the first n odd numbers is equal to n^2. The image shows only the first six cases of this theorem. However, the structure of the image provides evidence that the pattern will *necessarily* continue to every natural number. Specifically, the square shape of the image is preserved if and only if the next layer contains the next odd number of dots. Figure 1.2 details one way of demonstrating this using the image; while this argument wouldn't be accepted as a formal proof, it establishes that the pattern necessarily continues and thus could be considered a rigorous demonstration of the theorem.



(a) (b) (c) (d)

Figure 1.2: Using a visual proof to demonstrate the necessity of a theorem. (a) To construct the next layer, begin by considering the current outermost later. (b) Make a copy of this layer, and move it one unit up and one unit right. (c) This results in two open positions that must be filled in order to maintain the square shape. (d) Since the original outermost layer contained an odd number of dots, and since the difference between consecutive odd numbers is exactly 2, the next layer must contain the next consecutive odd number of dots. Thus, adding the next consecutive odd number will necessarily result in the next square.

The status of images such as these in mathematics is controversial; while some argue that some images can act as stand-alone proofs (e.g., Borwein & Jörgenson, 2001; Brown, 2008), most mathematicians and philosophers of mathematics would reject them as a valid means of mathematical justification (for a discussion of the role of visualization in proof, see Giardino, 2010; Hanna, 2000). In this article, we are neutral as to what is the status of these images in mathematical justification; it is not of central concern whether they should or should not be accepted as valid mathematical proofs. Instead, we consider them examples of "generic proof by figurate number" (Kempen & Biehler, 2019), in that they reduce the level of abstraction of the mathematical theorem, making them potentially accessible to viewers with no particular training in formal mathematics while still providing enough structure to demonstrate the necessity of the theorem.

In this study we use the visual proof in Figure 1.1(b) to investigate students' conceptual understanding of mathematical induction. As described above, the visual proof can be used to arrive at a comparable result as a formal proof by induction; that is, it can demonstrate the necessity of a theorem and the impossibility of counterexamples within the natural numbers. However, visual evidence uses an entirely different representational system than formal mathematics, and so the *procedure* of formal mathematical induction no longer applies (i.e., there is no algebraic notation like k and $k + 1$). Instead, in order to use a visual proof as justification of a general theorem, the viewer must apply a conceptual understanding of mathematical induction by recognizing that it requires establishing the truth of some initial instance and showing that the relationship between successive instances is invariant. By examining the conclusions that people draw from the image and the ways in which they do and do not use the visual proof to justify the theorem we can assess the extent to which students' understanding of formal mathematical induction depends on the specific algebraic procedure, and characterize some genuinely *conceptual* difficulties with formal mathematical induction that may have gone overlooked in previous studies. Specifically, we can address the following research questions:

- RQ1. How do participants who are unfamiliar with formal mathematical induction interpret the image? Specifically, do they recognize the necessity of the theorem it represents? If not, what conceptual obstacles keep them from doing so?
- RQ2. How do participants who *are* familiar with formal mathematical induction use the image to justify the theorem? Can they transfer their knowledge of mathematical induction to this new representational format, or is their knowledge of mathematical induction intimately linked to the algebraic procedure required by the formal method?

In regards to RQ1, we would expect all university undergraduates, regardless of familiarity with formal mathematical induction, to recognize the key features of the image (the odd numbers in each layer, and the squares formed at each iteration). We would also expect all undergraduates to recognize that the image represents the first six cases of a pattern that could be extended. Finally, in consideration of claims (Baroody, 2005; Smith, 2003) that even young children can engage in informal mathematical induction, we would expect undergraduate participants to do the same; that is, we expect that even participants unfamiliar with formal mathematical induction would recognize the necessity of the theorem represented by the image, and thus conclude that the theorem has no counterexamples in the natural numbers.

Previous work has shown that students who have been trained in formal mathematical induction frequently possess procedural (but not conceptual) understanding of the proof method (Baker, 1996; Harel, 2001; Lowenthal & Eisenberg, 1992; Woodall, 1981); based on this, we predict for RQ2 that undergraduates who are familiar with formal mathematical induction might have difficulty transferring this knowledge to a new representational format. If this is the case, we would expect to see these participants recognizing the necessity of the theorem but referring to formal mathematical induction to provide justification rather than using the image to provide a demonstration of the theorem.

Method

All participants were undergraduate students from a major research university and were tested individually. We recruited participants from two distinct student populations. The first group (n = 22, 11 males and 11 females) consisted of students (mostly mathematics majors) who had taken and received at least a B- in Mathematical Reasoning, an upper-division mathematics course that covers various proof techniques including formal mathematical induction. While there is some variation between class sections, instructors in this course cover a variety of examples of formal mathematical induction (including base cases other than $n = 1$ and strong induction) and link the proof technique to the Axiom of Induction. As these students had all received university-level instruction in formal mathematical induction (MI), we refer to this group as MI-Trained. Our second group of participants (n = 17, 9 males and 8 females) was recruited through the general subject pool and consisted of students with a variety of majors including psychology, cognitive science, and linguistics. None of these students had taken the Mathematical Reasoning course, or any other university-level course covering formal mathematical induction, and so we refer to this group as MI-Untrained. Importantly, all our participants were highly educated adults at a prestigious university, such that we expected them to be familiar with the mathematical concepts relevant to the task.

Open-Ended Explanation Task

Procedure. Participants were given a worksheet with the visual proof in Figure 1.1(b), and were instructed to explain how the picture was related to the statement, "The sum of the first n odd numbers is equal to n^2."[3] In the first phase of the study we asked each participant to create a "tutorial video" in which they explained their reasoning to an imagined third-party audience as clearly and completely as possible. Before filming their tutorial video each participant was given as much time as they needed to read the task and plan their response. During this time they had access to pencils, pens, colored markers, and additional blank paper, and were free to mark the task worksheet in any way they found helpful. Once they were ready, the participant filmed their tutorial video. Both the planning and the filming stages were entirely self-paced and occurred without the researcher present. The participant's speech, writing, and gestures towards the worksheet were recorded by a camera positioned directly above their workspace for subsequent analysis (Figure 1.3).

[3] There were three slightly different versions of the task: one in which participants were given the full statement, one in which they were given a fill-in-the-blank version ("The sum of the first n odd numbers is equal to _______"), and one in which they were asked to guess the entire statement that they thought the image was intended to represent. However, in this report we consider only those participants who demonstrated understanding of the complete statement "The sum of the first n odd numbers is equal to n^2", either because it was given in the task or because they successfully generated it. For these participants, there were no significant differences between task versions for any of the findings reported in this article.

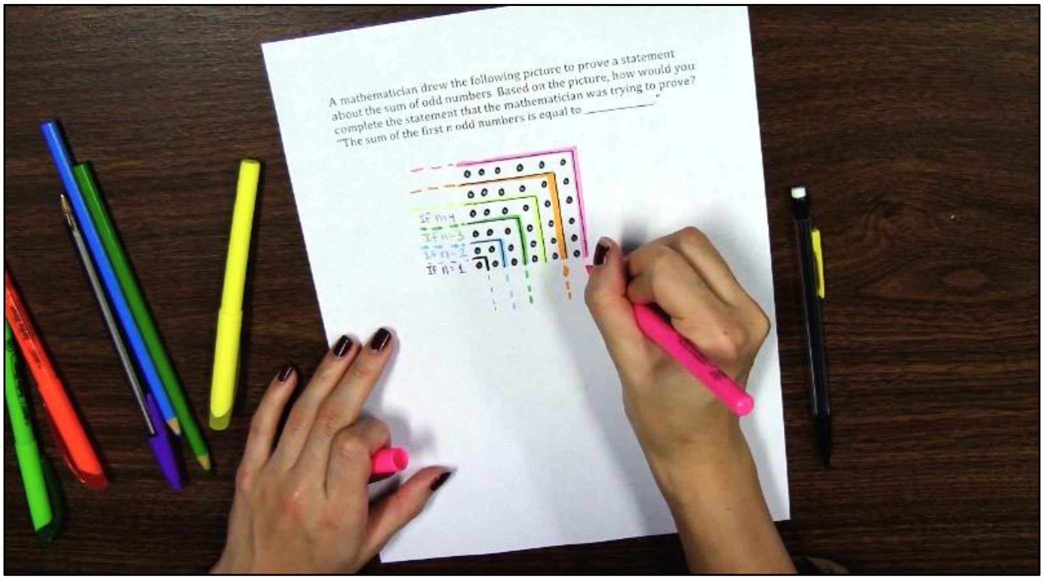


Figure 1.3: Participant workspace.

Analysis. Two raters independently coded the tutorial video footage based on specific criteria. First, the raters distinguished between *case-based* and *pattern-based* explanations. Case-based explanations used the image to describe one or multiple specific cases of the statement (e.g., showing how the first 3 layers of the image depict $1 + 3 + 5 = 3^2$). In contrast, pattern-based explanations offered a general description a pattern in the image (e.g., explaining that the picture shows consecutive odd numbers of dots in each layer, and that at each stage the layers form a square). Coders also noted responses which explicitly mentioned that the pattern could be extended beyond just the first six cases depicted in the image, and any instance of a rigorous justification of pattern extension (comparable, although not necessarily identical, to the argument in Figure 1.2).

Semi-Structured Interview

Procedure. Once the participant finished their tutorial video the researcher returned to the room and conducted a semi-structured interview. The purpose of this second phase of the study was to assess in a standardized way the conclusions that the participants had drawn from the image. Specifically, we were interested in determining whether after working with the visual proof the participant had generalized the statement to cases not depicted in the image (*generalization*), and if so, whether this conclusion was truly extended to *all* natural numbers (*necessity*). To assess generalization, we asked each participant two questions: (1) "Do you think the statement is true in all cases?", and (2) "What would be the sum of the first 8 odd numbers?" Importantly, these questions alone were not enough to determine whether the participant recognized the necessity of the statement. In daily life, the word "all" is used quite loosely, as when we say "All birds can fly" or "All Californians love the beach". In mathematics, however, the universal quantifier "all" is much stronger in that it implies the impossibility of counterexamples. In order to assess whether each participant recognized the necessity of the statement we asked a follow-up question: we suggested the existence of large-magnitude counterexamples ("Very large numbers where the statement actually isn't true") and asked what they thought about this possibility. Any participant who expressed significant doubt at this possibility was asked how they might argue against the existence of large-magnitude counterexamples. The interview was recorded in the same manner as the tutorial video.

Analysis. The interview footage was independently coded by the same two raters. Any participant who answered "Yes" to question (1) and quickly applied the rule to answer "64" to question (2) was considered to have generalized the statement. To assess whether this generalization extended to all natural numbers, two coders rated each participant's expressed doubt or resistance to the possibility of large-magnitude counterexamples on a 0-5 scale. Scoring criteria and corresponding sample responses are given in Table 1.1. For later analysis, scores of 0-3 were associated with low resistance to counterexamples, while scores of 4 and 5 were considered high resistance. For participants who expressed high resistance to counterexamples, the coders also noted how they argued against such a possibility, including whether they produced a rigorous image-based argument and/or mentioned or performed a formal proof by mathematical induction.

Table 1.1: Scoring criteria and sample responses for rating participants' resistance to the possibility of large-magnitude counterexamples

Score	Criteria, *Sample Response*
0	No resistance; clearly accepts the existence of counterexamples *"That makes sense, I believe that."*
1	Very little resistance; it is likely there are counterexamples *"That makes sense. I'd ask to see what the number was."*
2	Unsure or neutral to the existence of counterexamples *"Since I haven't explored the math I couldn't make a statement about it."*
3	Somewhat doubtful of the existence of counterexamples *"It seems like the pattern should hold, but when it gets to high numbers I guess that's possible."*
High Resistance to Counterexamples	
4	Very doubtful of the existence of counterexamples *"I'd be skeptical of that without evidence."*
5	Complete rejection; states counterexamples are impossible *"That seems impossible, that doesn't make sense."*

Questionnaire. Finally, each participant filled out a short questionnaire indicating their age, gender, major, and the names of any university-level mathematics classes that they had completed. Participants also were asked to indicate if they were familiar with the term "mathematical induction", and if so, to describe what they knew about it.

Results

Tutorial Video. There were no significant differences between the duration of the time spent planning for MI-Trained and MI-Untrained participants (Trained M = 7.51 minutes, SD = 5.2 minutes; Untrained M = 9.31 minutes, SD = 5.6 minutes; $t(37) = -1.03$, $p = 0.31$), or the duration of their tutorial videos (Trained M = 6.26 minutes, SD = 5.49; Untrained M = 4.12 minutes, SD = 2.16 minutes; $t(37) = 1.49$, $p = 0.15$). In their tutorial videos, 10/17 (58.8%) MI-Untrained participants relied on case-based strategies (using the image to describe one or multiple specific cases of the theorem), while 7/17 (41.2%) produced pattern-based explanations (describing a general pattern present in the image). MI-Trained participants overwhelmingly preferred pattern-based strategies (20/22, 90.9%), with only 2 MI-Trained participants producing case-based explanations (9.1%). MI-Untrained participants were significantly more likely to produce case-based explanations than were MI-Trained participants (Fisher Exact Test, p = 0.026; Odds Ratio = 6.63; 95% CI 1.02, 76.84; Figure 1.4). Only one of the 17 MI-Untrained participants (5.9%) mentioned that the pattern represented in the image could be extended beyond the first six cases. A greater number (12/22, 54.5%) MI-trained participants mentioned the possibility of pattern extension; five MI-Trained participants used the image to demonstrate that the pattern *necessarily* continues. MI-Trained participants were significantly more likely than MI-Untrained participants to mention the possibility of pattern extension (Fisher Exact Test, $p = 0.0018$; Odds Ratio = 17.82; 95% CI 2.07, 866.16).

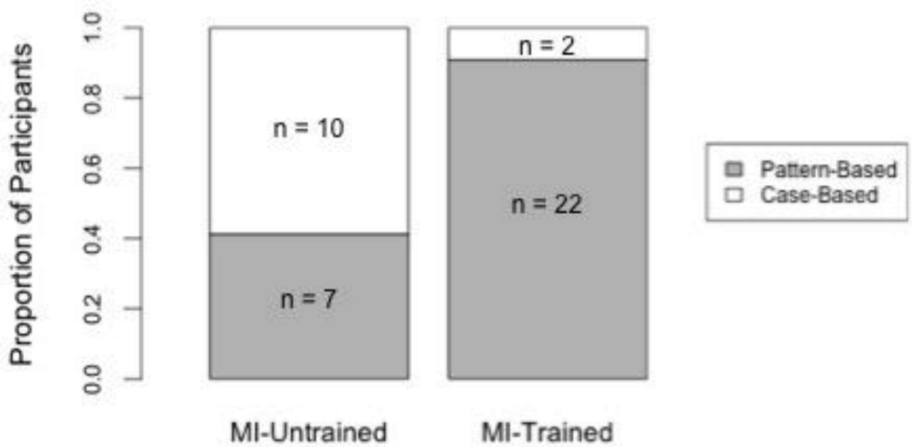
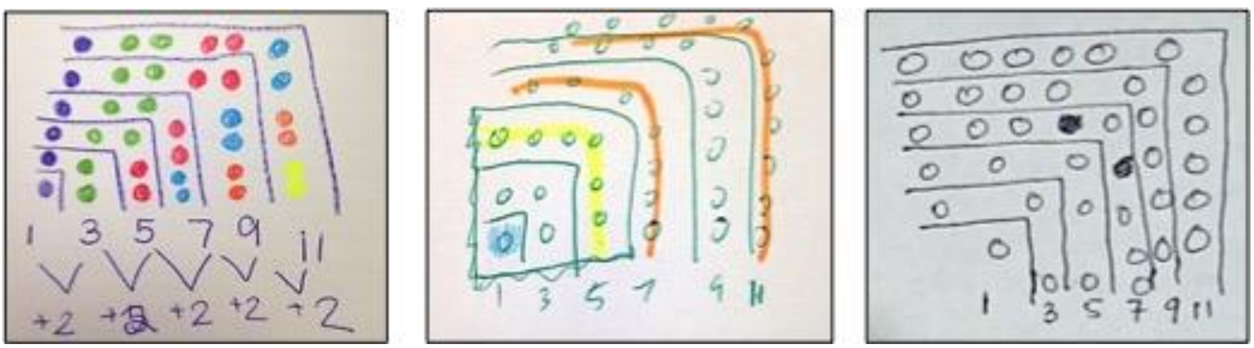


Figure 1.4: MI-Trained participants overwhelmingly preferred pattern-based explanations, while the majority of MI-Untrained participants provided case-based explanations.

Discussion of Tutorial Video. MI-Trained and MI-Untrained participants employed different strategies in their explanations. Untrained participants often used the image to walk the viewer through one or multiple specific cases of the theorem, while trained participants were more likely to describe a general pattern in the image. The untrained participants' reliance on case-based strategies suggests that they may have been viewing the image as a set of examples, which just happened to be presented visually rather than numerically, and not considering the possibility that the pattern necessarily continues. Furthermore, qualitative analysis suggests that many MI-Untrained participants were unaware of the invariance of the pattern represented in the image**.** Untrained participants often chose to re-draw the image as part of their explanation, and in some cases these drawings violated essential features of the original image (Figure 1.5).

Specifically, some untrained participants produced images that did not maintain the regular row-column structure, suggesting that these participants may have been genuinely unaware of the invariance of the pattern represented in the image. MI-Trained participants were more likely to describe the general pattern represented by the image, and to mention that the pattern could be extended indefinitely. However, it was still only a relatively small percentage of MI-Trained participants who mentioned the possibility of pattern extension, and an even smaller portion who used the image to justify why the pattern represented in the image *necessarily* extends to every natural number.



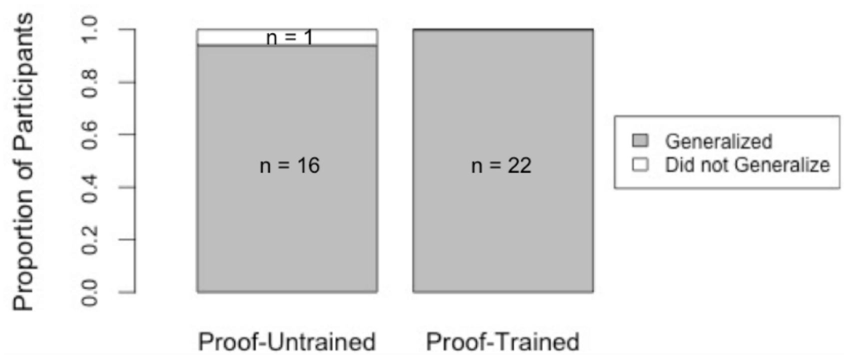
(a) (b) (c)

Figure 1.5: In their tutorial videos some MI-Untrained participants produced images that violated the essential row-column structure of the original figure.

Data from the tutorial video suggests two results. First, MI-Untrained participants do not explicitly describe the image as representing a pattern that could be extended indefinitely, and in some cases may be truly unaware of the necessity of pattern extension. Second, MI-Trained participants, while more likely to mention pattern extension, do not tend to spontaneously use the image to justify the necessity of the statement. However, the fact that many of our participants neglected to *mention* certain aspects of the image during their tutorial videos doesn't necessarily imply that they were unaware of these features. In the next part of the study we used a semi-structured interview to probe specific aspects of the conclusions our participants drew from the visual proof, including their assessment of necessity of the theorem.

Semi-Structured Interview. The vast majority of participants indicated a willingness to generalize the statement to cases not depicted in the image (Figure 1.6a). There was no difference between the MI-Untrained and MI-Trained groups in their willingness to generalize (Untrained 16/17, Trained 22/22, Fisher Exact Test, $p = 0.44$). However, MI-Trained participants were significantly more likely than Untrained participants to show high resistance to large-magnitude counterexamples (Fisher Exact Test, $p = 0.007$, Odds Ratio = 7.02; 95% CI 1.4, 41.2; Figure 1.6b). Seventeen of the 22 MI-Trained participants (77.3%) expressed a high degree of doubt regarding the existence of counterexamples (characterized by a resistance score of 4 or 5). In contrast, only 5 out of the 16 (31.3%) MI-untrained participants who generalized the statement expressed a high degree of doubt towards the possibility of counterexamples.

(a)



(b)

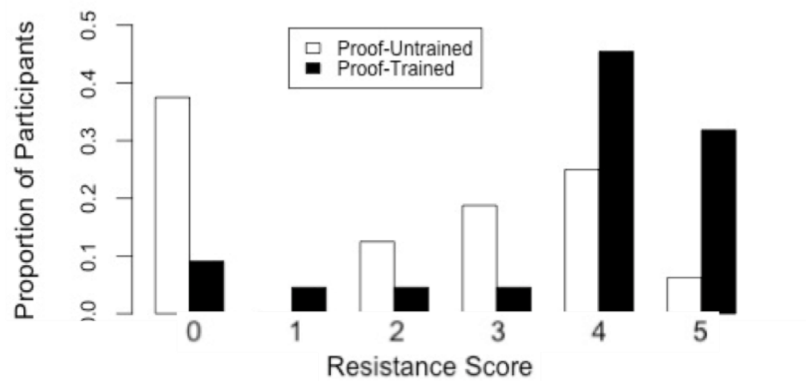


Figure 1.6: All participants were willing to generalize the theorem to nearby cases (a). However, MI-Untrained participants were significantly less likely to show a high degree of doubt that large-magnitude counterexamples to the statement are possible (b).
Figures from Relaford-Doyle & Núñez (2017).

Participants who expressed significant doubt about the existence of large-magnitude counterexamples were asked how they would argue against such a possibility. Two MI-Untrained participants used the image to produce a rigorous image-based justification, while 8 out of the 17 MI-Trained participants who showed high resistance to counterexamples did so. In general, MI-Trained participants showed a preference for the formal proof method; 75% mentioned that they could use formal mathematical induction to argue against counterexamples, and 25% actually completed a formal proof, even though it wasn't a required part of the task.

Discussion of Semi-Structured Interview. We observed that, while participants in both groups were willing to generalize the theorem to nearby cases, only MI-Trained participants subsequently showed a high degree of resistance to the possibility of counterexamples. This suggests that MI-Untrained participants were not engaging in informal mathematical induction (characterized by recognition of necessity), but were instead engaging in everyday inductive reasoning and using the visual evidence as the basis for an inductive generalization. Additionally, MI-Trained participants, while generally recognizing the necessity of the theorem, had a difficult time transferring their knowledge of formal mathematical induction to the new, non-algebraic representational system, which suggests that many trained students' understanding of formal mathematical induction may have been largely procedural in nature.

When responding to the possibility of large-magnitude counterexamples, many of our MI-Untrained participants made statements about natural numbers that were inconsistent with the formal characterization that is required for mathematical induction. Specifically, we observed that MI-Untrained participants frequently expressed a belief that very large-magnitude natural numbers may be unpredictable or follow different rules than smaller numbers. For instance, when asked about the possibility of large-magnitude counterexamples, three MI-Untrained participants responded as follows:

- "I guess that makes sense. Like the larger numbers could be, like, outliers, or something like that."
- "Based on my impression, just based on this observation, I think it would work, but when it gets to really high numbers, um, it's possible that, like *(pauses)*. I can see maybe it gets kind of fuzzy. Because at extremes things tend to not work as they do normally."

- “I guess this model proves to be true for, until, maybe like 99. I know it would be true. I don’t know, I consider 99 a big number…Like maybe the model deconstructs at a thousand or a million, I don’t know, but it’s too hard to draw a million dots.”

These responses all suggest that these participants believe that large numbers may have qualitatively different properties than small numbers, such that rules that apply to small numbers may no longer work at larger magnitudes. This is a reasonable conclusion to draw – in practice there are many differences between small and large numbers: small numbers (like *one* through *nine*) are encountered more frequently, have simple numerical notation and lexical structure, and are easier to use in computations. However, this finding is surprising in that it is in opposition to the widely-held assumption in developmental psychology that “mature” conceptualizations of natural number are consistent with the Dedekind-Peano axioms, in which the entire set of natural number is governed by the same logic (Cheung et al. 2017; Rips et al. 2008; Sarnecka & Carey 2008). In contrast, MI-Trained participants expressed formally-consistent conceptualizations of natural number, either by invoking technical notions like the inductive step or by referring to the regularity of counting (for a full qualitative analysis of participants’ comments regarding natural number, see Relaford-Doyle & Núñez, 2018).

Questionnaire and Individual Differences

Unsurprisingly, MI-Trained participants had taken significantly more university-level math classes than the MI-Untrained participants (Trained M = 6.6, SD = 2.20; Untrained M = 2.94, SD = 2.02; one-tailed $t(37) = 5.33$, $p < 0.01$). However, in neither group was number of math courses taken significantly related to any outcomes during either the tutorial video or interview phases of the study (median split in both groups, Fisher Tests all non-significant). The fact that no outcomes were related to either pattern completion ability or amount of exposure to general mathematics indicates that the difference in our two groups' performance is related specifically to differing levels of exposure to mathematical proof-writing in general, and perhaps to formal mathematical induction in particular.

Various studies have shown that female undergraduates tend to express lower confidence in their mathematical abilities than their male peers (Felder et al., 1995; Peters, 2013; Tariq et al., 2013); thus, we may expect female participants in our study to show relatively lower resistance to the suggestion of large-magnitude counterexamples than their male peers. To explore a possible effect of sex, participants' resistance to counterexamples was analyzed with a 2 (Sex: Male versus Female) x 2 (Training: Untrained versus Trained) between-subjects ANOVA. The main effect of training on resistance to counterexamples was significant ($F(1, 34) = 8.75$, $p < 0.01$). There was a marginal main effect of sex ($F(1, 34) = 3.26$, $p = 0.08$), and no interaction between sex and training ($F(1, 34) = 1.27$, $p = 0.27$). However, for our analysis we were not concerned with mean resistance scores, but rather whether the participant expressed high resistance to counterexamples (categorized by a resistance score of 4 or 5). We observed no significant differences in the likelihood of expressing high resistance to counterexamples between males and females in either group (Fisher Exact Test, Trained $p = 0.31$, Untrained $p = 1$). Furthermore, non-parametric testing revealed a significant effect of training on resistance to counterexamples (Mann-Whitney test, $U = 88.5$, $p < 0.01$), but no significant effect of gender ($U = 138.5$, $p = 0.2$).

General Discussion and Implications for Education

RQ1. How do participants who are unfamiliar with formal mathematical induction interpret the image? Specifically, do they recognize the necessity of the theorem it represents? If not, what conceptual obstacles keep them from doing so?

We observed that, while the majority of MI-Untrained participants were willing to generalize the target theorem to nearby cases, the majority showed relatively little resistance to the possibility that large-magnitude counterexamples may exist. This result – generalization without the recognition of necessity – suggests that these participants used the visual proof as the basis for an inductive generalization and did not engage in informal mathematical induction. This interpretation is further supported by the observation that the majority of MI-Untrained participants used case-based strategies in their explanations, thus treating the image as a set of discrete examples, rather than as the first cases in a pattern that could be extended indefinitely.

This result is inconsistent with claims in developmental psychology that children as young as five years old can spontaneously reason by an informal version of mathematical induction (Baroody et al., 2013; Smith, 2003). There are at least two possible explanations for this inconsistency. First, it is possible that people with no training in formal mathematical induction, including young children, *can* reason by informal mathematical induction in simple contexts (like recognizing that the natural numbers are infinite), but that this reasoning breaks down when the mathematical content is more sophisticated. In other words, there may be some limited capacity for genuine informal mathematical induction that pre-exists formal training. A second possibility is that, like our adult participants, the children in Baroody's and Smith's reports were simply engaging in standard inductive reasoning, and that it has been mischaracterized as "informal mathematical induction". As described earlier, the existing empirical work has failed to convincingly demonstrate that children can recognize necessity of the theorem, which is a critical component of informal mathematical induction. Further work is required to characterize the nature of generalizations that untrained people, both children and adults, make in different mathematical contexts. Specifically, future studies must carefully assess untrained people's recognition of the necessity of mathematical generalizations, thereby disentangling informal mathematical induction and everyday inductive reasoning.

Of central importance is the question, what obstacles kept our MI-Untrained participants – all highly educated adults who are presumably comfortable with addition, odd numbers, and squaring – from using the visual proof to recognize the necessity of the theorem? We were surprised by the number of MI-Untrained participants who made statements about the natural number system that were inconsistent with the formal characterization required for mathematical induction. Many participants expressed a particular misconception – that very large numbers behave differently or follow different rules than smaller, more familiar ones – and this may have impeded their ability to generalize the theorem to *all* natural numbers. This suggests that one conceptual roadblock that students may face when first encountering formal mathematical induction is a lack of understanding of the natural number system as it is formally characterized in the Dedekind-Peano axioms. Even at the college level, instructors should not assume that their students already possess formally-appropriate conceptualizations of the natural number system. Students may benefit explicit instruction in the Dedekind-Peano axioms and their implications in the natural number system, which are sometimes left out of formal instruction (Zazkis & Leikin, 2010).

RQ2. How do participants who *are* familiar with formal mathematical induction use the image to justify the theorem? Can they transfer their knowledge of mathematical induction to this new representational format, or is their knowledge of mathematical induction intimately linked to the algebraic procedure required by the formal method?

While our MI-Trained participants were significantly more likely than our Untrained participants to express a high degree of doubt about the existence of counterexamples, they frequently referred to the formal mathematical induction to justify this claim. While a few MI-Trained participants were able to use the image to demonstrate the necessity of the theorem, most did not provide any image-based argument for the general theorem. In other words, the majority of MI-Trained participants didn't transfer their knowledge of formal mathematical induction to the novel representational system. This is consistent with previous work (Baker, 1996; Harel, 2001; Lowenthal & Eisenberg, 1992; Woodall, 1981), which has shown that students often have only *procedural* knowledge of mathematical induction; they know how to perform the formal algebraic proof, but lack the general conceptual understanding that they would need to construct an argument in a different representational system.

However, it is also possible that our MI-Trained participants who did not provide an image-based justification *did* have conceptual understanding of mathematical induction, but simply rejected the visual representation as a valid means of justification. In the case of formal mathematical induction, it is the Axiom of Induction which allows for the conclusion that the theorem will hold for all natural numbers; mathematics students may have been wary of using a non-axiomatic representational system to justify such a conclusion. More generally, it could also be the case that this pattern of results is reflective of students' awareness of the general norms of mathematical justification. In most modern mathematics, pictures and other visual representations are considered useful psychological aids but are explicitly disallowed in formal proofs. Students are often suspicious of or reluctant to use purely visual representations in mathematics, particularly in contexts of justification (Eisenberg & Dreyfus, 1991; Inglis & Mejía-Ramos, 2009). The fact that so few of our MI-Trained participants produced rigorous image-based arguments may not indicate a lack of conceptual understanding, but might instead reflect negative attitudes towards the use of visual representations in mathematics. Therefore, we cannot conclude from our evidence alone that MI-Trained participants are genuinely *unable* to produce image-based arguments; they may simply be unwilling to do so.

This raises an important point about how mathematical induction, and mathematical proof in general, is taught. In virtually all other areas in mathematics, it is widely acknowledged that transferring between multiple representations helps students to develop conceptual understanding of mathematical content (Lesh et al., 1987; Pape & Tchoshanov, 2001; Schoenfeld, 1985). For instance, algebra teachers want students to learn that a function can be represented as an algebraic equation, as an input-output table, or as a graph (e.g. Brenner et al., 1997); indeed, possessing truly conceptual knowledge of functions implies that a student can transfer flexibly between these different representational systems. The same is not always the case for mathematical justification, where students often learn that a proof must be written as a symbolic, propositional argument. If students are exposed to only one means of representing a mathematical proof, they may struggle to develop deep conceptual understanding of the proof techniques they learn. Why should we expect students to know there's more to formal mathematical induction than an algebraic procedure, when every example they see consists of that procedure?

In order to develop genuinely conceptual, representation-independent knowledge about formal mathematical induction, students may benefit from being exposed to a variety of "proofs" in which the same concepts are applied in different representational systems. For instance, both formal mathematical induction and images such as the one used in this study rely on establishing the truth of the theorem for some starting value, and then demonstrating the invariance of the relationship between successive instances. A student who has seen only the formal proof may easily come to believe that the inductive step is simply an algebraic procedure involving k and $k + 1$. A student who has also seen and understood a visual proof by induction may be in a better position to understand that the inductive step demonstrates the constancy of the relationship between successors, and why this allows for the conclusion that the theorem will be true for all natural numbers. While our study doesn't demonstrate the pedagogical value of visual proofs by induction, future work could explore the potential educational benefits of supplementing instruction in formal mathematical induction with visual representations in order to foster deep conceptual understanding of the proof method.

Conclusion

This study used a novel method to investigate undergraduate students' conceptualizations of mathematical induction and explore the conceptual difficulties that students may face when learning this proof method. Using a 'visual proof by induction' – a simple image that represents a proof by formal mathematical induction in an accessible, non-symbolic representational system – we were able to explore both trained and untrained students' conceptualizations around the proof method. Our results suggest that MI-Untrained students used the image as the basis for an inductive generalization, but *not* informal mathematical induction; while they initially stated that the theorem was true for “all” numbers, most untrained students were willing to accept the possible existence of counterexamples, and thus did not recognize the necessity of the theorem. In some cases this may have been due to a lack of understanding of the key features of the image (as in participants who violated these features when they redrew the image; Figure 1.5). However, qualitative analysis of MI-Untrained students’ responses indicate another possible source of difficulty: many undergraduates may possess non-normative conceptualizations of natural number, thus making them more likely to believe that the theorem could break down for large-magnitude numbers. In contrast, MI-Trained students were highly resistant to counterexamples; however, most indicated a preference for formal mathematical induction and had difficulty using the image to provide a rigorous justification for the theorem. Consistent with previous findings, this suggests that these students’ knowledge of mathematical induction is largely procedural in nature, and reliant on applying a specific algebraic procedure. Based on these results, we make two recommendations for educators and researchers interested in fostering conceptual understanding of mathematical induction in students. First, instructors should provide novice students with explicit instruction in the Dedekind-Peano Axioms in order to ensure that all students possess the understanding of the natural number system that is required for recognizing *why* mathematical induction is a valid proof method. Second, instructors and researchers could explore the potential pedagogical benefits of supplementing instruction in formal mathematical induction with rigorous “proofs” in different representational formats, including visual proofs by induction. Future work should explore whether encouraging students to “translate” their knowledge between different representational systems – one algebraic, and the other visual – may help students develop deeper conceptual knowledge of formal mathematical induction.

Chapter 1, in full, is a reprint of the material as it appears in *The International Journal of Research in Undergraduate Mathematics Education*, 2021, Relaford-Doyle, Josephine; Núñez, Rafael, 2021. The dissertation author was the primary investigator and author of this paper.

References

Avital, S. & Libeskind, S. (1978): Mathematical induction in the classroom: Didactical and mathematical issues. *Educational Studies in Mathematics 9*, 429-438.

Baker, J.D. (1996). Students' difficulties with proof by mathematical induction. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Baroody, A. (2005). Discourse and Research on an Overlooked Aspect of Mathematical Reasoning. The American Journal of Psychology, 118(3), 484-489.

Baroody, A. J., Reid, E. E., & Purpura, D. J. (2013). An example of a hypothetical learning progression: The successor principle and emergence of informal mathematical induction. *WISDOMe Monograph*, *3*.

Borwein, P., & Jörgenson, L. (2001). Visible structures in number theory. *The American Mathematical Monthly*, *108*(10), 897-910.

Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Durán, R., Reed, B. S., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal*, *34*(4), 663-689.

Brown, J. R. (1997). Proofs and pictures. *The British Journal for the Philosophy of Science*, *48*(2), 161-180.

Brown, J. R. (2008). Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures. Second Edition. New York: Routledge.

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive psychology*, *92*, 22-36.

Eisenberg, T., & Dreyfus, T. (1991, February). On the reluctance to visualize in mathematics. In *Visualization in teaching and learning mathematics* (pp. 25-37). Mathematical Association of America.

Ernest, P. (1984). Mathematical induction: A pedagogical discussion. Educational Studies in Mathematics, 15(2), 173-189.

Felder, R. M., Felder, G. N., Mauney, M., Hamrin Jr, C. E., & Dietz, E. J. (1995). A longitudinal study of engineering student performance and retention. III. Gender differences in student performance and attitudes. *Journal of Engineering Education*, *84*(2), 151-163.

Fischbein E., & Engel, I. (1989): Psychological difficulties in understanding the principle of mathematical induction", in G Vergnaud, J. Rogalski & M. Artigue (Eds.): *Proceedings of the 13th international conference for the Psychology of Mathematics Education,* Vol I (pp. 276-282). Paris, France: CNRS.

Giardino, V. (2010). Intuition and visualization in mathematical problem solving. *Topoi*, 29(1), 29-39.

Hanna, G. (2000). Proof, explanation and exploration: An overview. Educational Studies in Mathematics, 44(1-2), 5-23.

Harel, G. (2001). The development of mathematical induction as a proof scheme: A model for DNR-based instruction. In S. Campbell & R. Zazkis (Eds.), *Learning and teaching number theory*. (pp. 185–212). Norwood, New Jersey: Ablex.

Inglis, M., & Mejía-Ramos, J. P. (2009). On the persuasiveness of visual arguments in mathematics. *Foundations of Science*, *14*(1-2), 97-110.

Kempen, L., & Biehler, R. (2019). Pre-service teachers' benefits from an inquiry-based transition-to-proof course with a focus on generic proofs. *International Journal of Research in Undergraduate Mathematics Education*, 5(1), 27-55.

Lesh, R., Post, T. R., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In *Problems of representations in the teaching and learning of mathematics*. Lawrence Erlbaum.

Lowenthal, F., & Eisenberg, T. (1992). Mathematical Induction in School: An illusion of Rigor? *School Science and Mathematics*, 92(5), 233-238.

Miller, S. A., Custer, W. L., & Nassau, G. (2000). Children's understanding of the necessity of logically necessary truths. *Cognitive Development*, 15(3), 383-403.

Morris, B. J., & Sloutsky, V. (2001). Children's solutions of logical versus empirical problems: What's missing and what develops?. *Cognitive Development*, 16(4), 907-928.

Movshovitz-Hadar, N. (1993): The false coin problem, mathematical induction and knowledge fragility. *Journal of Mathematical Behavior 12*, 253-268.

Pape, S. J., & Tchoshanov, M. A. (2001). The role of representation (s) in developing mathematical understanding. *Theory into practice*, *40*(2), 118-127.

Peters, M. L. (2013). Examining the relationships among classroom climate, self-efficacy, and achievement in undergraduate mathematics: A multi-level analysis. *International Journal of Science and Mathematics Education*, *11*(2), 459-480.

Relaford-Doyle, J. & Núñez, R. (2017). When does a 'visual proof by induction' serve a proof-like function in mathematics? In G. Gunzelmann, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), Proceedings of the 39th Annual Conference of the Cognitive Science Society (pp. 1004-1009), London, UK: Cognitive Science Society.

Relaford-Doyle, J. & Núñez, R. (2018) Beyond Peano: Looking into the unnaturalness of natural numbers. In S. Bangu (Ed.), *Naturalizing Logico-Mathematical Knowledge: Approaches from Psychology and Cognitive Science*. New York: Routledge.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences*, *31*(6), 623-642.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662-674.

Schoenfeld, A. H. (1985). *Mathematical Problem Solving*. Orlando, FL: Academic Press.

Smith, L. (2003). Children's reasoning by mathematical induction: normative facts, not just causal facts. *International Journal of Educational Research*, *39*(7), 719-742.

Stylianides, G. J., Stylianides, A. J., & Philippou, G. N. (2007). Preservice teachers' knowledge of proof by mathematical induction. *Journal of Mathematics Teacher Education*, *10*(3), 145-166.

Tariq, V. N., Qualter, P., Roberts, S., Appleby, Y., & Barnes, L. (2013). Mathematical literacy in undergraduates: role of gender, emotional intelligence and emotional self-efficacy. *International Journal of Mathematical Education in Science and Technology*, *44*(8), 1143-1159.

Woodall, D. R. (1981). Finite sums, matrices and induction. The Mathematical Gazette, 65(432), 92-103.

Zazkis, R., & Leikin, R. (2010). Advanced mathematical knowledge in teaching practice: Perceptions of secondary mathematics teachers. *Mathematical Thinking and Learning*, *12*(4), 263-281.

CHAPTER 2

Investigating Contexts in Which Undergraduates Demonstrate Reasoning by Informal Mathematical Induction

Introduction

In the previous chapter, we showed that undergraduates did not recognize the mathematical necessity of a statement even when provided a visual proof by induction. Specifically, undergraduates predicted that the pattern represented in the diagram would continue, but remained open to the possibility that there could be large-magnitude counterexamples to the statement. We took this as evidence that our participants were engaging in standard inductive reasoning, rather than "informal mathematical induction", which is characterized by generalization to all numbers *and* recognition of mathematical necessity. However, the previous study had some limitations which may have impacted our results. First, we presented participants with only one relatively complex mathematical statement; thus, it is possible that our untrained participants' performance reflected a general lack of comprehension, and that reasoning by informal mathematical induction may still occur in simpler contexts. Second, the previous study relied on a face-to-face interview-based protocol, leaving the door open for social dynamics to have influenced responses (for instance, participants may have felt uncomfortable contradicting the interviewer, or withheld statements out of politeness). In this follow-up study, we use a computer-based task and a wider variety of mathematical statements ranging in complexity to address these limitations and explore when, if ever, undergraduates without relevant formal mathematical training engage in informal mathematical induction.

Background

As discussed in the previous chapter, informal mathematical induction is distinct from both the formal proof method of mathematical induction - as it does not require that a formal proof be written - and standard inductive reasoning, which involves generalization but does not imply recognition of mathematical necessity. If counting adults possess a formally-consistent conceptualization of natural number, as characterized by a successor principle consistent with the Peano axioms, then we may expect to see evidence of this in their reasoning within the domain of natural numbers. Specifically, we might expect that adults, even without high-level mathematical training, could still engage in reasoning by informal mathematical induction. Previous research has attempted to provide evidence for such reasoning in children (Baroody, 2005; Baroody et al., 2013; Smith, 2003), but as described in the previous chapter, this research does not provide compelling evidence for children's recognition of mathematical necessity, but rather based conclusions on children's "yes" or "no" responses to questions about statements regarding numbers. This leaves open the possibility that the children weren't in fact reasoning by mathematical induction, but rather engaging in standard inductive reasoning within the domain of natural numbers (Rips et al., 2008). However, informal mathematical induction implies much more than inductive reasoning about mathematics; thus, any study which claims to shed light on reasoning by informal mathematical induction must assess participants' awareness of mathematical necessity. This is a central goal of the present study.

Additionally, informal mathematical induction may be distinguished from inductive reasoning by making use of known effects in inductive reasoning, which we would not expect to see if participants are engaging in informal mathematical induction. For instance, *similarity effects* are a well-documented phenomenon in inductive reasoning (Osherson et al., 1990;

Osherson et al., 1991; Rips, 1975). This extensive body of research has shown that people are more likely to generalize a hypothesis to a new case when it is more similar to the observed cases, and less likely to generalize to less similar targets. For instance, Rips (1975) asked participants to make inductive judgments about pairs of animal categories (e.g., if all the robins on a small island have a communicable disease, what proportion of mice on the island did participants think had the disease). Another group provided pairwise similarity ratings between the animal categories (e.g., robins and mice). Rips found that similarity judgments strongly predicted inductive generalization judgments, suggesting that participants were more likely to generalize when the categories in question were deemed more similar to one another. Importantly, we would *not* expect to see such effects in informal mathematical induction; awareness of mathematical necessity would imply that, once the generalization occurs, it is fully and equally applied to all natural numbers, no matter their distance or "similarity" to the observed cases. Thus, by examining when and where we observe similarity effects, we can begin to disentangle informal mathematical induction and inductive generalization within the domain of natural numbers.

In this study we assess whether there is evidence of similarity effects when participants make generalizations within the natural numbers. We equate similarity with numerical distance, and ask whether, when given evidence about the numbers 1, 2, and 3 participants are more likely to generalize to closer (more similar in magnitude) numbers than to farther (less similar) numbers. To assess participants' recognition of mathematical necessity, we use a wagering paradigm in which we ask participants how much they would bet that a generalization is true, based on the provided evidence. Bet amount was chosen as an appropriate index of participants' awareness of necessity for a variety of reasons. Betting is a commonly used measure in behavioral economics (Sauer, 1998; Suhonen & Saastamoinen, 2018; Wohl et al., 2014) where it is viewed as "more objective" than confidence ratings, which are known to vary quite a bit based on task instructions (Koch & Preuschoff, 2007). In psychology, researchers have used post-decision wagering as a measure of people's awareness, with some preferring it to "subjective" measures like confidence ratings because it "measures awareness directly", rather than indexing "awareness of awareness" (Persuad et al., 2007). Drawbacks with using bet amount as a measure of awareness have been identified but were deemed irrelevant for this study. For instance, some researchers have argued that wagering behavior is not a straightforward proxy for confidence since it is subject to risk-aversion and economic context (Studer et al., 2015). Indeed, studies that have included both confidence ratings and wagering have shown participants are more likely to give high confidence ratings than they are to wager high amounts (Persuad & Macleod, 2008). This suggests that bet amount may be a more conservative measure of participants' certainty and is thus more appropriate for the present study. Similarly, we are not concerned if confidence ratings and bet amount actually index different phenomena; this study assesses perceived degree of mathematical necessity, which we would expect to be reflected – but not directly measured – in subjective confidence ratings. We are operating with the assumption that if participants recognize the mathematical necessity of a statement, they should be willing to "put their money where their mouth is" and bet the maximal amount (Koch & Preuschoff, 2007). For these reasons, bet amount is an appropriate indicator of our participants' degree of awareness of mathematical necessity and thus the scope and robustness of informal mathematical induction. Further, this method allows us to make clear predictions: If participants engage in standard inductive reasoning, we would expect to see bets below the maximal amount, and which decrease for father/less similar targets. However, if participants engage in informal mathematical induction, we would predict that they bet the maximal amount, and, importantly, that they do so regardless of similarity/numerical distance between observed and target cases.

Method

Participants: All participants were undergraduate students at a major research university, and were tested individually. Participants (n = 80; 40 female, 40 male) were recruited through a general subject pool that consisted of students with a range of majors including psychology, cognitive science, and linguistics. Participants received course credit for their participation.

Materials: A web-based study was designed and deployed using Qualtrics. Participants accessed and completed the study on their home or personal devices.

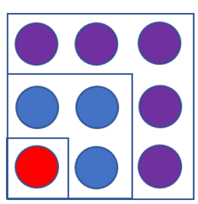
Stimuli: The study consisted of 9 separate evidence sets, each presented twice (once with a close/more similar prediction target, and once with a far/less similar prediction target), for a total of 18 trials. Six of the evidence sets were mathematical in nature and ranged in complexity; for these sets, the evidence always presented information about the numbers 1, 2, and 3, and the close prediction target was always a prototypical counting number between 4 and 9 (see below). Three evidence sets were non-mathematical in nature, and included statements about animals and plants. These non-mathematical trials served both as distractors (to guard against possible demand effects) as well as a methods check (as we can assess whether our procedure replicates known similarity effects in general inductive reasoning tasks).

Table 2.1 presents evidence sets and predictions (both close and far) for all six mathematical items, as well as one example non-mathematical item for reference.

Table 2.1: Evidence sets and predictions for each of the six mathematical items, and one example non-mathematical item.

Item Label	**Evidence Set**	**Close Prediction**	**Far Prediction**
Successor	The number after 1 is 1+1. The number after 2 is 2+1. The number after 3 is 3+1.	The number after 8 is 8+1.	The number after 896 is 896+1.
Addition	$a = a$ $a + 1 = a + 1$ $a + 2 = a + 2$ $a + 3 = a + 3$	$a + 9 = a + 9$	$a + 827 = a + 827$
Powers of Two	2^1 is greater than 1. 2^2 is greater than 2. 2^3 is greater than 3.	2^5 is greater than 5.	2^{750} is greater than 750.
Powers of Five	5^1 (= 5) ends in a 5 5^2 (= 25) ends in a 5 5^3 (= 125) ends in a 5	5^6 ends in a 5	5^{632} ends in a 5
Prime	$1^2 – 1 + 5$ is a prime number. $2^2 – 2 + 5$ is a prime number. $3^2 – 3 + 5$ is a prime number.	$7^2 – 7 + 5$ is a prime number.	$335^2 – 335 + 5$ is a prime number.

Table 2.1: Evidence sets and predictions for each of the six mathematical items, and one example non-mathematical item. (Cont.)

Item Label	**Evidence Set**	**Close Prediction**	**Far Prediction**
Sum of Odds	**No Visual Proof:** Half of the participants received purely numerical evidence: The first 1 odd number is equal to 1^2. ($1 = 1^2$) The sum of the first 2 odd numbers is equal to 2^2. ($1 + 3 = 4 = 2^2$) The sum of the first 3 odd numbers is equal to 3^2. ($1 + 3 + 5 = 9 = 3^2$)	The sum of the first 6 odd numbers is equal to 6^2.	The sum of the first 439 odd numbers is equal to 439^2.
	W/ Visual Proof: For the other half of the participants, the information above was accompanied by this visual proof: 		
Non- Mathematical (Example)	Robins are susceptible to a new disease. Sparrows are susceptible to a new disease. Pigeons are susceptible to a new disease.	Chickens are susceptible to the new disease.	Whales are susceptible to the new disease.

The “Successor” and “Addition” items represent basic facts regarding number and arithmetic, and it is in the context of these two facts - the successor principle and addition of equals - that Smith (2003) and Baroody and co-authors (2013) have argued that young children can reason by informal mathematical induction. By including these items in the present study we can directly assess these claims and determine whether there is evidence that adults reason by informal mathematical induction in the simplest mathematical contexts. The “Powers of Two”, “Powers of Five”, and “Sum of Odds” provide increasing mathematical complexity, and all correspond to mathematical statements which could be formally proven using mathematical induction (though time constraints, detailed in the next section, prohibited participants from actually completing a proof). Based on this, these statements were identified as good candidates for a study of informal mathematical induction. Finally, the “Prime” item represents a mathematical statement which is true in most cases, including those presented in the evidence set, but which is actually *false* for particular values (specifically multiples of 5, including the target value in the Far prediction). As this item could *not* be formally proven using mathematical induction, it is not a good candidate for reasoning by informal mathematical induction; including this item thus allows us to assess whether “informal mathematical induction” may in fact be over-applied to contexts in which it is not actually appropriate.

Finally, as elaborated in the previous chapter, the “Sum of Odds” item has a corresponding visual proof which provides diagrammatic evidence that the pattern will necessarily continue to all natural numbers. To determine whether the format of evidence impacts the likelihood of informal mathematical induction, for this item half of our participants saw only propositional/numerical evidence, and the other were additionally provided a reduced version of the visual proof showing the first 3 cases of the theorem (aligning with the numerical evidence of cases n = 1, 2, and 3 that was provided to the participant).

Procedure: Participants first read the study instructions and indicated when they were ready to begin. In order to make sure participants understood instructions and timing requirements, they completed two practice trials before moving on to the study. In each study trial, participants first saw a screen which presented an evidence set for them to read (labeled "Information"; see Table 2.1, Column 2 for details). This portion of the trial was untimed. The participant indicated via mouse click when they were ready to move on. On the next screen, a prediction appeared below the evidence set, which remained visible on screen. In one block of trials, the prediction asked participants about a target that was close in numerical value or relatively similar to the evidence they had read (Table 2.1, Column 3), while in the other block of trials the prediction asked about a target that was farther from or relatively less similar to the evidence (Table 2.1, Column 4). The order of specific evidence sets was randomized within each block, and the block order was counterbalanced. Upon seeing the prediction, participants were asked "From $0 to $100, how much would you bet that the prediction is true?", and input their answer into a field on the screen. In order to discourage participants from performing calculations in order to directly verify the truth or falseness of the prediction, they were given a 15-second time limit in which to enter their bets. On each trial we recorded the time spent reading the evidence set (unlimited), the time spent inputting the bet (0 seconds to 15 seconds), and the amount of the bet (which ranged between $0 and $100).

Upon completion of the trials, the participant was asked a comprehension question which checked their understanding of one of the more complex mathematical statements used in the study (the "Sum of Odds" item, which was explored in Chapter 1). They then completed a brief survey which asked them to report their major, gender, familiarity with formal mathematical proofs, and whether they had taken Mathematical Reasoning, a university-level course covering formal mathematical induction. When the survey was complete, a screen appeared thanking the participant for their time.

Analysis: First, as we were interested in the generalization behaviors of undergraduates *without* formal training in mathematical induction, we excluded participants who indicated on the exit survey that they had taken the Mathematical Reasoning course (n = 3). To help ensure that we were analyzing meaningful data, we also excluded participants (n = 3) who spent on average less than 3 seconds reading each mathematical evidence set upon their first presentation, as this indicated a possible "rush" through the study and lack of comprehension. This left us with a total of n = 74 participants whose responses were included for analysis. We then dropped trials in which the participant did not enter a trial within the 15-second time window (10/1332 trials). Finally, for the Odds item only, we excluded the responses of participants who provided an incorrect answer to the comprehension check on the post-study survey (13/74 participants (17.6%); thus, for the Odds question we analyzed responses from 61 participants, 29 of whom saw the visual proof).

Results

Methods Checks: We first ran a few analyses to check the efficacy of our methods. To begin, we checked whether betting patterns for our three non-mathematical distractor items did indeed replicate known similarity effects in inductive reasoning. A linear mixed-effects model which included participant, item, and order condition as random effects revealed a significant effect of target similarity (Close/Far) on bet amounts in non-mathematical trials ($\chi^2(1) = 118.9$, p < 0.001): on average, participants bet $76 when the prediction target was more similar to the evidence set (standard deviation = 25.5), and $50 when the prediction target was less similar (standard deviation = 35.2). This replicates similarity effects in inductive reasoning (Osherson et al., 1990; Osherson et al., 1991; Rips, 1975) and suggests that our methods are an appropriate means of studying inductive reasoning and generalization patterns.

Next, to determine whether bet amount might be able to detect understanding of necessity in generalization, we examined participants' maximal betting behaviors: were all participants willing to bet the full amount of 100 on at least one trial? We found a significant difference in participants' betting patterns based on evidence type: 100% of participants bet the maximal amount at least one time on mathematical trials, while 39/74 participants (52%) bet 100 on at least one trial presenting non-mathematical information (Fisher's Exact test, p<0.001). The difference in participants' willingness to bet the maximal amount in mathematical versus non-mathematical contexts suggests that our methods may be detecting recognition of mathematical necessity. That approximately half of our participants were willing to bet the maximal amount even in non-mathematical contexts indicates that for some this betting behavior may simply reflect a very high degree of confidence, and not necessarily mathematical certainty. If this is the case, then our results would overestimate the number of participants recognizing mathematical necessity, and thus should be seen as giving an upper limit (rather than an exact measurement) of proportion of undergraduate participants reasoning by informal mathematical induction in each condition.

Finally, to assess whether we were successful in including a range of complexity in our mathematical evidence sets, we examined the amount of time that participants spent reading each mathematical evidence set upon first exposure (taking reading time as an indicator of relative cognitive complexity of the information; Just & Carpenter, 1980). Mean reading times ranged from 5-6 seconds on first exposure to the Addition, Powers of 2, and Powers of 5 evidence sets; 12-15 seconds for the Successor and Prime items; and 22-23 seconds for the Odds question (both with and without the visual evidence). A linear mixed effects model revealed a significant impact of item on reading time when including participant as a random effect ($\chi^2(6) = 56.1$, $p < 0.001$), suggesting that we were successful in identifying items which represented a range of mathematical complexity. We were somewhat surprised to see that the Successor evidence set had a mid-range mean read time, and hypothesize that perhaps this statement was so elementary that seeing it written out in detail caused confusion or surprise and therefore took a longer time to process. The fact that participants spent the longest time reading the Odds evidence set is consistent with our hypothesis that this represents a mathematical statement with a relatively high degree of complexity.

Mean Bet Amounts

To assess the extent to which generalization in mathematical contexts resembled standard inductive reasoning, we analyzed the impact of prediction target magnitude (close/far) on bet amounts. If generalization on a mathematical item was consistent with informal mathematical induction, we would expect participants to bet the maximal amount of 100, regardless of target magnitude. However, if the generalization reflects standard inductive reasoning, we would expect bets below the maximal amount, and which decrease for larger-magnitude (i.e., less similar) prediction targets.

Figure 2.1 shows descriptive statistics of betting behaviors on each mathematical item, for both close and far prediction targets. To investigate the impact of prediction target magnitude, we created a linear mixed effects model which included participant and order condition as random effects. Subsequent model comparisons revealed a significant effect of target magnitude on bet amounts ($\chi^2(1) = 6.59$, $p = 0.01$), as well as an interaction between target magnitude and item ($\chi^2(12) = 285.3$, $p < 0.001$). Post-hoc pairwise tests were conducted using the emmeans package (1.7.0, Lenth, 2021) and showed a significant difference between mean bet amounts for close and far targets for the Prime item only ($t(781) = 3.42$, $p = 0.04$). Pairwise tests by items revealed that there were no significant differences overall between mean bets for the Successor, Addition, and Powers of 2 items, and that participants bet significantly higher amounts on these items than on the Powers of 5 item, which itself saw significantly higher bet amounts than the Prime item and both Odds questions.

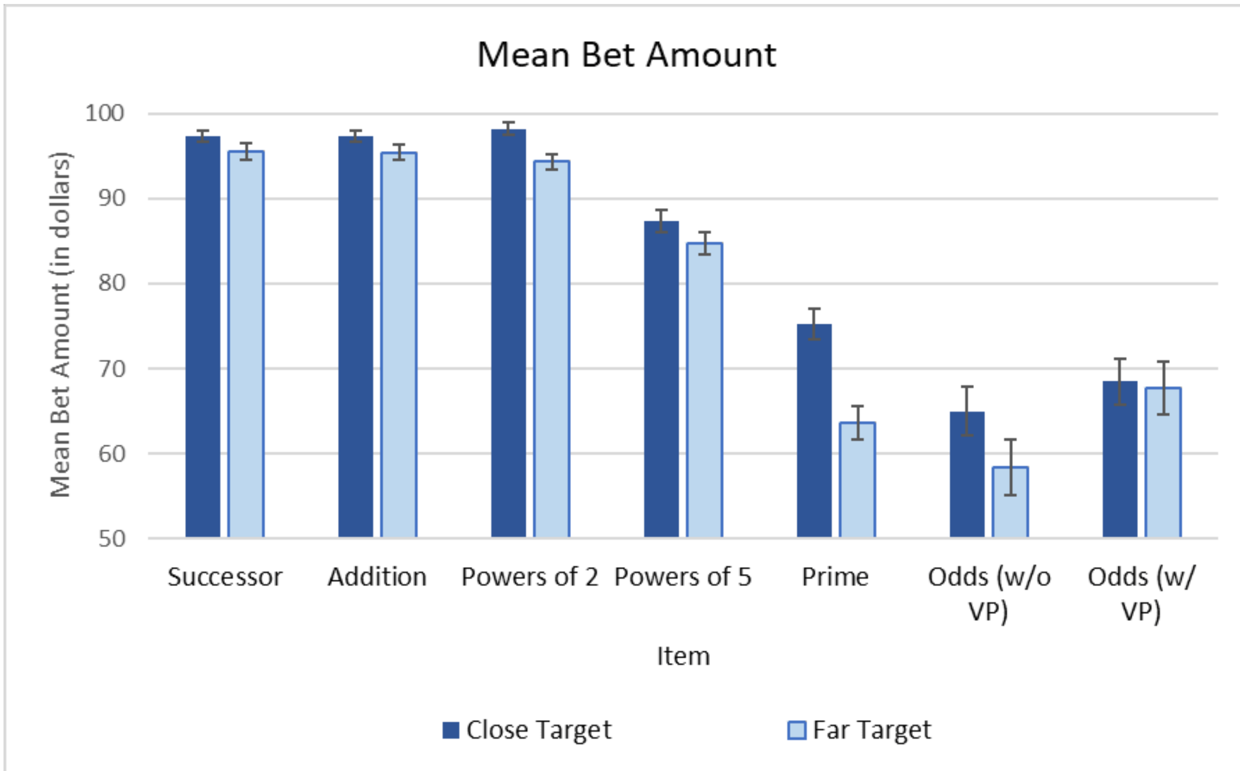


Figure 2.1: Mean bet amounts by item. Error bars show standard error of the mean.

Individual Betting Behaviors: Similarity Effects

In addition to examining mean bet amounts, we were also interested in examining the impact of prediction target magnitude on individual betting behaviors. To identify similarity effects at the individual level, we determined the percent of participants who bet a lower amount on the far prediction target than on the close prediction target for each item. Results are summarized in Figure 2.2. A generalized linear mixed effects model revealed significant differences between the likelihood of betting less on far targets than on close targets on each item, when including Participant and Order as random effects ($\chi^2(6)= 59.65$, $p < 0.001$). Participants were most likely to demonstrate similarity effects (i.e., bet less on far targets than on close targets) in the Prime item (35/74, 47.3%), followed by the Odds items (without visual proof 11/31, 35.5%; with visual proof 9/29, 31%) and the Powers of 5 item (22/74, 29.7%). Participants were relatively less likely to demonstrate similarity effects for the Powers of 2 item (12/74, 16.2%), and particularly the Successor (6/74, 8.1%) and Addition (5/74, 6.8%) Items.

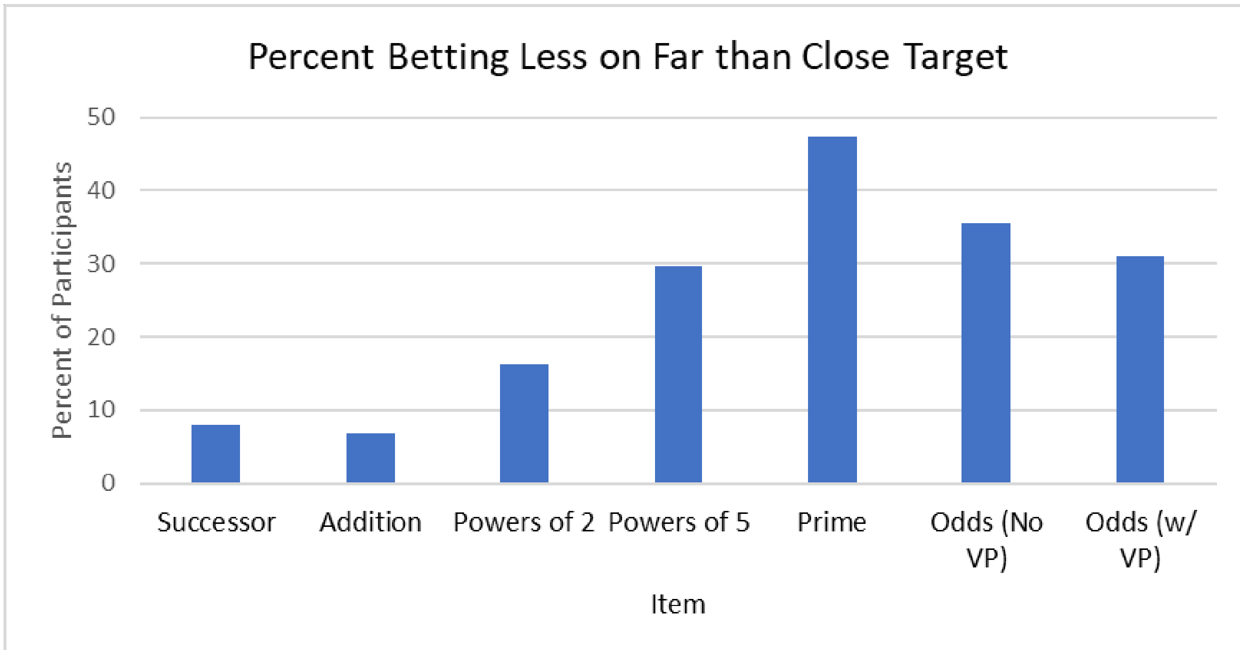


Figure 2.2: Percent of participants on each item who bet a smaller amount on the far prediction target than on the close prediction target.

Individual Betting Behaviors: Maximal Betting

In order to assess participants' recognition of mathematical necessity in each context, we examined the percentage of participants who bet the maximal amount of $100 on each item for close and far targets. This measure is meaningful because we are using betting the maximal amount as a proxy for recognition of mathematical necessity (although, as described above, this likely gives an upper bound of this percent, rather than a direct measure). Figure 2.3 shows percent of participants who bet the maximal amount of $100 in each condition. A generalized linear mixed effects model for binomial data revealed a significant impact of target magnitude on likelihood of betting the maximal amount, when including Participant, Item, and Order as random effects ($\chi^2(1)= 10.76$, p = 0.001). Thus, in mathematical contexts participants were significantly less likely to bet the maximal amount when the prediction target was farther in magnitude from (i.e., less similar to) the observed evidence. Additionally, participants were less likely to bet the maximal amount for more complex mathematical statements; in particular, less than a third of participants bet $100 on predictions for the Odds item (for both close and far targets).

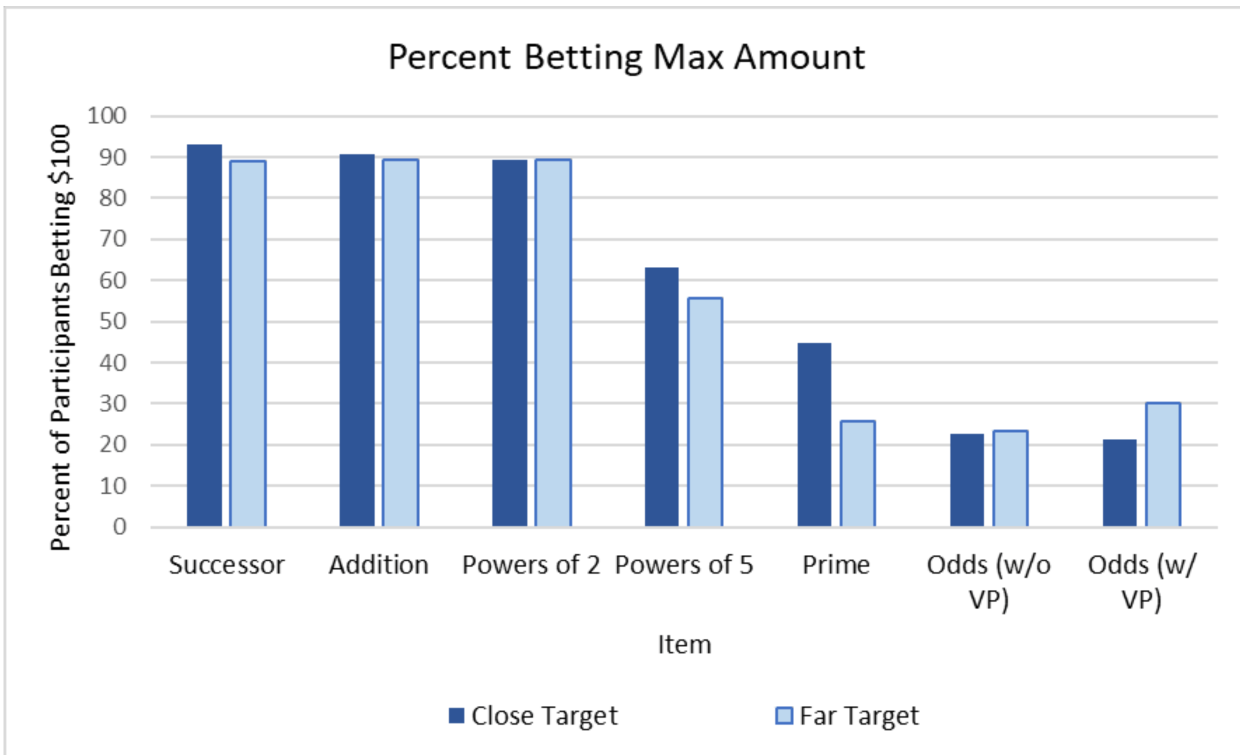


Figure 2.3: Percent of participants who bet the maximal amount of $100, by item and target condition.

Impact of Visual Evidence

Finally, for the odds question we were interested in examining the role of providing additional evidence in the form of a visual proof (this was examined qualitatively in Chapter 1). To investigate the impact of visual evidence on betting behaviors, we created a linear mixed effects model of mean bet amount which included participant and order condition as random effects. When examining only data from the Odds Item, we found that there was no significant improvement in the model when including evidence type (numerical only, or numerical and visual; $\chi^2(1) = 0.66$, $p = 0.42$).

Discussion

In this study we were interested in exploring contexts in which undergraduates may demonstrate reasoning by informal mathematical induction. Informal mathematical induction is distinct from standard inductive reasoning in that it implies recognition of mathematical necessity - an awareness that the generalization must be true for all possible cases. We used two key indicators to distinguish informal mathematical induction from standard inductive generalization: first, the presence of similarity effects, which are a hallmark of inductive reasoning and which we would not expect to observe in informal mathematical induction; and second, patterns of maximal betting, which we take as a proxy for recognition of mathematical necessity and therefore a possible marker of informal mathematical induction. Here we discuss our results with respect to these two key markers; additionally, we discuss the impact of visual evidence on reasoning, as well as limitations and areas for future research.

Presence of Similarity Effects

In inductive reasoning, similarity effects refer to the well-documented phenomenon that people are more likely to generalize an observed pattern to a new case when the prediction case is similar to the observed evidence. Consistent with this, we observed similarity effects in our non-mathematical distractor items. To examine reasoning in mathematical contexts, we presented our undergraduate participants with prediction targets that were either low-magnitude prototypical counting numbers (i.e., more similar to the observed cases of n = 1, 2, 3) or higher-magnitude numbers that were farther from (i.e., less similar to) the observed cases. In general, our observed mean bets (summarized in Figure 2.1) demonstrate varying degrees of similarity effects in contexts of mathematical generalization. We found a main effect of prediction target magnitude on bet amount, which showed that participants tended to bet lower amounts when the prediction target was farther from the observed evidence; however, it is important to note that this effect was much less pronounced for mathematical items than for our non-mathematical distractor items. We also identified a significant interaction of target and item; while mean bets were lower for far targets on all items, this effect only reached statistical significance for the Prime evidence set. This is noteworthy as this was the only item in which the pattern actually does not hold true for all natural numbers and therefore is not an appropriate context for reasoning by informal mathematical induction.

We can further inform our investigation of similarity effects by examining this phenomenon at the level of individual participants. As seen in Figure 2.2, participants were the least likely to demonstrate similarity effects in the simplest mathematical contexts, the Successor and Addition items, which have been identified in previous literature as contexts in which children reason by informal mathematical induction (Baroody, et al., 2013; Smith, 2003; it is worth noting that even for these items, there were still some undergraduates - highly educated adults - who bet less for far targets than for close targets). For more complex mathematical statements, however, we observed an increased number of participants betting a lower amount on the far target than they had for the close target. For the Powers of 5 item and the Odds item, almost a third of our undergraduate participants bet a lower amount for the far prediction target than for the close prediction target. Participants were most likely to demonstrate similarity effects on the Prime item, again suggesting some possible sensitivity to the fact that this statement was not, in fact, true for all natural numbers.

Overall our results show that similarity effects are relatively small in mathematical contexts compared to non-mathematical contexts, are less likely to be observed in simple mathematical contexts, and become more likely and more pronounced as mathematical complexity increases. The fact that similarity effects were strongest for the Prime item suggests that many undergraduates may be able to detect contexts in which it is not appropriate to generalize to all natural numbers. However, as we did not check for comprehension of the Prime item (as we did for the Odds item), it is also possible that these results reflect a general lack of understanding of this particular item; future studies could include more comprehensive comprehension checks, in order to rule out the possibility that similarity effects are the result of general confusion regarding the mathematical statement.

These results leave open the possibility that many undergraduates may reason by informal mathematical induction, particularly in simple mathematical contexts; however, to assess this more thoroughly we must examine not only whether we observe similarity effects, but also whether we have evidence for recognition of mathematical necessity. In what contexts, if any, are participants willing to bet the maximal amount?

Maximal Betting Behaviors as Evidence of Mathematical Necessity

As mentioned above, in this study we used maximal betting as a proxy for recognition of mathematical necessity; thus, in order to determine in which contexts participants may be reasoning by informal mathematical induction, it is critical to examine when and to what extent participants are willing to bet the maximal amount on the target prediction. Overall, our results (summarized in Figure 2.3) showed that most - though not all - undergraduate participants bet the maximal amount regardless of prediction target magnitude in the simplest mathematical contexts: Successor, Addition, and Powers of 2. As mathematical complexity increases, participants become significantly less likely to bet the maximal amount; for instance, in the Odds item, only between 20-30% of undergraduate participants bet $100, even for close prediction targets and when visual evidence was provided. This pattern of results suggests that, while many undergraduates may engage in informal mathematical reasoning in the simplest of contexts, they become significantly less likely to demonstrate awareness of mathematical necessity as complexity increases. This is particularly noteworthy, as our participants were all highly educated young adults and none of our mathematical statements involved concepts more advanced than arithmetic and basic number knowledge.

These results call into question various claims which have been made in developmental psychology regarding the emergence of a capacity to reason by a form of mathematical induction which does not require training in formal mathematical proof. Our results do not rule out Baroody et al.’s (2013) and Smith’s (2003) assertions that children may be reasoning by informal mathematical induction in the context of the successor principle and other foundational arithmetic principles. However, the fact that undergraduates do not consistently engage in this form of reasoning in even slightly more complex arithmetic contexts should give us pause in considering informal mathematical induction to be a general or widespread phenomenon, even among highly educated adults. Instead, our results support the conclusion that informal mathematical induction is restricted to the most basic of mathematical concepts, and that recognition of mathematical necessity in more advanced contexts likely requires having been trained in formal mathematical proof, including formal mathematical induction.

Effect of visual evidence

Finally, we were interested in assessing the potential role of visual evidence in promoting reasoning by informal mathematical induction in more complex mathematical contexts. In our previous chapter, we found that even when presented with a 'visual proof by induction' for a theorem - specifically, the statement that "The sum of the first n odd numbers is equal to n^2" - undergraduates did not recognize the mathematical necessity of the statement. Our results in this study are consistent with that finding - though mean bets were somewhat higher on the Odds question when visual evidence was provided (Figure 2.1), we found no statistically significant impact of providing visual evidence on mean bet amounts. Interestingly, we did find that participants who saw visual evidence were slightly less likely to demonstrate similarity effects at the individual level (Figure 2.2), and in fact slightly more likely to bet the maximal amount on far targets than close targets (Figure 2.3). However, none of these results support the notion that this form of visual evidence allows all educated adults, regardless of formal proof experience, to recognize the necessity of the mathematical statement, as has been claimed by various philosophers of mathematics (Brown, 2008; Chihara, 2004)

Alternatively, the fact that the impact of visual evidence on mean bets did not reach statistical significance, and more generally did not seem to support reasoning by informal mathematical induction in the majority of participants, may also reflect a lack of attention to the evidence. Mean reading times showed that our participants spent approximately the same amount of time (around 22-23 seconds) reading the evidence for the Odds item, regardless of whether the visual evidence was present or not. Thus, it is possible that our participants simply did not put in the extra effort required to grasp the visual proof and its connection to the rest of the evidence being presented. Future studies could be designed in such a way as to require participants to demonstrate understanding of the visual evidence, in order to more accurately determine its impact on reasoning. We predict, however, that such studies would show similar results (for instance, in our previous chapter we discovered that even participants who accurately and thoroughly explained the image were unlikely to assert the necessity of the theorem it represented in a follow-up interview).

Future Work

In addition to the improvements noted above, there are some additional avenues for future research which could be fruitful. First, in this study we included only participants who had no university-level classroom training in formal mathematical induction. It would be informative to conduct the same study using participants with knowledge of formal mathematical induction; this would provide a relevant "baseline" regarding how those with explicit knowledge of necessity in mathematical generalization would respond to our particular task. In particular, examining this group of participants could shed light on the use of maximal betting as a proxy for recognition of mathematical necessity.

Second, future work could make use of other phenomena in inductive reasoning - which we would not expect to see in informal mathematical generalization - in order to continue to disentangle these two forms of reasoning. This study exploited a known phenomenon in inductive reasoning - similarity effects - to distinguish between contexts in which undergraduates reasoned by standard inductive reasoning and those in which they may engage in informal mathematical induction. However, similarity effects are only one of many documented characteristics of inductive reasoning. A particularly useful phenomenon to explore in future work would be prototype effects, in which participants are more likely to generalize to other members of the category when their evidence includes prototypical category members (for instance, we are more likely to generalize to all birds based on observations of robins, rather than penguins; Carey, 1985; Rips, 1975; Rosch, 1983). A similar line of research making use of prototype effects could provide further information about the unique role of the prototypical counting numbers in generalization in mathematics, including reasoning by informal mathematical induction. More generally, this work could help to further illuminate the varied mechanisms at play in mathematical generalization in the domain of natural number, and the ways in which generalization practices of even highly educated adults deviates from formalized mathematical induction.

Conclusion

Developmental researchers (Smith, 2003; Baroody et al., 2013) have assumed, on the basis of very little evidence, that young children are able to engage in informal mathematical induction. Specifically, this body of work has claimed that even without training in formal mathematical proof, children are able to recognize the necessity of generalizations within the domain of natural number. However, existing research has failed to thoroughly assess subjects' recognition of mathematical necessity, and therefore leaves open the possibility that participants are simply engaging in standard inductive reasoning within the domain of natural number. In Chapter 1, we found that undergraduates without training in formal mathematical induction did not engage in informal mathematical induction, even when provided with a visual proof that demonstrated the necessity of the theorem. However, that study included only one relatively complex mathematical statement and relied on an interview-based protocol, and thus did not rule out (1) social dynamics as a confounding factor in undergraduates' responses, and (2) the potential that undergraduates may reason by informal mathematical induction in simpler contexts. The present study was designed to address these shortcomings. Our web-based task used a betting paradigm to investigate contexts in which undergraduates demonstrated evidence of reasoning by informal mathematical induction, including two key markers: first, a willingness to bet the maximal amount on target predictions, and second, a resistance to similarity effects (a known phenomenon in inductive reasoning). Our results suggest that most - though not all - undergraduates recognize the mathematical necessity of the most basic arithmetic relationships (e.g., the successor principle). However, in contexts of even slightly increased mathematical complexity, many undergraduates chose not to bet the maximal amount on target predictions, and demonstrated similarity effects in their betting behaviors (i.e., bet lower amounts for far targets than for close targets). Overall, these findings call into question the notion that informal mathematical induction is a widespread form of reasoning which can be deployed in novel contexts. Rather, for highly educated adults without formal training in formal mathematical proof, informal mathematical induction appears to be quite an isolated phenomenon which occurs only in the most foundational and familiar contexts of basic number facts. Future research in number concept development should take care not to over-apply labels like "informal mathematical induction" before performing systematic examinations of the phenomena, including subjects' recognition of mathematical necessity.

References

Baroody, A. (2005). Discourse and research on an overlooked aspect of mathematical reasoning. *The American Journal of Psychology, 118*(3), 484–489.

Baroody, A. J., Reid, E. E., & Purpura, D. J. (2013). An example of a hypothetical learning progression: The successor principle and emergence of informal mathematical induction. WISDOMe Monograph, 3.

Brown, J. R. (2008). *Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures (Second ed.)*. New York: Routledge.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press, Bradford Books.

Chihara, C. S. (2004). *A structural account of mathematics.* Clarendon Press.

Length, R.V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.0. https://CRAN.R-project.org/package=emmeans

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329.

Koch, C., & Preuschoff, K. (2007). Betting the house on consciousness. *Nature Neuroscience, 10*(2), 140– 141.

Osherson, D. N., Stern, J., Wilkie, O., Stob, M., & Smith, E. E. (1991). Default probability. *Cognitive Science, 15*(2), 251-269.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185.

Persaud, N., & McLeod, P. (2008). Wagering demonstrates subconscious processing in a binary exclusion task. *Consciousness and Cognition, 17*(3), 565– 575.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience, 10*(2), 257– 261.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 665-681.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences, 31*(6), 623-642.

Rosch, E. (1983). Prototype classification and logical classification: The two systems. New trends in conceptual representation: *Challenges to Piaget's theory*, 73-86.

Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature, 36*(4), 2021-2064.

Smith, L. (2003). Children's reasoning by mathematical induction: Normative facts, not just causal facts. *International Journal of Educational Research, 39*(7), 719–742.

Studer, B., Limbrick-Oldfield, E. H., & Clark, L. (2015). 'Put your money where your mouth is!': effects of streaks on confidence and betting in a binary choice task. *Journal of Behavioral Decision Making*, *28*(3), 239-249.

Suhonen, N., & Saastamoinen, J. (2018). How do prior gains and losses affect subsequent risk taking? New evidence from individual-level horse race bets. *Management Science*, *64*(6), 2797-2808.

Wohl, M. J., Branscombe, N. R., & Lister, J. J. (2014). When the going gets tough: Economic threat increases financial risk taking in games of chance. *Social Psychological and Personality Science*, *5*(2), 211-217.

CHAPTER 3

Are There Infinitely Many Natural Numbers? Differential Prominence of the Successor Principle in Math- and Non-Math Majors' Justifications

Introduction

How we come to have abstract concepts of number is a central question in the study of the mind. In particular, developmental psychologists have sought to characterize how children come to develop concepts of natural number, which is assumed to arise from mastery of counting. Central to this research program is the claim that it is understanding of the successor principle – the fact that for any natural number n, the next natural number (and the next item in the verbal count list) is n + 1 – which is key to "mature" conceptualizations of natural number. Specifically, developmental psychologists suggest that it is understanding of this principle that allows children to develop genuinely abstract concepts of larger-magnitude numbers (Sarnecka & Carey, 2008). It is also thought that an inductive generalization of the successor principle allows children to recognize that the natural numbers are infinite (since any number, no matter how large, has a successor; Cheung et al., 2017; Rips et al., 2008).

In the preceding chapters, we have presented evidence that college undergraduates – highly educated adults at a prestigious university who certainly have "mature" conceptualizations of number – do not always make use of higher-level implications of the successor principle in spontaneous, real-time reasoning (Relaford-Doyle & Núñez, 2018). This calls into question the centrality of the successor principle in adults' conceptualizations of the natural number system. In the present study, we further examine the relationship between the successor principle and the concept of infinity in college undergraduates both with and without formal mathematical training in the axioms of natural number. We address limitations of existing research of the successor principle by empirically examining the "mature" (i.e., fully developed) concept in adults, and by using a novel method that allows us to examine a diversity of conceptualizations.

Background

According to major developmental accounts, concepts of natural number arise from mastery of the counting procedure. By the bootstrapping account (Carey, 2000), children initially learn the count sequence as a memorized verbal list ("one, two, three"). With counting practice, and building on innate preconditions for number concepts such as subitizing ability (Nuñez, 2017; Kaufman et al., 1949), they eventually match the smallest items in the list, "one", "two", and "three" to the quantities they represent. By around 5 years of age children notice a pattern: that for the small numbers, moving up in the count list represents an increase of exactly 1 item. They then apply this rule to the numbers in their full count list, and eventually to "all possible numbers" (Cheung et al., 2017), realizing that any number in the count sequence, no matter how large, has a successor that is one greater. This insight is considered to be the key to children's realization that the natural numbers are an infinite set (Rips et al., 2008), marking the emergence of "mature" conceptualizations of natural number and the end of the developmental trajectory investigated by most researchers of number concepts.

However, research into the role of the successor principle in natural number concepts has been limited in key ways. First, the majority of experiments that have focused on children's understanding of the successor principle have relied on tasks involving particular numbers, frequently focusing on numbers smaller than ten. For instance, Sarnecka & Carey (2008) investigated when children develop an understanding of the successor principle using a task which required children to identify successors for the numbers one through six. Studies such as these tell us very little about children's understanding of the successor principle as a general rule characterizing the entire set of natural numbers. When researchers have assessed children's understanding of the successor principle in "large numbers" (24 and 25; Davidson et al., 2012) and "very large numbers" (53, 57, 76, 77; Cheung et al., 2017) they have found evidence suggesting that children initially generalize the successor principle only to the smaller numbers in their count list, and only later to larger numbers. This piecewise generalization of the successor principle – apparent even in the set 1 - 100 – calls into question the claim that children's generalization of the successor principle to "all possible numbers" is a process that straightforwardly gives rise to concepts of infinity.

Second, when researchers have tried to assess children's understanding of the successor principle as a general rule governing the entire set of natural numbers, they have relied on highly constrained interview protocols that include leading questions. For example, Cheung et al. (2017) supplemented their study of children's knowledge of the successor principle in "very large numbers" with an interview probing whether they understood that the natural numbers extend infinitely. The following transcript (Cheung et al., 2017, Appendix B) shows a conversation between an experimenter and a 5-year-old child. On the basis of this interview, the child was classified as understanding that every natural number has a successor and thus that the natural numbers are infinite.

> (1) Experimenter: So if we thought of a really big number, could we always add to it and make it even bigger, or is there a number so big we couldn't add any more?
>
> (2) Child: No, we can add to it.

(3) E: Why is that?

(4) C: Because there would be a million, and then you could keep adding numbers if you know them.

(5) E: So you said that the biggest number you know is a million. Is it possible to add one to a million, or is a million the biggest number possible?

(6) C: You can add one to it.

(7) E: Why?

(8) C: Because it would be like a million one, a million two.

(9) E: Could I keep adding one?

(10) C: [child nodded]

(11) E: Why?

(12) C: Because, um, because if you want to you can keep adding and adding and make it fun.

Suppose this child had also earlier demonstrated that they could consistently identify the successor for all the numbers in their count list, including 'very large numbers' like 77. In that case, the authors would conclude that this child understood "the successor function as defined in the Peano axioms—i.e., that every number n has a successor n + 1, such that numbers never end" (p. 25). Does the evidence support this claim? Notice that it is the experimenter, not the child, who first mentions adding one as a way to make a number bigger (line 5)—that is, the experimenter explicitly mentions the successor function in their question. In five separate responses regarding how to make bigger numbers (lines 2, 4, 6, 10, 12), the child only explicitly mentions adding one once (6), immediately following the experimenter's mention of it in (5). In most of their responses, the child doesn't specify a number to be added, and the child's response in line 4 ("keep adding numbers") indicates that the child may in fact recognize that any number will do the job. Thus, by our reading, this child seems to understand two separate facts: first, that counting represents an increase by one each time ("a million one, a million two", line 8), and second, that you can always generate bigger numbers by adding. We see no evidence that this child's understanding of the unbounded nature of natural numbers relies uniquely on the successor principle. That is, this child may understand the logic of counting, but there is no evidence that the same logic governs the child's understanding of the infinite set of natural numbers.

Additionally, in interviews with children researchers have frequently sought only positive evidence for understanding of the successor principle, and ignored responses suggesting that children may actually have a variety of alternate conceptualizations of the number system. For instance, consider the following interview with a 7-year-old child (Hartnett & Gelman, 1998):

> (1) Experimenter: If I thought of a really big number, could I always add to it and get a bigger number? Or is there a number so big that I couldn't add anymore; I would have to stop?
>
> (2) Child: You could always make it bigger and add numbers to it.
>
> (3) E: If I count and count and count, will I ever get to the end of the numbers?
>
> (4) C: Uh uh.
>
> (5) E: Why not?
>
> (6) C: Because there isn't one. . . .
>
> (7) E: How can there be no end to the numbers?
>
> (8) C: Because you see people making up numbers. You can keep making them, and it would get higher and higher. . . .
>
> (9) E: You mean there's no last number?
>
> (10) C: Uh uh. 'Cause when you get to a really high number, you can just keep on making up letters and adding one to it.
>
> (11) E: Okay. Well, what if somebody came up to you and said that a googol was the biggest number there could ever be? . . . When you make a googol, you write a one and a hundred zeroes. It's a really big number. So if somebody came up to you and said, "I think a googol is the biggest number there could ever be", would you believe him?
>
> (12) C: Uh uh. Because you could put one and a thousand zeroes on it.

Based on this interview, this child was classified as understanding the successor principle. However, the child's responses clearly show that their understanding of the unboundedness of numbers does not rely *exclusively* on the successor principle. In fact, in this short exchange the child actually offers four different explanations for how to generate bigger and bigger numbers: adding numbers (line 2), making them up (line 8; this might refer to children's use of words like "a bazillion" and "a gajillion" to refer to bigger and bigger numbers); adding one (line 10), and appending digits to create longer numeral strings (line 12). This child clearly understands that there is no largest number, and that adding one is just one way to produce a larger number, but this interview does not demonstrate that the child's understanding of the unboundedness of natural number relies exclusively on the successor principle.

Finally, existing studies have focused only on the development of number concepts in very young children, and generally take understanding of the successor principle as evidence for a "mature" conceptualization and thus the endpoint of this developmental trajectory. But the developmental story does not end here. After learning to count, children soon encounter a variety of other numbers systems that follow different rules than the natural numbers: for instance, in even and odd numbers the successor is obtained by adding 2; the rational numbers have no successor principle at all. As studies into number concept development typically stop around age 6, very little is known about adults' understanding of the infinite nature of natural number, and the role of the successor principle in this knowledge.

In the present study we address the limitations outlined above to provide a more comprehensive account of the role of the successor principle in mature conceptualizations of natural number. Do adults use the successor principle to justify that there are infinitely many natural numbers? If not, what other types of justifications do they provide? We examine the types of justifications provided by undergraduates with no particular mathematical background as well as the types of justifications offered by mathematics and computer science majors who have received explicit training on the successor principle (as formalized in the Dedekind-Peano axioms of natural number). This allows us to compare knowledge of the successor principle in people who (according to developmental accounts) would have spontaneously arrived at this insight through inductive generalization, to that of people who received explicit mathematical instruction on this formal property of natural number. If learning the successor principle is a key part of learning how to count, and if understanding of the unboundedness of natural number relies uniquely on this principle, then we would expect both groups to use the successor principle to justify that the natural numbers are infinite. However, if spontaneously-arising conceptualizations of natural number as an infinite set do not rely exclusively on the successor principle, we would expect undergraduates who have not received explicit instruction to offer a wider variety of justifications (similar to the child interviewed in Hartnett & Gelman, 1998, described above).

Method

Materials: A training task made use of a set of 9 rectangular white paper cards, each of which was printed with a simple black line drawing of a well-known animal (e.g., cat, whale, chicken). On the main task, participants were given a set of 32 rectangular cards, each of which displayed a black numeral printed on white paper. The numerals represented numbers from multiple categories, including natural numbers of varying magnitudes, rational numbers written as fractions, negative numbers, and decimals. The list of numerals is 1; 2; 4; 8; ½; ¾; 123 ; 1310; 3.5; 0.4; 9.75; 0.01; -3; -8; -1; -0.2; -½; -152; -645; -0.03; 0.0000004; 11; 5.333…; 9,274,976; 3; 102,375; -0.18; -9; 625; 56; 1.191919…; and 150. Participants were seated at a table next to the researcher in a quiet room. A video recording camera was positioned on the table to capture the participants' workspace, as well as the audio and manual gestures during each interview.

Participants: Participants were drawn from two distinct populations. The first group (n = 17) was recruited through the mathematics department and consisted of mathematics and computer science majors. All of these students had taken "Mathematical Reasoning", an upper-division mathematics course that covers proof-writing (or an equivalent course at a previous institution). This course includes explicit instruction in the successor principle as formalized in the axioms of natural number used in the proof method of mathematical induction. Because they had received explicit instruction at the college level in the successor principle, we refer to this group as Successor-Principle-Trained (SP-Trained). The second group (n = 26) was recruited from the general university subject pool and consisted of undergraduates from a wide variety of majors including Cognitive Science, Psychology, Engineering, Biology, and Economics. While many of these students had taken college-level courses in calculus and statistics, none had taken a college-level course covering the axioms of natural number. Because these participants had not received explicit instruction in the successor principle at the college level we refer to this group as Successor-Principle-Untrained (SP-Untrained). SP-Untrained participants received course credit for their participation; SP-Trained participants received $10 as compensation.

Procedure: Participants first completed a brief open-ended sorting task, which served as the basis for the subsequent semi-structured interview.

Sorting Task: The researcher presented participants with a set of printed cards and asked them to arrange the cards into groups based on whatever criteria they chose ("in whatever way makes sense to you"). The researcher emphasized that there were no right or wrong answers, and that the participant could take as long as they wanted to complete the task. In order to familiarize participants with the task, and to alleviate possible anxieties surrounding a mathematical task, participants were first given the set of nine cards showing pictures of different animals. Participants arranged these cards into groups, and then explained their groups to the researcher. The researcher then asked the participant about other possible arrangements and emphasized that there were multiple possible arrangements and that any grouping would have been acceptable.

Next, the participant was given the set of 32 cards in random order, each of which showed a different number (written as Arabic numerals). Once again, participants were asked to arrange these cards into groups in whatever way they chose, and given as much time as they needed to complete the task. When they indicated they had finished, the researcher asked them to describe each of the groups they had created. If the participant had created a unique group containing the natural numbers, the researcher began the interview. If the participant had not created a group for the natural numbers, the researcher said, "I'm going to arrange the cards a different way and see how you'd describe the groups I make." The researcher then arranged the cards such that all the natural numbers were in a group and asked the participant what they would call this group. From here, the researcher began the interview, using the terminology that was provided by the participant in their explanations.

Interview: The researcher told the participant that they were now going to be talking about just one of the groups, which they referred to using the participant's language (for instance, "positive whole numbers", "positive integers", or "natural numbers"). To allow the participant to focus on just this group, all other cards were removed from the workspace and the cards showing natural numbers were arranged centrally on the desk, grouped randomly (Figure 3.1). The researcher then proceeded with the interview using the following protocol:

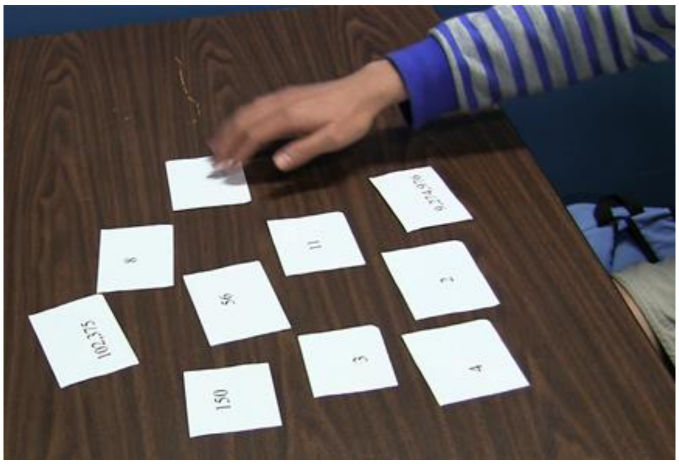


Figure 3.1: Cards showing natural numbers arranged haphazardly in front of participant.

- (Q1) "I'm going to ask you some questions about this group, the one you called the [natural numbers/positive integers/positive whole numbers]. This group has numbers in it like 2 and 8 and 56 [points to these cards on table]. What are some other numbers that would go into this group?"
- (Q2) "Great. So you said that some other numbers like [repeated participants' responses]] would go into this group. My next question is, how many numbers are there that could go into this group?"

If the participant's response indicated knowledge that there are infinitely many numbers that could go into the group, the researcher continued with the next question; otherwise, the interview terminated.

- (Q3) "Okay, so you said there are infinitely many numbers that could go into this group. Let's say you had a friend who disagreed with you, who didn't think that there were infinitely many [natural numbers/positive integers/positive whole numbers]. You want to convince your friend that you're right. How would you prove to your friend that there really are infinitely many numbers that would go into this group?"

If the participant referred to the successor principle in any form ("adding one", "plus one", "give you one more", etc.) in their answer, the interview terminated here. If the participant did not refer to the successor principle, the researcher rearranged the cards on the table so that they were laid out horizontally in order of magnitude, beginning with the cards 1, 2, 3, 4 (Figure 3.2). This was intended to make counting salient, and to highlight the +1 relationship between successive numbers. The researcher then said,

- (Q4) "Okay, great! Now I'm going to arrange the cards a different way to see if that gives you any other ideas about what you might say to your friend. [Rearrange cards] So now we have the cards in order, starting with 1, 2, 3, 4, then 5, 6, 7, 8 [pointing successively to each card or space], and on and on [sweeping gesture upwards and to the right]. Does seeing the cards like this give you any other ideas about what you could say to your friend?"

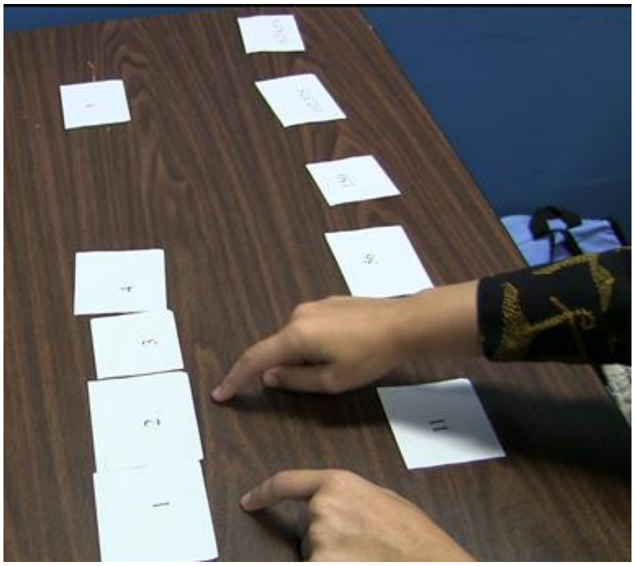


Figure 3.2: Cards rearranged in order of magnitude, to highlight the +1 relationship between successors. The researcher used speech and gesture, pointing successively at each card as seen above, to make salient the idea of counting.

During the conversation, the semi-structured nature of the interview allowed the researcher to clarify questions as needed and ask additional follow-up questions to probe participants' reasoning as thoroughly as possible. For instance, if the participant answered Q3 or Q4 by simply restating that there are infinitely many natural numbers, the interviewer asked, "Is there a rule about numbers that you could use to explain why this is true?" If a participant said that they would ask their friend to name the biggest number, the interviewer responded (as in Hartnett & Gelman, 1998): "Okay, let's say your friend says that the biggest number is googol. That's a one with a hundred zeroes after it. How would you respond to your friend?" In this way, the researcher provided as many opportunities as possible for the participant to refer to the successor principle in their justifications. Once the participant indicated that they had nothing further to add to their explanation, the researcher thanked them for their time and the interview terminated.

Analysis

Two raters, the researcher and a research assistant who had been trained in the coding criteria, independently coded interview responses for three measures:

1. Infinity: In their response to Q2, did the participant state that there were infinitely many natural numbers? This was coded as either yes or no.

2. Initial Justification: In response to Q3, how did the participant initially justify that the natural numbers are infinite? We identified and coded for four distinct strategies:

- **Successor Principle**: The participant refers to the successor principle in their justification (e.g., “You can always add 1 so you can never get a biggest number.”)
- **Other Operation**: The participant refers to a different arithmetic operation (e.g., “Take whatever number’s biggest and square it.”)
- **Notation-Based**: The participant refers to appending digits to the numeral to create a longer numeral string (e.g., “Put another 0 on the end.”)
- **None/Restating**: Participant restates that numbers are infinite but does not provide a mathematical justification for this fact (e.g., “The numbers never end, so there’s no biggest number.”)

If a participant offered multiple justifications in their response, each justification was noted and included in our analysis.

3. Post-Rearrangement Justification (if applicable): For participants who were asked Q4, upon seeing the cards rearranged in order of magnitude, what additional explanations did the participant provide (if any)? This was coded in the same manner as Item 2, above.

On Item 1 (infinity), the two raters showed 100% agreement. On Item 2 (Initial Justification) and 3 (Post-Rearrangement Justification), there was strong inter-rater reliability (92% agreement, Krippendorff's Alpha = 0.897). Disagreements in coding were resolved through conversation.

Results

All SP-Trained participants and all but one SP-Untrained participant (96.3%) stated that there were infinitely many natural numbers. However, when asked how they would prove that fact, the two groups showed different patterns of responses (Figure 3.3).

When first asked how they would prove that there are infinitely many natural numbers, SP-Untrained participants were significantly less likely than SP-Trained participants to refer to the successor principle (SP-Untrained: 5/26, SP-Trained: 12/17; $\chi^2 = 11.34$, $p < 0.001$). SP-Untrained participants were equally likely to mention another type of operation (such as squaring, multiplying, or adding two successive numbers) or to provide a notation-based explanation as they were to mention the successor principle. The majority (9/17) SP-Untrained participants did not provide a mathematical justification in their initial response, instead simply restating that there are infinitely many natural numbers (see next section for illustrative examples of participant responses).

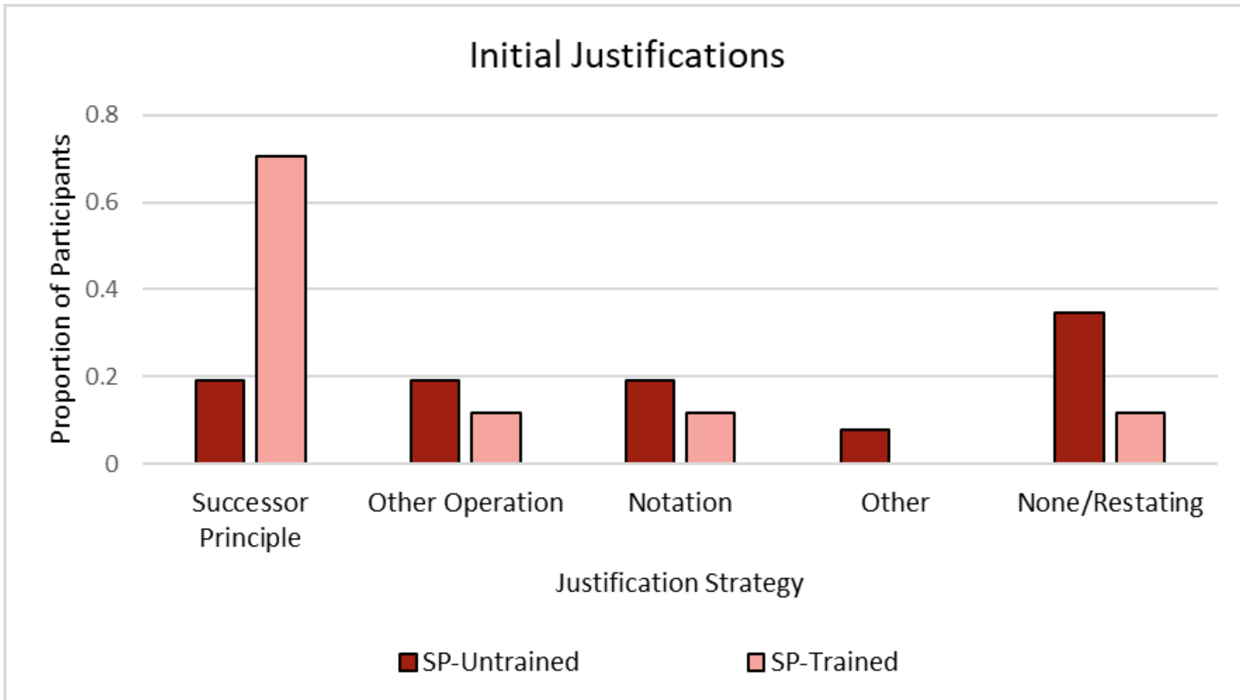


Figure 3.3: Initial justifications provided by SP-Untrained and SP-Trained participants when asked how they would prove that there are infinitely many natural numbers. SP-Trained participants overwhelmingly referred to the successor principle. SP-Untrained participants showed wider variability in their responses and were less likely than SP-Trained participants to refer to the successor principle in their justifications.

Participants who did not mention the successor principle in their initial justifications were given another opportunity to respond after the cards had been rearranged by order of magnitude, thus highlighting the +1 relationship between successors. All 5 of the remaining SP-Trained participants mentioned the successor principle at this point; thus, over the course of the interview 100% of SP-Trained participants referred to the successor principle as a means of proving that the natural numbers are infinite. On the other hand, of the 21 SP-Untrained participants who saw the cards rearranged, only an additional 5 mentioned the successor principle in their responses (Figure 3.4). Thus, even when the +1 relationship between successors was made salient through card placement, speech, and gesture, only 10/26 (38%) of SP-Untrained participants ever referred to the successor principle as justification that the natural numbers are infinite. This is significantly below what we'd predict if we expected all college undergraduates to refer to the successor principle in their justifications (Binomial Test, $p < 0.001$). In fact, this is significantly less than what would be predicted even if we expected, say, only 60% of undergraduates to use the successor principle (Binomial Test, $p = 0.02$). Notably, 4 out of the 26 SP-Untrained participants (15%) never provided a mathematically valid justification for why the natural numbers extend infinitely, even after having seen the cards rearranged to make the idea of counting more salient.

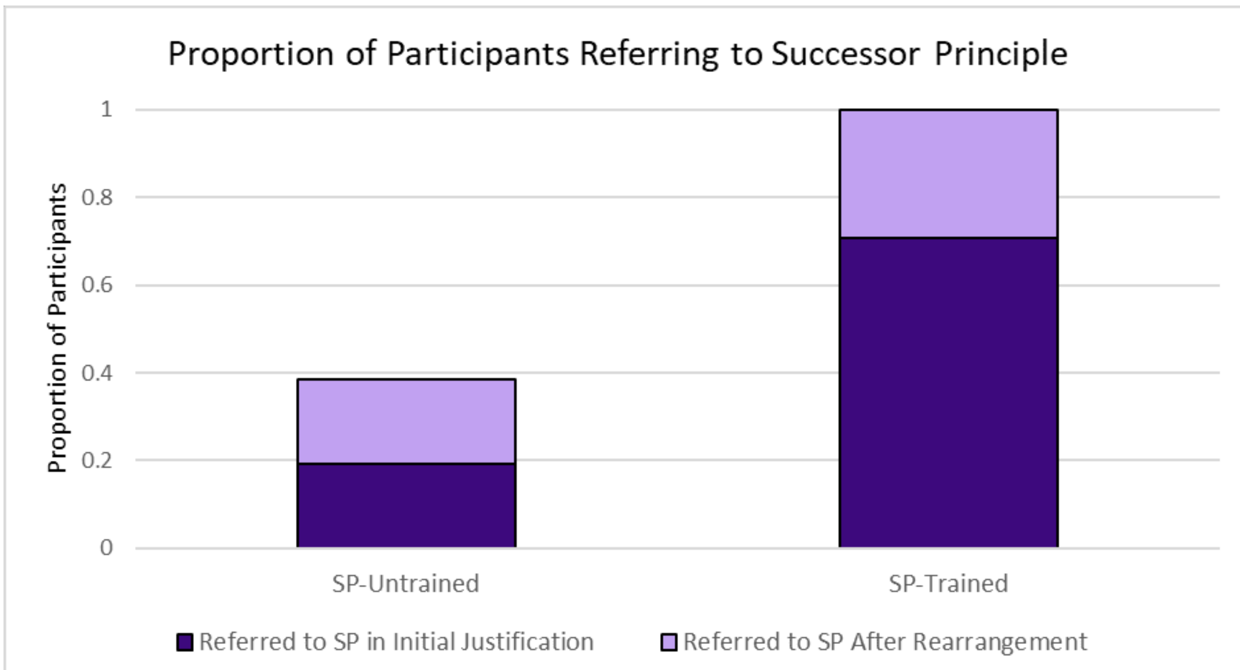


Figure 3.4: Proportion of participants referring to the Successor Principle (SP) during the interview (during initial justification or after rearrangement). All SP-Trained participants referred to the Successor Principle at some point during the interview, while only 38% of SP-Untrained participants did so.

To demonstrate the diversity of explanations, as well as themes which emerged in each group, below we provide some illustrative examples of the justifications that participants provided.

SP-Trained Participants. The majority of our SP-Trained participants referred to the successor principle in their initial justifications. In some cases, participants provided simple explanations which did not refer directly to any technical or formal mathematical notions. For instance, when asked how he would prove to his doubtful friend that there are infinitely many natural numbers, SP-Trained Participant A (PA) responded:

> *PA:* "I'd ask my friend to think of the absolute largest number that they can. Once they gave the answer, I'd say 'Okay in that case, there are that many numbers that can go into this group.' If the person agreed with that, because that's the largest number, I'd say, 'Okay, can you add one to that number?' And if they can, which they should be able to, I'd say 'Okay so in that case you're wrong, *now* we have the largest group'. And add on to that again. And then [gestures with the right hand, iteratively circling rightward] - proof."

When the interviewer followed up by asking the participant to consider a case where their friend had claimed that googol was the largest number, Participant A responded, "I'm going to take out the last zero and just put a one there and ask, okay, is this number larger than googol?" Thus, this participant referred to the successor principle in both general terms and when given a specific example to consider.

Other SP-Trained participants whose initial justifications made use of the successor principle referred to technical mathematical concepts in their explanations. For instance, SP-Trained Participant B (PB) provided a similar argument as Participant A, but included a reference to the formal strategy of proof by contradiction:

> *PB:* "I would just go to the classic proof, I would do a rigorous mathematical proof, you know, use contradiction. *N*+1 must belong, right, since *n* is supposed to be the biggest. That's it."

When prompted by the interviewer to walk through the explanation in more detail, Participant B responded:

> *PB:* "It's gonna be like, to prove that there are infinite positive integers in the set... what we usually do is like assume there are finite, and then say let *n* be the largest in that group. But we can always add one or two, whatever, to that *n* to get this contradiction, then when you go back the other one must be true instead."

Notably, while this participant's first explanation clearly refers exclusively to the successor principle ("*N*+1"), their follow up explanation makes it clear that any value could be added to create a larger number ("add one or two, or whatever"), and therefore arrive at the needed contradiction. This was a theme among SP-Trained participants, who frequently noted that there are infinitely many ways to generate a larger number.

Multiple SP-Trained participants explicitly referred to formal mathematical induction when providing their explanations. For instance, SP-Trained Participant C explained:

> *PC*: "I guess I would use a mathematics proof, induction step. So assume like if you have [pauses] - technically this is very, like, if you have never learned mathematics you might find this really counterintuitive or something, but I guess I would still try. So if you have a number, like an unknown *x*, you can still find another number which is larger. So for any number that is a positive integer *x*, you can find a number, a positive integer that is I guess maybe one larger than *x*, which is *x*+1. So *x*+1 is larger than *x*. And for *x*+1 you can still find another number, which is *x*+2, which is larger than *x*+1. So...for any number you can find a larger number which means this never ends, this rule never ends, for any numbers. So I guess there would be infinitely many positive integers."

Finally, a relatively small number of SP-Trained participants did not refer to the successor principle in their initial explanations, doing so only once the cards were rearranged to make the relationship between successors more salient. For instance, SP-Trained Participant D (PD) initially provided an explanation based on multiplication (and like Participant B, noted that there are infinitely many ways to generate a larger integer):

> *PD*: "If it is like finite number of integers, that just doesn't make sense. I can multiply all the integers [gestures to cards on table] and that will give me another integer. And then I can multiply them again, give me another integer. There are multiple ways, infinite ways to create an integer based on my group."

After the interviewer rearranged the cards, Participant D offered an additional explanation which referred to the successor principle and also made reference to formal mathematical induction:

> *PD*: "This is kind of like an induction. If the biggest number of integer is x, there must be another number x+1 which is also an integer, and x+1 is bigger than x. So it means there are infinitely many."

These examples demonstrate that, in addition to reliably referring to the successor principle, SP-Trained participants frequently noted that there are in fact an infinite number of ways to prove that there is no largest number. SP-Trained participants were also likely to refer to technical mathematical concepts in their explanations, including proof by contradiction and mathematical induction.

SP-Untrained Participants

As mentioned above and shown in Figure 3.3, only a relatively small number of SP-Untrained participants - highly educated undergraduates - referred to the successor principle in their initial justifications. While SP-Trained participants generally mentioned the abstract arithmetic operation of adding one to produce a bigger number, SP-Untrained participants who *did* refer to the successor principle were more likely to provide examples of counting or enumerating concrete objects. For instance, when asked how he would respond to a friend's claim that googol was the biggest number, SP-Untrained Participant E (PE) offered the following:

> *PE*: "I would say, okay, so say you had a googol cookies. If I gave you one more, would that shatter the fabric of reality?"

Some SP-Trained participants provided initial justifications which did not rely on the successor principle, but instead referred to other mathematically valid ways of producing larger numbers. SP-Untrained Participant F (PF) initially offered a notation-based explanation of why there are numbers larger than googol:

> *PF*: "These are just names for numbers. Numbers, you can simply invent a new name for numbers if you like...You can even have zeroes beyond googol. I'd just ask him to write googol, or pull it on the computer. And I'd just copy the number and add zeroes, and say, 'Okay, is this not a number?'"

Upon seeing the cards rearranged on the table in order of magnitude, PF provided an additional explanation, this time referring to the successor principle:

> *PF*: "The arrangement, hm. I guess I could, if we actually had this [gestures towards cards],...we have him write cards and see what the maximum number is that he can write. And then I'll take one card and write one bigger every time he

does that….So if he's smart he'll write one googolplex. And I'll write one googolplex plus one. And that would be bigger, and he can keep on going on and I can just respond until he stops."

SP-Untrained participants often expressed uncertainty about their ability to prove that the natural numbers extend infinitely. For example, when first asked how she would prove that there are infinitely many natural numbers, SP-Untrained Participant G (PG) expressed hesitance before eventually arriving at an explanation relying on the notion of squaring to produce a larger number:

PG: "That's a good question. Um, there probably is some kind of formula that can do that. Um, but proving just off the top of my head I don't think it's possible. There's definitely theorems or something that goes into that, I don't know how to do that, but what I would tell him would be, I guess one thing I could say as a counterargument would be like, What's the last number you could name, or the highest number you could name? And he doesn't have an answer and then I'll be like, it just goes on forever."

Int: "So lets say they do give you an answer, they say the highest number is a googol. How would you respond?"

PG: "I'd be like, what's googol squared?"

Unlike the participants above, a large number of SP-Untrained participants did not initially provide any mathematical justification for why the natural numbers are infinite, instead simply restating the claim. In some cases these participants provided a justification upon seeing the cards rearranged. For instance, SP-Untrained Participant H (PH) first restated that numbers are infinite, but eventually provided an operation-based explanation, noting that both addition and multiplication could be used to produce larger numbers:

PH: “I mean, it’s pretty simple, there’s an infinite amount of numbers, it goes on and on and on, and they can still be positive. Like, starting from 0 and all the way down the line to infinity basically, you’ll have integers which can go in this group according to the criteria.”

Int: “Let’s say you had one of those friend’s who’s like, But why? If your friend asked why they go on and on, is there something you could say about how you know that it goes on forever?”

PH: “I know from, I guess, all I can really say is from taking math courses, you learn that it’s a fact that numbers can go on infinitely. And that’s been known since like forever, so, I can’t, it’s not really disputable.”

[after cards are rearranged]

PH: “If it pertains to this specific group, so long as it’s - you could add any positive integer and it would still go in here. If you multiplied it by any positive integer, except for subtraction, subtraction would be the only exception....You can multiply any number in this, in these set of numbers right here [gestures towards cards on the table] and it would still be positive, and whatever the answer is could fit into this group.”

Many SP-Untrained participants expressed uncertainty about how to prove that there are infinitely many natural numbers. For instance, SP-Untrained Participant J (PJ) was not immediately sure how to respond to a claim that googol is the largest number; her uncertainty remained even after providing a notation-based explanation:

PJ: “Hm, I’ve never heard of googol. I would argue that there’s still more numbers after that, because - how would I do that? [pause] I’m thinking of how I would prove that, because proving is a different thing. I don’t have one single object that is one with ten zeroes after it, I don’t have anything, so I would say it’s hard in so far as bringing actual objects into it. Even printing out cards would take forever. So I think if this friend is just not believing me, I would just say why do they think it ends at that point? Why don’t they think there is a 2 that precedes that?”

Int: “Okay, so just so I know what you mean, you mean a 2 with a hundred zeroes?”

PJ: “Uh huh, yeah. So yeah it really depends on the scenario, I’m trying to think how to prove that there are infinitely many possibilities. But if someone’s really adamant about that I don’t know how I would prove it. But I guess that’s also something that, you know, you learn in class and you just believe that there are

> infinitely many numbers, which I feel is solid, like, truth, but I can see where maybe someone wouldn't believe that."
>
> *Int:* "Can you think about a rule about numbers that might be convincing? Like a rule about how these numbers work?"
>
> *PJ:* "Wow you're putting me on the spot like for math. I should know this. I'm thinking about different math concepts, and I can't think of anything necessarily, I'm sure there is something out there, but I can't actually think of anything right now."

These examples provide a sense of the wide variability in justifications provided by SP-Untrained participants, as well as the uncertainty that many expressed about what such a justification could look like. It is clear that, for at least some undergraduates without formal training in the axioms of natural number, the infinite nature of this set is accepted as a mathematical fact, but one for which they cannot immediately provide a clear and straightforward explanation.

Discussion

While, expectedly, virtually all undergraduates recognized that the natural numbers are an infinite system, it was only students who had received explicit college-level instruction on the formal axioms of natural number who consistently relied on the successor principle to justify this fact. SP-Untrained students - who, importantly, were highly educated adults at an elite university - generally did not refer to the successor principle in their justifications. Instead, SP-Untrained students offered a wider variety of justifications, including other arithmetic operations and notation-based strategies. Importantly, these other strategies do provide valid justifications that there are infinitely many natural numbers; for instance, one could continue to add 0's to the end of numerals to create an infinite list of larger and larger natural numbers (in this case, powers of 10). Thus, the majority of SP-Untrained students did provide mathematically valid justifications that there is no largest natural number. What these other justification strategies do not do, however, is account for the entire infinite system of natural numbers, which is uniquely generated by the successor principle.

Our results indicate that spontaneously-arising conceptualizations of the natural number system – those that arise without explicit mathematical instruction – may not rely uniquely on an understanding of the successor principle as is currently assumed in most developmental accounts. Scientists interested in concepts of number should devote more attention to empirically assessing "mature" conceptualizations of natural number in adults, in order to more accurately characterize the "endpoint" (or "endpoints") of developmental trajectories and assess the impact of formal mathematical training. Researchers should acknowledge the wider variety of conceptualizations of number that may exist, in both children and adults, and ask not only *when* understanding of the successor principle emerges but also *in what contexts* this conceptualization is recruited.

The present study has limitations which should be addressed in future work. For instance, the format of the task, which required participants to sort cards containing printed numerals, may have led SP-Untrained participants to focus on numerals and thus be more likely to give notation-based explanations (rather than operative ones). Mature concepts of number are likely to be highly flexible and context-dependent; additional work is needed to explore the contexts in which untrained adults do rely on the successor principle to explain features of the natural number system. If such conceptualizations are found to be elicited only in highly constrained contexts (for instance, when interview protocol explicitly mention the successor principle), then theories of number concept development should be updated to account for the wider variety of conceptualizations used in real-time, spontaneous reasoning.

Finally, these results have important implications for mathematics education. Mathematics instructors have to make assumptions about their students' prior knowledge in order to design effective instruction, and work such as ours can help instructors to better calibrate these assumptions. Existing work has revealed differences between everyday concepts and formal mathematical characterizations in high-level domains (e.g., continuity in calculus Marghetis & Núñez, 2013; transfinite numbers Núñez, 2005); however, relatively little work has explored the extent to which conceptualizations of even very simple mathematical concepts (like the natural numbers) might deviate from their formal characterizations. Our findings suggest that prior to explicit training undergraduates possess a variety of conceptualizations of natural number and may not immediately recognize that the set of natural numbers is uniquely generated by the successor principle. This might be one reason why mathematical induction – a formal proof method that relies on the successor principle – is a notoriously difficult method for students to learn (Avital & Libeskind, 1978; Movshovitz-Hadar, 1993; Stylianides et al., 2007). By being aware of the variety of conceptualizations that undergraduates may possess for even the simplest mathematical ideas, mathematics educators can create learning experiences and design curricula that more effectively bridge between their students' existing conceptualizations and formal mathematical knowledge. Finally, mathematics teacher education at all levels should incorporate thoughtful discussions regarding assumptions being made about students' entering mathematical knowledge, and the differences between everyday conceptualizations and mathematical formalisms.

Conclusion

We used a novel method to characterize undergraduates' understanding of the relationship between the successor principle and the set of natural numbers as an infinite system. We found that only undergraduates who have received explicit instruction in the successor principle at the college level were likely to use it to justify the fact that the natural numbers are infinite. Undergraduates without explicit training offered a wider variety of justifications, suggesting that spontaneously-arising conceptualizations of natural number may not primarily rely on understanding of the successor principle. These results cast doubt on the assumption, widespread in developmental psychology, that understanding of the successor principle is the central component and defining feature of "mature" conceptualizations of natural number. Further empirical work is needed to more fully characterize the variety of mature conceptualizations of natural number, and the contexts in which these different conceptualizations are recruited.

A portion of Chapter 3 is a reprint of selected material as it appears in *Naturalizing Logico-Mathematical Knowledge* 2018. Relaford-Doyle, Josephine; Núñez, Rafael, Routledge, 2018. The dissertation author was the primary author of this chapter.

References

Avital, S., & Libeskind, S. (1978). Mathematical induction in the classroom: Didactical and mathematical issues. *Educational Studies in Mathematics*, 429-438.

Carey, S. (2000). The origin of concepts. *Journal of Cognition and Development, 1*(1), 37-41.

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive Psychology, 92*, 22-36.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*(1), 162-173.

Hartnett, P., & Gelman, R. (1998). Early understandings of numbers: Paths or barriers to the construction of new understandings?. *Learning and Instruction, 8*(4), 341-374.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*(4), 498-525.

Marghetis, T., & Núñez, R. (2013). The motion behind the symbols: A vital role for dynamism in the conceptualization of limits and continuity in expert mathematics. *Topics in Cognitive Science, 5*(2), 299-316.

Movshovitz-Hadar, N. (1993). The false coin problem, mathematical induction and knowledge fragility. *Journal of Mathematical Behavior*, *12*(3), 253-268.

Núñez, R. E. (2005). Creating mathematical infinities: Metaphor, blending, and the beauty of transfinite cardinals. *Journal of Pragmatics, 37*(10), 1717-1741.

Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends in Cognitive Sciences, 21*(6), 409-424.

Relaford-Doyle, J., & Núñez, R. (2018). Beyond Peano: Looking into the unnaturalness of natural numbers. In Bangu, S. (Ed.) *Naturalizing Logico-Mathematical Knowledge* (pp. 234-251). Routledge.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences, 31*(6), 623-642.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*(3), 662-674.

Stylianides, G. J., Stylianides, A. J., & Philippou, G. N. (2007). Preservice teachers' knowledge of proof by mathematical induction. *Journal of Mathematics Teacher Education, 10*(3), 145-166.

CHAPTER 4

Formal Consistency of Undergraduates Number-to-Line Mappings for Large-Magnitude Scales

Introduction

In the three previous chapters we have used novel methods to explore contexts in which adults' conceptualizations of natural number appear to deviate from the formal characterization given in the Dedekind-Peano axioms, and specifically in which adults' reasoning in the domain of natural numbers appears not to be aligned with the logic of the successor principle. In this final study, we use a well-established paradigm: the bounded number line task. A large body of work has used bounded number line tasks to track the development of number concepts through childhood. A prominent theory emerging from this research program argues that "intuitive" number representations are logarithmic (i.e., the perceived difference between successive numbers becomes "compressed" as magnitude increases), and shift to "formally-appropriate" linear representations over the course of development and with schooling (Dehaene et al., 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003; Siegler et al., 2009). The eventual linearity of the number line mappings produced by older children and adults is taken as evidence for formally-consistent natural number concepts; however, this body of work has left some key questions unexamined. As in much of the developmental work examining number concepts, number line tasks frequently test performance of scales using relatively small, familiar magnitudes. Additionally, in focusing only on linearity and not the specific features of the linear relationships, researchers have spotlit positive evidence for formally-consistent representations and ignored aspects of the data which actually indicate non-normative number-to-line mappings. In this study we seek to address these shortcomings, and ask: In what ways do educated adults' number-to-line mappings *deviate* from formally-appropriate patterns? What is the impact, if any, of mathematical expertise and formal training in the structure of the natural number system?

Background

Various researchers have used number line tasks to investigate number concepts in children and adults (e.g., Dehaene et al., 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003), including the extent to which number concepts are consistent with formal characterizations of natural number. While there are multiple variations of this task in use, one of the most common versions is the bounded number line task, in which participants are given a number and asked to locate its position along a line segment with specified endpoints (e.g., place 37 on a line with endpoints 0 and 100). An underlying assumption of this work is that number-to-space mappings, and in particular mapping numbers to a line segment, provides a relatively straightforward and "pure" glimpse into the participant's mental representation of number (Siegler & Booth, 2004, p. 428).

Siegler and colleagues (Laski & Siegler, 2007; Siegler & Opfer, 2003; Siegler & Booth, 2004) have examined performance in number line tasks from early childhood to adulthood, and reported a shift from logarithmic to linear patterns of reporting. Drawing from Piagetian developmental theory (Inhelder & Piaget, 1958), Siegler situates these findings in the context of parallel sequences of development, in which similar developmental sequences repeat themselves in different contexts and over different time periods. Siegler and Opfer (2003) presented 2nd-graders, 4th-graders, 6th-graders, and undergraduates with two bounded number line tasks, one using a 0-100 scale and one using a 0-1000 scale. They found that, for all age groups, responses on the 0-100 task were best described by a linear model. However, they saw a different pattern of results on the 0-1000 scale: 2nd graders' estimates were logarithmic, while 6th-graders' and undergraduates' estimates were best described by a linear model (4th graders' responses were equally well-fit by the two models). These results led Sieger & Booth (2004) to ask whether they could find a similar developmental trajectory for younger children on the 0-100 task. They tested kindergarteners, 1st, and 2nd-graders, hypothesizing that 2nd graders, having been exposed to two-digit arithmetic in their school curriculum, would be more likely to produce linear estimates. Their results were consistent with this hypothesis: median estimates of kindergarteners and 1st graders followed a logarithmic pattern, while 2nd graders' median estimates were linear; individual results followed the same pattern, with 2nd graders being significantly more likely than younger children to produce linear mappings. The authors also found a significant correlation between math achievement scores and estimation accuracy for all age groups, as well as math achievement and linearity of responses.

From these findings, Siegler and colleagues (Siegler & Opfer, 2003; Siegler & Booth 2004; Laski & Siegler, 2007) argue that children and adults have multiple representations of number which are recruited differently based on age and task context. Initial "intuitive" number representations, available from infancy, are logarithmic; with schooling and experience with the formal number system, including exposure to arithmetic, children come to develop "formally appropriate" number representations and produce linear number-to-space mappings (Siegler & Opfer, 2003, 236). This shift does not happen all-at-once; the same child may recruit a linear representation while mapping numbers on a 0-100 line segment, and a logarithmic representation when given a 0-1000 scale (as was the case for Siegler and Opfer's 2nd-graders). Adults reliably produce linear number-to-space mappings (at least up to a 0-1000 scale), but continue to recruit logarithmic and non-linear mappings in other contexts (Banks & Coleman, 1981; Banks & Hill, 1974), suggesting that even in adulthood we continue to possess multiple number representations which are deployed based on task conditions. From this Siegler concludes "with age and experience, children may rely increasingly on the most appropriate representation for the situation" (Siegler & Booth, 2004, 430) - on number line tasks, this looks like a shift from logarithmic to linear responses. This account - that initial intuitions of number are logarithmic, and with schooling shift to linear - has formed the basis of (and seemingly been corroborated by) influential cross-cultural research which reports logarithmic number-to-space mappings in unschooled indigene adults (Dehaene et al., 2008; though see Núñez 2008 for a critique).

While the studies described above provide compelling evidence that children's number-to-line estimations undergo a shift from logarithmic to linear response patterns, there are some shortcomings in the research. First, Siegler's theories are based on evidence using, at the largest, a 0-1000 scale, leaving open the question of how adults perform on higher-magnitude number line tasks. Based on the framework of parallel developmental sequences, Siegler & Booth (2004) speculate that on scales of much larger magnitude, and particularly under time constraints precluding calculation of percentages, adults would produce logarithmic patterns of responses. This raises the general question: is there a point at which adults' general concept of number - of the *natural number system* itself - becomes wholly "formally appropriate"? Or will there always be magnitudes for which even schooled adults produce non-linear, formally-inconsistent mappings? Additional data is needed to shed light on these questions.

Second, as is common in studies using number line mappings, Siegler and colleagues have exclusively tested children's and adults' performance on scales in which the large-magnitude endpoint is a power of 10 (in particular, 0-100 and 0-1000). Evidence from a variety of studies suggests that participants of all ages rely on proportional reasoning on bounded number line tasks (Cohen et al., 2018; Slusser et al., 2013; Sullivan et al., 2011), and that analogical reasoning between scales of different orders of magnitude is an effective strategy for producing relatively accurate responses on higher-magnitude scales (Thompson & Opfer, 2010). Thus, linear responses on large-magnitude scales may reflect older children's and adults' increased skill with proportional reasoning and use of more sophisticated, task-specific techniques, rather than an underlying conceptualization of numbers within that range. Siegler's work leaves unaddressed the question of how patterns of responses might look when proportional and analogical reasoning becomes less accessible.

Finally, in Siegler and colleagues' work and within the larger body of research using number line tasks, linearity is taken as an indicator of "formally-appropriate" number representations, regardless of the specific features of the linear relationship. The formal characterization of natural numbers doesn't just specify a linear relationship; the Dedekind-Peano axioms of natural number stipulate that the difference between successive numbers is exactly 1 (and it is this - the successor principle - which developmental psychologists claim underlies mature concepts of natural number). Thus, a truly "formally-appropriate" response pattern would be described graphically by the line y=x, in which the slope of the linear relationship is exactly 1. While such clean response patterns are unlikely to be elicited from human participants, significant deviations from this model are worthy of consideration in any question of the formal consistency of natural number concepts. Siegler & Booth (2004) acknowledge that the "ideal" slope of a linear response pattern is 1, and report that slopes move closer to 1 over the course of development. However, the oldest children in this study produced estimations that were best fit by a line with a slope of 0.6; the authors do not explore this further, or offer an explanation of how this deviation impacts their claims of a shift to "formally-appropriate" representations. Notably, many studies using bounded number line tasks don't even report the slopes of lines-of-best-fit, instead reporting only the total variance captured by the linear model (e.g., Bull et al., 2012; Siegler & Opfer, 2003; Sullivan et al., 2011; Thompson & Opfer, 2010). Overall, the lack of attention to the features of the linear relationship suggests that, in this body of research, it is linearity itself - in any form - which is taken as evidence of formally-consistent number concepts. This misses key aspects of the formal characterization of natural number, specifically that the successor principle dictates a slope of 1, which should be considered in any such claims.

In the present study, we seek to provide a more thorough exploration of adults' number-to-line estimations by addressing the gaps outlined above. We also consider the role of mathematical expertise, and examine how explicit training in formal mathematical induction and the axioms of the natural number system impacts number-to-space mappings in a large-magnitude context.

Method

This study used methods based on Laski & Siegler's (2007) number line estimation task, but extended their design by adding additional high-magnitude conditions and testing adult participants with different mathematical backgrounds.

Materials: The study was programmed in MatLab2013b, and presented to participants on a 12"x18" flatscreen monitor in a quiet room. Participants used a wireless mouse to click a location along a line segment to indicate the position of a given number. Throughout the study, numerals and number lines were displayed in black against a white background. On every trial, the number line extended horizontally and was 13” long, 1mm thick, centered on the screen. On both ends of the line there were 5mm vertical tick marks, and the endpoints were labeled with numerals centered above the tick marks. The numerals which were presented as stimuli had a height of 0.5” and were presented approximately 2.5” above the number line, centered horizontally.

Participants: All participants were undergraduate students at a major research university, and were tested individually. Participants were recruited from two distinct populations. The first group (n=90; Male = 34, Female = 54, Nonbinary/Genderfluid = 2) was recruited through a general subject pool that consisted of students with a range of majors including psychology, cognitive science, and linguistics. Many of these participants had taken university-level mathematics courses, including calculus and statistics. However, none had taken a university-level mathematics course covering formal mathematical induction, and so we refer to this group as MI-Untrained. Participants from this group were randomly divided between conditions A, B, and C of the task (n=30 in each condition), which are described below. The second group of participants (n=15) consisted of students (mostly Mathematics majors) from the same university who had taken Mathematical Reasoning, an upper-division mathematics course focusing on proof techniques including formal mathematical induction. We refer to these participants as MI-Trained. Due to recruitment limitations and our interest in specific comparisons, all 15 of these participants completed Condition C of the task (described below).

Procedure: All participants completed a timed number line estimation task. In this task, a number was displayed in the top center of the screen, and a number line with marked endpoints was positioned below. Upon seeing the number, participants had two seconds to indicate via mouse-click the position of that number on the line segment; we included this time-limit in order to preclude participants from calculating proportions or doing other arithmetic transformations. To advance to the next trial, participants needed to click a box on the lower-middle portion of the screen (thereby resetting their mouse position). This portion of the task was untimed, allowing participants to take breaks between trials as needed. To orient participants to the timed nature of the task, participants completed four training trials at the beginning of the study, in which they were asked to indicate by mouse-click the location of numbers corresponding to the endpoints of the line segment. During the experimental portion of the study, 22 numbers were displayed in random order 3 times each, resulting in 66 total trials. Following the number line estimation task, participants completed a computer-based survey in which they reported their gender, handedness, major, list of mathematics courses taken at the university, and a brief description of any strategies they used during the task.

Conditions: Each condition employed a different number scale, as described below:

Condition A: Low-magnitude. This condition replicates Laski & Siegler's (2007) task, which used a 0-100 scale and included the following 22 numbers, which were chosen to maximize discriminability between linear and logarithmic response patterns: 2, 3, 5, 8, 12, 17, 21, 26, 34, 39, 42, 46, 54, 58, 61, 67, 73, 78, 82, 89, 92, 97. Thirty MI-Untrained participants were given the Condition A task.

Condition B: High-magnitude. This condition used a 0-100,000 scale; to generate target numbers, each number in Laski & Siegler's list was multiplied by 1000 (e.g., 2000, 3000, …, 97000). Thirty MI-Untrained Participants were given the Condition B Task.

Condition C: High magnitude, difficult proportions. Various studies have shown that both children and adults often rely on proportional and analogical reasoning on bounded number line tasks (Sullivan et al., 2011; Thompson & Opfer, 2010). To examine response patterns when proportional reasoning is less accessible, in this condition all values were multiplied by 1183, resulting in a 0-118,300 scale (and values 2366, 3549, …, 114751). Thirty MI-Untrained and 15 MI-Trained participants were given the Condition C task.

Analysis: Responses on each trial were automatically recorded. Trials in which participants look longer than 2 seconds to respond were excluded from analysis (291/6930 trials, or 4.2%). Prior to analysis we removed trials in which the participants' response was greater than 3 standard deviations from the mean of all participants' responses for that target number within that condition (127/6639, or 1.9%). Finally, as described above, each of the 22 target values were displayed three times to each participant; for the analysis, we averaged the responses provided on each of the three trials, producing a single mean response value per target for each participant.

Results

Figure 4.1 displays participants' mean responses in each condition. To determine whether response patterns were better fit by linear or logarithmic models, linear and logarithmic least-squares regressions were run separately for each condition. Comparisons of R^2 coefficients of determination indicate that for all four conditions, patterns of responses are, unsurprisingly, better fit by a linear model (see Figure 4.1 and Table 4.1).

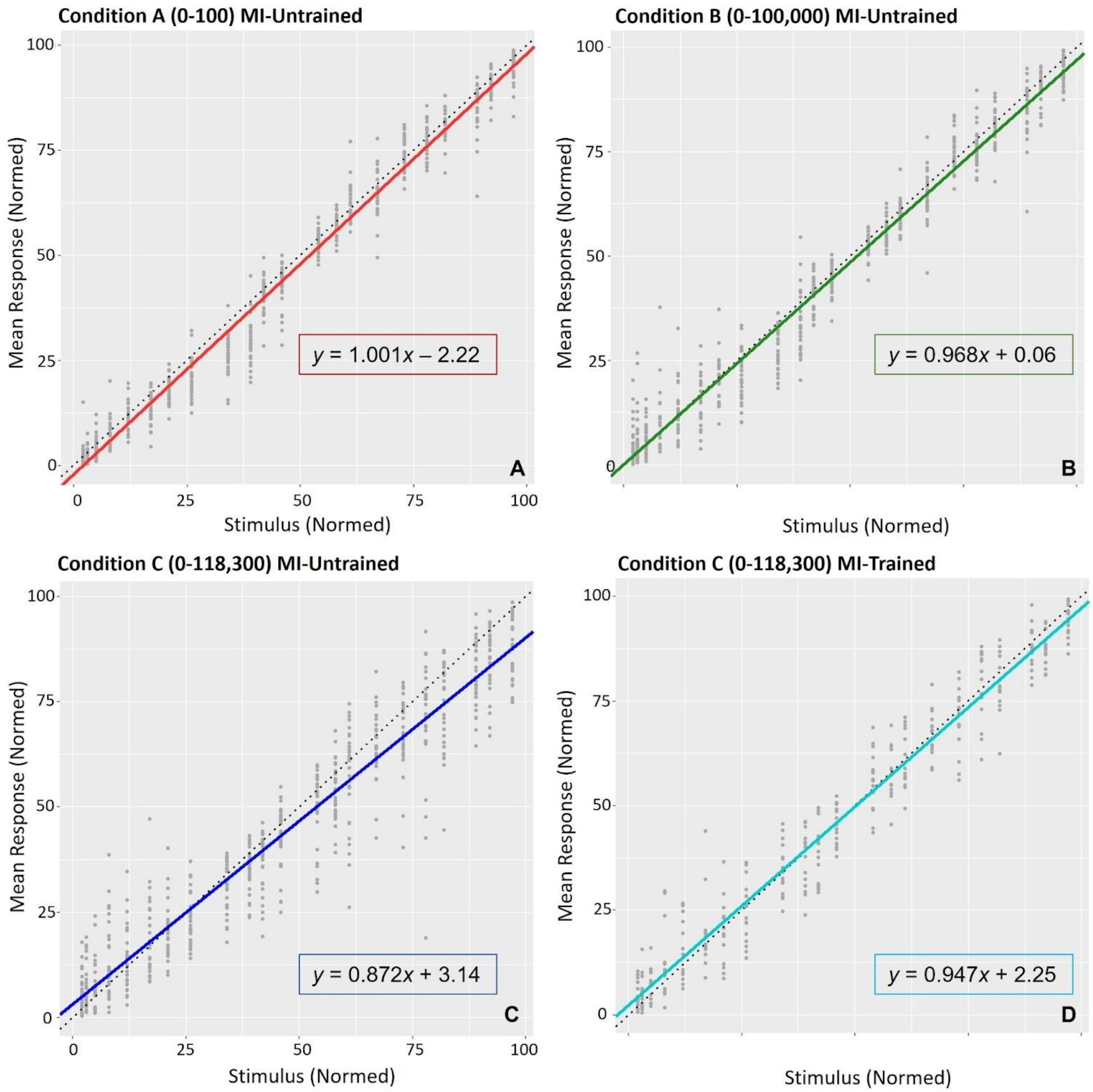


Figure 4.1: Mean participant responses and lines-of-best-fit by Condition. Each gray dot represents a single participant's mean response for the given target value. The colored line represents the least-squares regression line for all participant data within that Condition. The line y=x, representing accurate number-to-space mappings, is shown as a dotted line for reference. For ease of visualization, all responses have been normed to a 0-100 scale.

Table 4.1: Fit of linear and logarithmic models for all conditions.

Participants	Condition	Linear Model	R^2	Log Model	R^2
MI-Untrained	A(0-100)	y = 1.001x - 2.22	0.9775	y = 24.28ln(x) - 39.07	0.7733
	B (0-100,000)	y = 0.968x + 0.06	0.966	y = 23.38ln(x) - 35.2	0.7562
	C (0-118,300)	y = 0.872x + 3.14	0.9147	y = 21.24ln(x) - 29.27	0.7317
MI-Trained	C (0-118,300)	y = 0.947x + 2.25	0.9523	y = 23.23ln(x) - 33.5	0.7704

To investigate whether this result held for individuals, regressions were run separately for each participant. Again, unsurprisingly, all 105 participants produced patterns of responses which were better fit by the linear model than the logarithmic model (as determined by comparison of R^2). These results are consistent with the existing literature showing that counting adults produce linear number-to-space mappings on bounded number line tasks (Siegler & Opfer, 2003).

To more thoroughly investigate the linear response patterns produced by participants, and specifically the extent to which patterns were consistent with a formal successor principle, we next determined lines of best fit for each individual participant and compared the slope coefficients in each condition. A one-way analysis of variance test revealed significant differences between the slopes produced by participants by condition, $F(3, 101) = 7.84$, $p < 0.001$. A post-hoc Tukey test showed that the slopes produced by MI-Untrained participants in Condition C (mean = 0.871) differed significantly than slopes produced in both Conditions A (mean = 1.001) and B (mean = 0.968) at the $p < 0.01$ level. No significant differences were found between slopes in Conditions A, B, and C with MI-Trained participants, or in Conditions C between MI-Trained and MI-Untrained participants (although this approached marginal significance; see Figure 4.2).

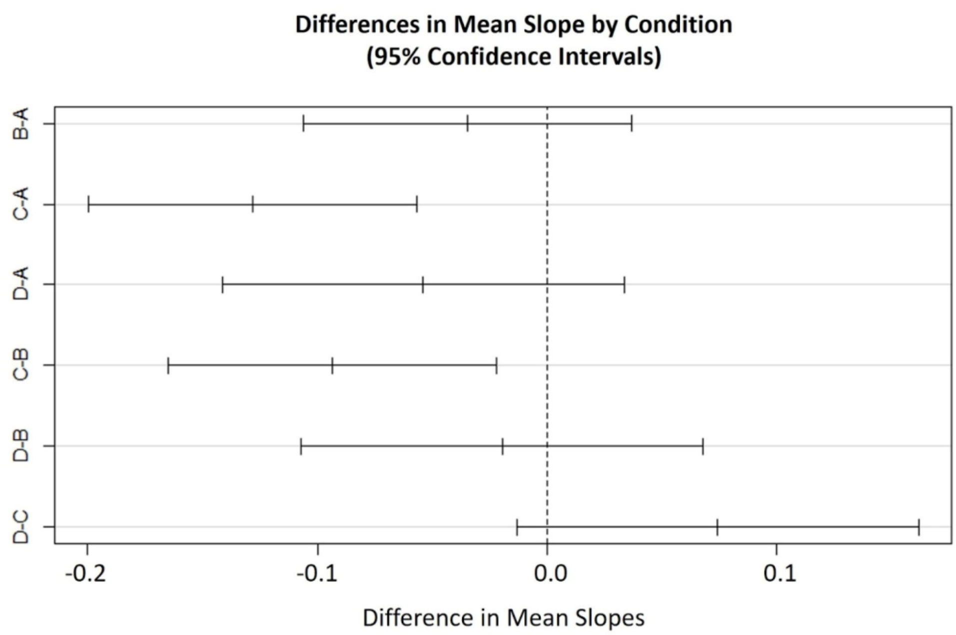


Figure 4.2: Summary of Tukey post-hoc tests for differences in slopes by condition. Bars show 95% confidence intervals of differences between mean slope values for each comparison. Note: On the y-axis, C refers to Condition C (MI-Untrained), while D refers to Condition C (MI-Trained).

Figure 4.3 shows frequency distributions of slopes by condition. For Conditions A, B, and C (MI-Trained), slopes were distributed approximately normally around the mean slope values of 1.00, 0.968, and 0.947, respectively (Figure 4.3A,B,D). For Condition C with MI-Untrained participants the distribution of slopes was bimodal and contains two clusters, one centered around 1.0 and another centered around 0.72 (Figure 4.3C). To determine whether the two clusters represented participants with differing levels of mathematics expertise, we conducted a median split of participants by number of university-level math courses taken, resulting in a low-math-exposure group and a high-math-exposure group. There was no relationship between math exposure and slope of regression line (Fisher's Exact Test, $p = 1$).

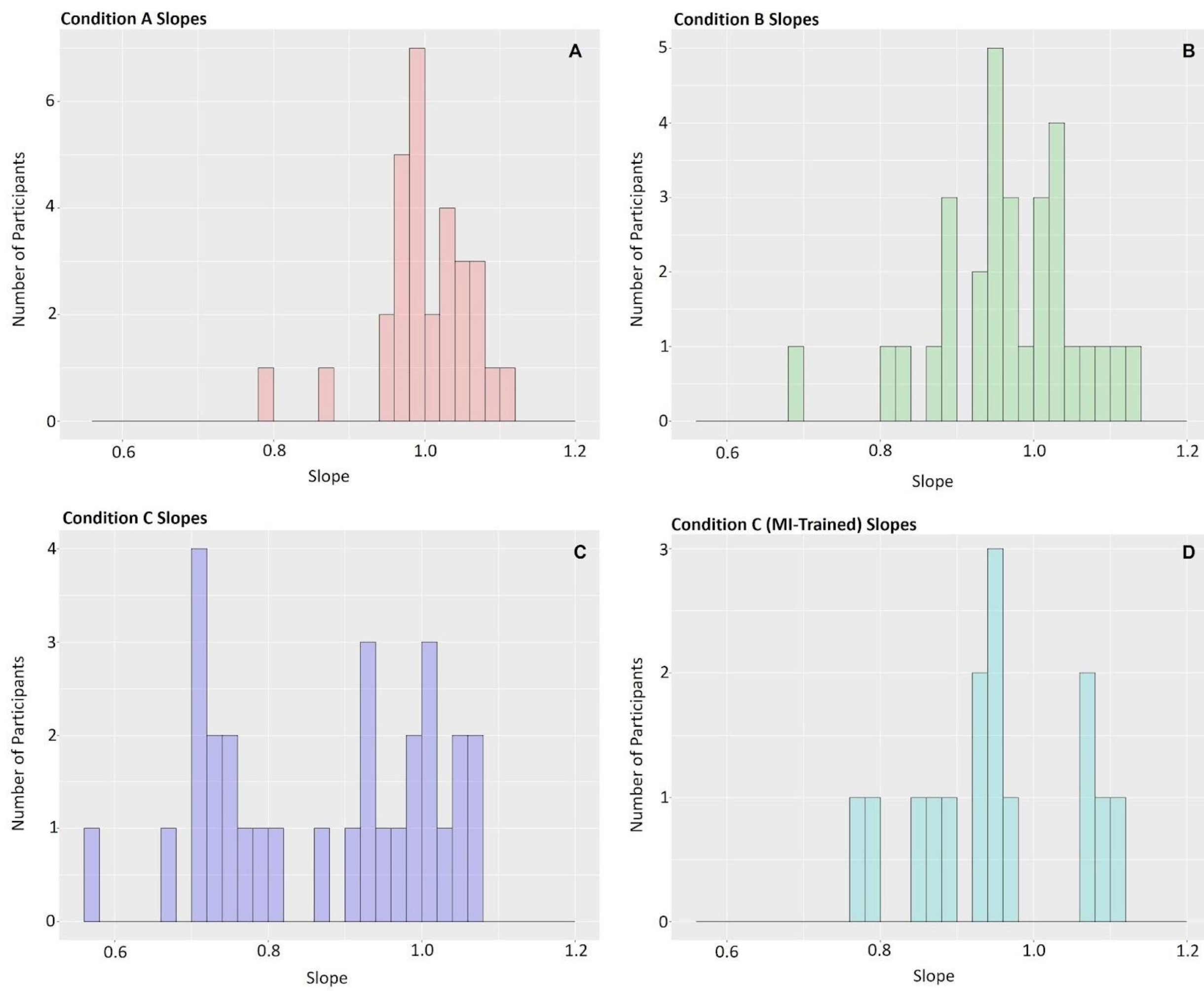


Figure 4.3: Frequency distributions of slopes of individual participants' least-squares regression lines. (A) Slopes in Condition A, mean = 1.001, sd = 0.064. (B) Slopes in Condition B, mean = 0.968, sd = 0.095. (C) Slopes in Condition C, MI-Untrained participants, mean = 0.872, sd = 0.144. (D) Slopes in Condition C, MI-Trained Participants, mean = 0.947, sd = 0.106.

Finally, to examine patterns which may underlie the observed differences in slope, we examined patterns of estimation for particular spans along the line. Figure 4.4A shows the mean responses averaged over all participants for each target value. Visual inspection suggests that participants in larger-magnitude conditions may have tended to overestimate the location of target numbers at the lower end of the scale (Figure 4.4B), specifically stimuli that when scaled correspond to the prototypical counting numbers (PCNs) between 1 and 9. Additionally, participants seemed to underestimate the location of numbers at the upper end of the scale (Figure 4.4C); together, this pattern of over- and underestimation could have driven the reduction of slope in larger-magnitude conditions, and particularly for MI-Untrained Participants in Condition C.

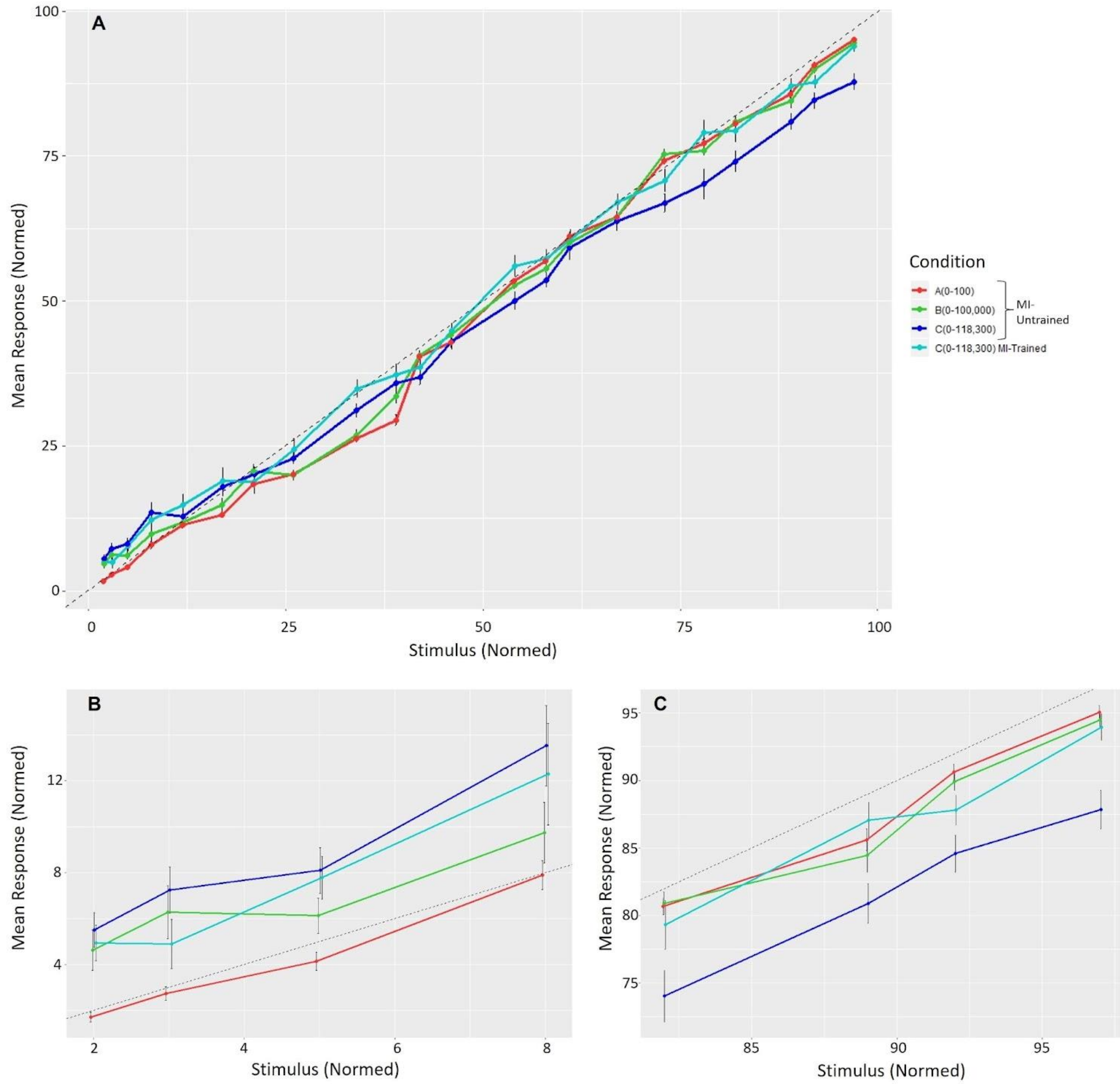


Figure 4.4: Mean responses by condition. Results over full span of range (A), and for smallest (B) and largest (C) stimulus values. For ease of visualization, all responses have been normed to a 0-100 scale. Error bars show standard error of the mean. The black dashed line shows y = x, representing perfectly accurate number-to-space mappings.

To explore these patterns of over- and underestimation, we calculated a "coefficient of deviation" for each participant and target value by subtracting [*mean response* - *target value*], both normed to a 0-100 scale (for instance, if a participant's mean response for target value 2000 was 3500, the corresponding coefficient of deviation would be 3.5 - 2 = 1.5, reflecting overestimation for that target value). We then calculated each participant's average coefficient of deviation within two ranges of interest: the PCNs (stimuli corresponding to 2, 3, 5, and 8), and numbers within PCN-range of the upper limit (stimuli corresponding to 92 and 97). Multiple t-tests were used to identify whether participants' mean deviation differed significantly from 0 (this resulted in a total of 8 t-tests reflecting the two regions-of-interest and four conditions; the Holm-Bonferroni procedure was used to correct levels of significance to account for the multiplicity of tests). At the lower end of the range, there was significant overestimation of target values corresponding to the PCNs in all three large-magnitude conditions: Condition B (M = 2.61, SD = 5.59, $t(29)$ = 2.56, p = 0.016); Condition C for MI-Untrained participants (M = 4.06, SD = 5.43 , $t(29)$ = 4.10, p = 0.0003); and Condition C for MI-Trained participants (M = 2.97, SD = 3.9, $t(14)$ = 2.95, p = 0.01). At the upper end of the range, there was significant underestimation in all four conditions: Condition A (M = -1.97, SD = 3.12, $t(29)$ = -3.46, p = 0.0017); Condition B (M = -2.31, SD = 2.98, $t(29)$ = -4.24, p = 0.0002); Condition C for MI-Untrained Participants (M = -8.13, SD = 6.49, $t(29)$ = -6.86, p < 0.0001); and Condition C for MI-Trained participants (M = -3.62, SD = 3.69, $t(29)$ = -3.81, p = 0.0019). Visual inspection of mean response patterns over the entire span of the line (Figure 4.4A) indicates that, while underestimation of large values may have been a more general pattern (e.g., occurring for MI-Untrained participants in Condition C in the entire upper quartile of the range), systematic overestimation of target values occurred only at the lowest end of the range, for stimuli corresponding to the PCNs.

Discussion

Are response patterns linear at larger magnitudes? First, we were interested in assessing whether educated adults produce linear number-to-space mappings for larger-magnitude and less-familiar scales. We found that in all conditions, patterns of responses were better fit a linear model than a logarithmic one; this was true for mean responses across all participants (Figure 4.1) and for each individual participant. This is consistent with Siegler and colleagues' theory that there is a shift towards linearity that occurs as individuals become more familiar with larger-magnitude scales. Our results here are unsurprising, as existing work has demonstrated linear response patterns in adult participants on bounded number-line tasks (Siegler & Opfer, 2003). However, this still leaves open the possibility that adults might produce logarithmic (or otherwise non-linear) patterns of responses for even larger-magnitude scales (see e.g. Landy et al., 2017, who found piecewise-linear responses within categories arranged by order-of-magnitude when testing up to 1 billion). In the present study, we tested only up to magnitudes on the order of 100,000, which may in fact be quite familiar to participants. Future work can continue to assess larger-magnitude scales in order to more thoroughly assess the limits of Siegler and colleagues "parallel development" of linear representations at larger magnitudes and over the course of development.

Differences in slopes between MI-Untrained Participants. Existing research has focused exclusively on linearity as the indicator of "formally-appropriate" number representations, and has failed to take into consideration the specific features of the linear pattern. In the formal set of natural numbers, the difference between any two successive numbers is exactly 1; thus, if natural number concepts are aligned with the formal characterization, we would expect to see, not just linearity, but a slope of exactly 1. In our data we saw significant differences in the slopes of the linear mappings produced in by MI-Untrained participants. Condition A presented a familiar and relatively low-magnitude 0-100 scale, and here we saw that our participants' linear mappings were indeed remarkably aligned with the formal characterization, with a mean slope of 1.001 with a relatively small standard deviation of 0.063. In Condition B, all values were scaled up by 1000, making proportional and/or analogical reasoning easily accessible, and here we saw no significant change in the mean slope that participants produced (though there was a slight increase in variability between subjects, Figure 4.1B). However, in a larger magnitude scale in which proportional and analogical reasoning were less accessible, we saw significant deviations from formal consistency; specifically, MI-Untrained participants in Condition C produced a mean slope of 0.872; we also saw an increase in variability within participants in this Condition (Figure 4.1C). These results provide evidence that adults are likely to produce formally-consistent linear mappings only on relatively small scales, or when the task allows participants to access proportional reasoning and/or analogical mapping to a smaller, more familiar scale.

By demonstrating key differences in the slopes produced at different magnitudes, these results add important nuance to Siegler and colleague's assertion that mental representations of number shift from logarithmic to "formally-appropriate" linear representations over the course of development. Siegler & Booth (2004) speculated that slopes approach 1 as children become more familiar with different number scales; our results suggest that this may only occur in the smallest, most familiar number scales (as in our Condition A, 0-100) or when the scale allows for straightforward proportional and analogical reasoning (Condition B). Additionally, our results are inconsistent with assumptions made in developmental literature, which argue that children come to understand the successor principle and that this knowledge is eventually applied to "all possible numbers" (Cheung et al., 2017). Our data show that, in the context of the number line task, educated adults produce number-space mappings that are consistent with the successor principle when the scale is small and familiar, but that the difference between successive numbers (as captured by the slope of the line-of-best-fit) deviates from 1 in higher-magnitude contexts, and particularly when proportional reasoning is inaccessible.

Role of mathematical expertise. We were interested in the role of mathematical expertise, and specifically whether familiarity with the formal characterization of the natural number system would be associated with a greater degree of "formally appropriate" number-to-line mappings. In our Condition C (0-118,300) we found no significant difference between the slopes produced by MI-Trained and MI-Untrained participants, nor between the mean slopes produced by MI-Trained participants in Condition C and untrained participants in Conditions A and B, where performance approached formal consistency. Thus, on the grounds of mean slopes, it is difficult to assess the role of mathematical expertise based on our data. However, it is worth noting that the distribution of slopes produced by MI-Trained participants in Condition C (Figure 4.3D) appears qualitatively different than the distribution of slopes produced by MI-Untrained participants in the same condition (Figure 4.3C). Specifically, MI-Untrained participants showed higher variability and had a large proportion of participants (13/30, 43%) produce slopes that were less than 0.85; MI-Trained participants showed less variability and only 2 of the 15 participants (13.3%) produced slopes less than 0.85 (a Fisher's Exact test reveals a marginally significant difference between these ratios, $p = 0.053$). Thus, while there was no difference in the mean slopes produced by MI-Trained and MI-Untrained participants in Condition C, inspection of individual responses reveals that MI-Trained participants responses in Condition C (0-118,300) more closely resembled MI-Untrained participants' performance in lower-magnitude, more familiar scales (Conditions A and B). In particular, MI-Trained participants in Condition C

(0-118,300) produced a distribution of slopes which most closely resembles those produced by MI-Untrained participants in Condition B(0-100,000).

The similarity in performance between our Condition C MI-Trained and Condition B MI-Untrained raises the question of whether our MI-Trained group's performance was related to their knowledge of the formal natural number system, or simply a reflection of general quantitative skills, including proportional reasoning ability. We investigated this question in our Condition C (MI-Untrained) group, using a median split based on number of math classes taken at the university level as an indication of general quantitative ability, and found no significant difference in the slopes generated by participants who had taken relatively more mathematics classes, suggesting that there may be a unique impact of training in mathematical induction. However, the number of math classes an undergraduate has taken at the university is certainly not a direct indicator of their quantitative skills and proportional reasoning. A limitation of the present study is that we did not include a test of general quantitative ability, including proportional reasoning. Future studies could include such a measure, in order to tease apart the contributions of general mathematical ability and number of the natural number system on math experts' number-to-line mappings.

Why are slopes less than 1 in higher-magnitude contexts? It is notable that in high-magnitude contexts, slopes didn't just deviate from 1, but became systematically less than 1. This was particularly clear in our Condition C (0-118,300, MI-Untrained) in which the reduction of mean slope to 0.872 was driven by a large number (13/30, 43%) of participants who produced patterns of responses with slopes less than or equal to 0.85. This raises the question - why does the slope become less than 1 in larger-magnitude, less-familiar contexts? An examination of mean responses for each stimulus value (Figure 4.4) indicates that this is the result of systematic overestimation of values at the lower end of the scale and underestimation of values at the higher end. This phenomenon has been observed in other studies; for instance, DiLollo & Kirkham (1969) saw similar patterns when asking participants to report the number of shaded squares in a black-and-white grid. Ross & DiLollo (1971) theorize that this pattern of results - overestimation at the low end, and underestimation at the high end, can be expected whenever the scale is bounded and unidimensional, and when participants are less familiar with the task context (see also Cohen & Sarnecka, 2014, for a discussion of this phenomenon in the context of number line tasks). On number line tasks, this phenomenon has been referred to as "edge-avoidance" (Landy et al., 2017). Landy and colleagues explain this effect as an instance of category bias. In their account, on a bounded number line task participants first map regions of the line to appropriate categories, which depend on the scale being presented. For instance, when using a 0-100 scale, participants might categorize the line into four regions representing the four quantiles 0-25, 25-50, 50-75, and 75-100; or, on a task involving multiple orders of magnitude, participants may categorize the line segment into regions corresponding to thousands, millions, etc. This categorization of the line into regions may be normative (for instance, if the line is divided into four equal regions to represent the four quartiles), or non-normative (for instance, placing a boundary of "1 million" approximately halfway between 1 thousand and 1 billion). When presented with a stimulus, participants first categorize it (i.e., select the corresponding region of the line), and then use magnitude information to place it relative to the endpoints of the subrange. Landy and co-authors suggest that edge-avoidance occurs because responses are shifted towards the "prototypical" value in the range (in their account, assumed to be the center of the subrange) in order to mitigate noisy magnitude information. By this account, we would expect to see more edge-avoidance on higher-magnitude and less familiar scales due to increased "noisiness" of the magnitude information, and therefore more bias towards the central prototype value. Indeed, this pattern is reflected between our MI-Untrained participants in Conditions A, B, and C.

While consistent with the data, Landy et al.'s explanation does not account for why a particular number within the range is considered "prototypical", and simply assumes that the central value is treated as the prototype. In this study, for instance, why would 59,015 be the "prototypical" number on a 0-118,300 scale? In fact, when looking at the patterns of participant responses in Condition C (Figure 4.4A), it would appear that a shift from overestimation to underestimation occurred quite a bit lower on the scale, below the normed stimulus value of 20, with targets near the endpoints seeing larger bias effects. Based on these results, we propose a separate account of edge-avoidance on large-magnitude scales. Rather than considering categorical regions of the line segment to be entirely context-dependent, as in Landy et al.'s (2017) account, we claim that there is a static and persistent category of numbers that impacts reasoning on number line tasks, namely, the category of prototypical counting numbers (PCNs). As described in earlier chapters, the PCNs are the smallest, most familiar numbers, which are the first we encounter as children and the ones most basic to daily experience. The PCNs contain at least the single-digit numbers 1-9 (and may extend higher for some people). Previous studies reported in this dissertation have demonstrated that in certain tasks, the PCNs may be treated as qualitatively distinct from other numbers; for instance, an individual may have a sense that the number 3 is of a different type than the number 100,000,000 (even though these are both formally classified as natural numbers). On the number line task, we propose that individuals may associate the left-most region of the line with the prototypical counting numbers, regardless of the scale being used. When the scale is low-magnitude (e.g., 0-100), the PCNs are accurately mapped to this region of the line segment (see Figure 4.4B, red line corresponding to Condition

A). On larger-magnitude scales, however, participants may avoid placing larger-magnitude numbers in the region associated with PCNs, even when such placement would be normative given the scale. This avoidance results in a shift to the right, or overestimation of the stimulus value (Figure 4.4B). This is consistent with our data, in which we saw significant overestimation of stimulus values corresponding to PCNs in all three larger-magnitude conditions. Underestimation at the higher end of the scale occurs by a symmetric process, as participants avoid the region at the right-most end of the segment which is associated with numbers that are within PCN range of the upper endpoint (this is consistent with multiple studies that have found that participants rely on a subtractive process to locate numbers relative to the higher endpoint, e.g., Cohen & Sarnecka, 2014; Cohen et al., 2018). Values that are deemed much smaller than the right-most endpoint are thus shifted to the left, resulting in underestimation at the highest end of the scale (a pattern we observed in our data, in all four Conditions). As above, this underestimation is most pronounced on higher-magnitude, less familiar scales (this is reflected in our Condition C data from MI-Untrained participants, Figure 4.4C).

Our account differs from Landy and colleagues' in that it assumes a stable, persistent category arising from everyday experience - the Prototypical Counting Numbers - which is automatically mapped to a region of the line segment, regardless of the scale. However, this account leaves aspects of our data unexplained: for instance, why do participants in Condition C significantly underestimate values well below the upper endpoint of the scale, and not only those that, when normed, are within PCN-range of the upper endpoint? And why do participants in Condition A (0-100) show accurate mappings of the PCNs, but underestimate numbers at the high end of the scale? These results are not explained by our account, but may be explained by Landy et al.'s category bias account. Thus, we believe that the edge-avoidance we see in our data is likely the result of both processes outlined above: category bias based on context-dependent categories (Landy et al., 2014), and categorical reasoning based on a persistent, pre-existing PCN category. Future studies could attempt to identify the specific contributions of each of these processes to performance on number line tasks, particularly those which use large-magnitude scales.

Do these results actually bear on number concepts? The logarithmic-to-linear shift in mental representation of number has been widely studied using bounded number line tasks (Dehaene et al., 2008; Siegler et al., 2009). This research program rests on the assumption that performance on such number line tasks provides an accurate "readout" of the individual's internal representation of number, such that a shift from logarithmic to linear mappings over the course of development suggests a comparable shift in mental representation of number. This assumption is not without controversy, and a separate body of work pushes against this "Assumption of Pure Numerical Estimation" (Cohen et al., 2018). Researchers in this camp argue that the bounded number line task requires a set of specific skills, including subtractive and proportional reasoning, and that performance on such tasks is a reflection not only of number concepts but also the participant's mensuration skills (their ability to scale numbers to lines). In line with this, Cohen & Sarnecka (2014) found that, while children's performance on bounded number line task becomes more linear over time, there is no such change in performance on an unbounded number line task (which requires only additive reasoning). Based on this, the authors argue that early logarithmic performance is actually the result of low mensuration skills, and encourage the use of unbounded number line tasks (in which only the smaller endpoint is given) are a more appropriate and transparent tool for studying the development of number concepts (Cohen & Sarnecka, 2014; Cohen et al., 2018).

In the present study, our goal was to engage with an existing and influential body of work, which has used bounded number line tasks to study internal representations of number. Within this framework, we have demonstrated important nuances to the logarithmic-to-linear shift theory, which we believe are important for researchers in this camp to consider when making any claims about the formal consistency of natural number concepts. However, we ourselves are not tied to the assumption that our task provides a direct window into our participants' internal representation of number; in fact, the data from our MI-Trained group suggests that proportional reasoning and mensuration skills likely play an important role in undergraduates' performance on bounded number line tasks. Future work could use unbounded number line tasks with high-magnitude scales, and participants with varying degrees of mathematical expertise, to determine whether the results we found are observed only in bounded tasks, or whether they carry over into other contexts in which mensuration demands are lower.

Conclusion

In this study we employed a well-known paradigm - the bounded number line task - to explore adults' number-to-line mappings in high-magnitude contexts. Undergraduate participants produced linear number-to-line mappings in all conditions, even in high-magnitude contexts when proportional reasoning was inaccessible; this is consistent with Siegler and colleagues' log-to-linear shift theory of mental number representation (Siegler et al., 2009). However, closer examination of the linear relationships revealed differences in the formal consistency of the mappings. Specifically, on larger-magnitude scales the slopes of the linear mappings deviated from 1, the normative slope which would be predicted by the axioms of natural number and the successor principle. This effect was greatest when proportional reasoning was inaccessible and when participants did not have formal training in the natural number system. We suggest that the reduction in slope on high-magnitude scales may be due in part to categorical reasoning about the prototypical counting numbers, which may be associated with the left-most side of the line segment, resulting in overestimation and underestimation at the extremes when the scale is increased. By revealing significant differences in the normativity of linear response patterns, we demonstrate that linearity itself is not a sufficient indicator of a "formally-appropriate" mental representation of number. Future research within this paradigm should take care to fully investigate and report features of the linear patterns produced by participants, in order to more accurately assess the formal consistency of number-to-line mappings.

References

Banks, W. P., & Coleman, M. J. (1981). Two subjective scales of number. *Perception & Psychophysics, 29*(2), 95-105.

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology, 102*(2), 353.

Bull, R., Marschark, M., Sapere, P., Davidson, W. A., Murphy, D., & Nordmann, E. (2011). Numerical estimation in deaf and hearing adults. *Learning and Individual Differences, 21*(4), 453-457.

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive Psychology, 92*, 22-36.

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology, 50*(6), 1640.

Cohen, D. J., Blanc-Goldhammer, D., & Quinlan, P. T. (2018). A mathematical model of how people solve most variants of the number-line task. *Cognitive Science, 42*(8), 2621-2647.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science, 320*(5880), 1217-1220.

Di Lollo, V., & Kirkham, R. (1969). Judgmental contrast effects in relation to range of stimulus values. *Journal of Experimental Psychology, 81*(3), 421.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). Psychology Press.

Landy, D., Charlesworth, A., & Ottmar, E. (2017). Categories of large numbers in line estimation. *Cognitive Science, 41*(2), 326-353.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78*(6), 1723-1743.

Núñez, R. E. (2008). Reading between the number lines. *Science, 321*(5894), 1293-1294.

Ross, J., & Di Lollo, V. (1971). Judgment and response in magnitude estimation. *Psychological Review, 78*(6), 515.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428-444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237-250.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3*(3), 143-150.

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142*(1), 193.

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review, 18*(3), 557-563.

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81*(6), 1768-1786.

CONCLUSION

The preceding studies presented university undergraduates with a variety of tasks in order to explore so-called “mature” conceptualizations of natural number in a wider range of contexts than has been included in existing research. Taken together, our findings show that undergraduates who have not received formal instruction in the axioms of natural number frequently recruit conceptualizations of natural number which are inconsistent with the formal characterization. Specifically, conceptualizations of natural number did not allow for spontaneous reasoning by informal mathematical induction beyond only the most elementary of mathematical contexts (Chapters 1 and 2). This was the case even when undergraduates were provided with visual evidence which demonstrated the necessity of the theorem for all natural numbers. Qualitative analysis showed that undergraduates frequently expressed ideas about natural number that were at odds with the formal characterization, including a belief that very large magnitude numbers may follow different rules than smaller numbers (Chapter 1). Additionally, while all undergraduates recognized that the natural numbers are infinite, it was only students who had received formal instruction in the axioms of natural number who consistently referred to the successor principle to justify this fact (Chapter 3). Undergraduates who had not received such instruction provided a wider variety of explanations, and in many cases were not able to provide any justification. Finally, an investigation of undergraduates’ number-to-space mappings revealed that, while, as expected, all undergraduates produced linear response patterns, the slopes produced by untrained undergraduates on unfamiliar large-magnitude scales varied significantly from 1, the slope that would be predicted by accurate recruitment of a formally-consistent successor principle (Chapter 4). Notably in each task context - informal mathematical induction, justification of infinity, and number-to-space mappings - we found that it was only undergraduates who had received training in the formal axioms of natural numbers (mostly mathematics majors) who reliably produced responses that would be predicted by formally-consistent conceptualizations of natural number. Taken together, these results call into question the assumption, widely held in developmental psychology, that formally-consistent knowledge of the natural number system emerges spontaneously from mastery of counting and does not require explicit instruction.

Based on these studies, we don't claim to have fully characterized "mature" conceptualizations of natural number. Rather, our studies suggest that there are likely a variety of conceptualizations, and that the formally-consistent notions assumed in major developmental accounts are likely only recruited in the simplest and most constrained task contexts. Additional work is needed to continue to characterize the variety of conceptualizations of natural number that counting adults possess and the contexts in which they are recruited. Developmental psychologists exploring concepts of "natural number" should reconsider the assumption that mastery of counting implies formally-consistent conceptualizations of natural number. Instead of seeking to determine *when* formally-consistent knowledge appears, developmental psychologists should also examine the limits of this knowledge and carefully explore what children do and do *not* understand about the natural number system. In such work, researchers should take care to avoid the use of leading questions and exclusively relying on tasks involving small, familiar numbers. As the studies detailed in this dissertation have revealed, using a wider array of task contexts is essential to more fully characterizing the relationship between counting ability and knowledge of the natural number system.

The findings of this dissertation also have important practical implications for mathematics education, including classroom instruction and curriculum development. To design appropriate instruction, teachers need to make assumptions about what their students already know – and students' learning suffers when these assumptions turn out to be wrong. Consider the case of formal mathematical induction, a proof method that hinges on the formal characterization of natural number and which is notoriously difficult for students to learn (Avital & Libeskind, 1978; Movshovitz-Hadar, 1993; Stylianides et al., 2007). Various studies in mathematics education have investigated students' conceptual difficulties with this method, offering explanations that focus on high-level issues like not understanding nuances in the meanings of algebraic variables (Fischbein & Engel, 1989), or believing that the inductive step assumes the theorem that is to be proven (Ernest, 1984). However, a very simple potential cause has been overlooked: instructors frequently do not present the Dedekind-Peano Axioms before introducing mathematical induction (Zazkis & Leikin, 2010). Instead, instructors may make the same assumption we see in developmental psychology: that students walk into the classroom already possessing the conceptualization of natural number that is necessary for mathematical induction. If students actually do *not* possess such formally-consistent concepts, it is easy to see why they might struggle with the formal proof method which builds upon them. Students' understanding of formal mathematical induction may improve considerably simply by providing explicit instruction in the formal definition of natural numbers.

As part of a larger program examining the relationship between everyday concepts and mathematical knowledge, this research carries important implications for teacher education and development. High quality mathematics teachers must be aware of the ways in which the objects of formal mathematics differ from our everyday conceptualizations, as failing to highlight these differences for students can result in damaging misconceptions and confusion (Lakoff & Núñez, 2000). Existing work has revealed differences between everyday concepts and formal mathematics in high-level domains (e.g., continuity in calculus Marghetis & Núñez, 2013, and transfinite numbers Núñez, 2005). However, mathematics teachers might be surprised to realize that differences can exist even for the simplest mathematical objects: the numbers we use for counting are *not* the same as the set of natural numbers. By demonstrating that deviations between everyday reasoning and formal mathematics exist *at all levels*, this research should inform new approaches to mathematics education. Early career mathematics educators should be encouraged to develop a reflective practice around the ways in which mathematical objects differ from students' everyday experience, in order to design instructional experiences which carefully and effectively bridge the gap between everyday life and genuine mathematical knowledge.

Finally, it is my hope that this dissertation contributes to ongoing critical reflection regarding the notion that mathematics is "everywhere around us", a real and objective reflection of the universe and building block of our everyday lives. This "mythology" is widespread in popular culture and mathematics education, and is reflected in the assumptions and interpretation of scientific research into mathematical thinking and learning. Recently, some scholars have turned a critical eye to these notions, pointing out ways in which this orientation towards mathematics actually serves to perpetuate exclusionary and elitist practices. The insistence that "math is everywhere" obscures the historical and current reality of mathematics as a privileged form of knowledge, which has allowed only an elite few to participate (Barany, 2016). Gutiérrez (2017) argues that the perceived "purity" of mathematics has allowed mathematical achievement to become a proxy for general intelligence, instilling feelings of anxiety and inferiority in students and perpetuating exclusionary practices within the field of mathematics and society at large. These scholars have argued, and I agree, that a more realistic understanding of the role of mathematics in society, including aspects of identity, power, and privilege, would help educators deliver more inclusive and responsible instruction. This dissertation provides another source of support for the need for a reorientation towards mathematics in both research and education, by providing empirical evidence that formally-consistent knowledge of one of the simplest mathematical objects - the set of natural numbers - does not emerge spontaneously from everyday experience. This serves to remind us that the objects of mathematics themselves are cultural artifacts, the products of specific norms, goals, and values internal to the field of mathematics, which are wholly distinct from the goals of everyday life. This distinction is crucial both for researchers seeking to accurately characterize the development of mathematical knowledge, and for educators who wish to provide honest, clear, and inclusive instruction in mathematics.

References

Avital, S. & Libeskind, S. (1978): Mathematical induction in the classroom: Didactical and mathematical issues. *Educational Studies in Mathematics 9*, 429-438.

Barany, M. J. (2016, August 16). *Mathematicians are overselling the idea that "math is everywhere"*. Scientific American Blog Network. Retrieved August 15, 2020, from https://blogs.scientificamerican.com/guest-blog/mathematicians-are-overselling-the-idea-that-math-is-everywhere/

Ernest, P. (1984). Mathematical induction: A pedagogical discussion. *Educational Studies in Mathematics, 15*(2), 173-189.

Fischbein E., & Engel, I. (1989): Psychological difficulties in understanding the principle of mathematical induction", in G Vergnaud, J. Rogalski & M. Artigue (Eds.): *Proceedings of the 13th international conference for the Psychology of Mathematics Education, Vol I* (pp. 276-282). Paris, France: CNRS.

Gutiérrez, R. (2017). Political conocimiento for teaching mathematics. In S.E. Kastberg, A.M. Tyminski, A.E. Lischka, W.B. Sanchez (Eds), *Building support for scholarly practices in mathematics methods* (pp. 11-37). Information Age Publishing.

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from*. New York: Basic Books.

Marghetis, T., & Núñez, R. (2013). The motion behind the symbols: A vital role for dynamism in the conceptualization of limits and continuity in expert mathematics. *Topics in Cognitive Science, 5*(2), 299-316.

Movshovitz-Hadar, N. (1993): The false coin problem, mathematical induction and knowledge fragility. *Journal of Mathematical Behavior*, *12*, 253-268.

Núñez, R. E. (2005). Creating mathematical infinities: Metaphor, blending, and the beauty of transfinite cardinals. *Journal of Pragmatics, 37*(10), 1717-1741.

Stylianides, G. J., Stylianides, A. J., & Philippou, G. N. (2007). Preservice teachers' knowledge of proof by mathematical induction. *Journal of Mathematics Teacher Education*, *10*(3), 145-166.

Zazkis, R., & Leikin, R. (2010). Advanced mathematical knowledge in teaching practice: Perceptions of secondary mathematics teachers. *Mathematical Thinking and Learning, 12*(4), 263-28.