Characterizing students' conceptual difficulties with mathematical induction using visual proofs

Josephine Relaford-Doyle [1] ⬤ · Rafael Núñez [1]

Abstract

This paper describes a study that used a novel method to investigate conceptual difficulties with mathematical induction among two groups of undergraduate students: students who had received university-level instruction in formal mathematical induction, and students who had not been exposed to formal mathematical induction at the university level. The method used a 'visual proof by induction' – a simple image that is designed to demonstrate a theorem that would be formally proven using mathematical induction – to investigate students' conceptualizations of mathematical induction. We report two major findings: first, the majority of students who were familiar with formal mathematical induction had difficulty using the image to justify the theorem, suggesting that their knowledge of the proof method was intimately linked to the algebraic method and thus largely procedural in nature. Second, students who had not studied formal mathematical induction generally used the image as the basis of a standard inductive generalization and did not recognize that the image could be used to establish the necessity of the theorem. Surprisingly, these students often expressed conceptualizations of natural number that were inconsistent with the formal characterization that forms the basis of formal mathematical induction. Implications for education are discussed.

Keywords Mathematical induction · Visual representations · Visual proof

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

✉ Josephine Relaford-Doyle
  jrelaford@ucsd.edu

[1] Department of Cognitive Science, University of California, San Diego, 9500 Gilman Dr, La Jolla,
     CA 92093, USA

they are necessary, and why they allow for the conclusion that a theorem is *necessarily* true for all natural numbers (Lowenthal and Eisenberg 1992; Woodall 1981; Baker 1996; Harel 2001). 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 (Smith 2003; Baroody 2005).

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 actually 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 and 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 (Ernest 1984; Fischbein and Engel 1989; Avital and Libeskind 1978; 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 and 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 and 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 (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 Fig. 1.

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 Fig. 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 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.

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 and 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 and 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.
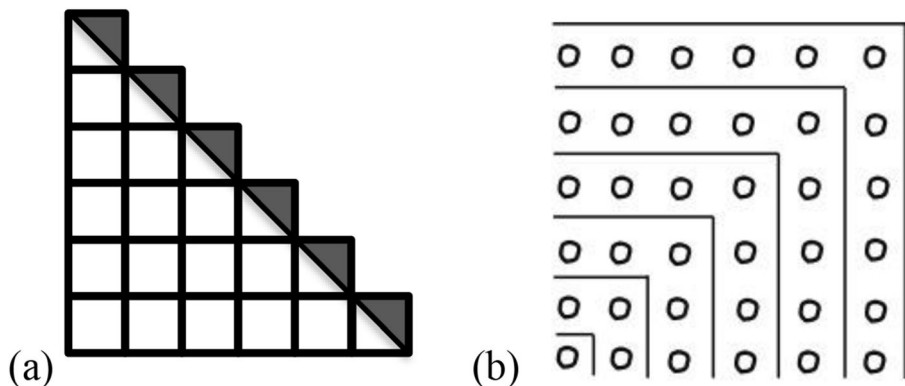

Fig. 1 Visual proofs of theorems that could be formally proven using mathematical induction. (a): $1 + 2 + 3 + \ldots + n = (n^2 + n)/2$. (b): $1 + 3 + 5 + \ldots + (2n - 1) = n^2$. Figures adapted from Brown (1997)
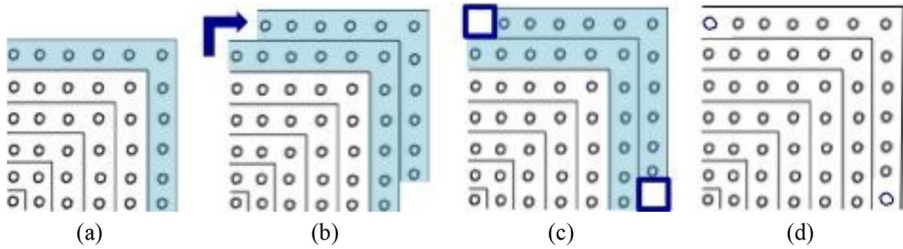
Fig. 2 (a) To construct the next layer, begin by considering the current outermost later. (b) Make a copy of this layer, and move it one unit up and one unit right. (c) This results in two open positions that must be filled in order to maintain the square shape. (d) Since the original outermost layer 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

In this study we use the visual proof in Fig. 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 regard 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 (Smith 2003; Baroody 2005) 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 and 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 Fig. 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."[1] In the first phase of the study we

---

[1] 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.

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 (Fig. 3).

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 Fig. 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.
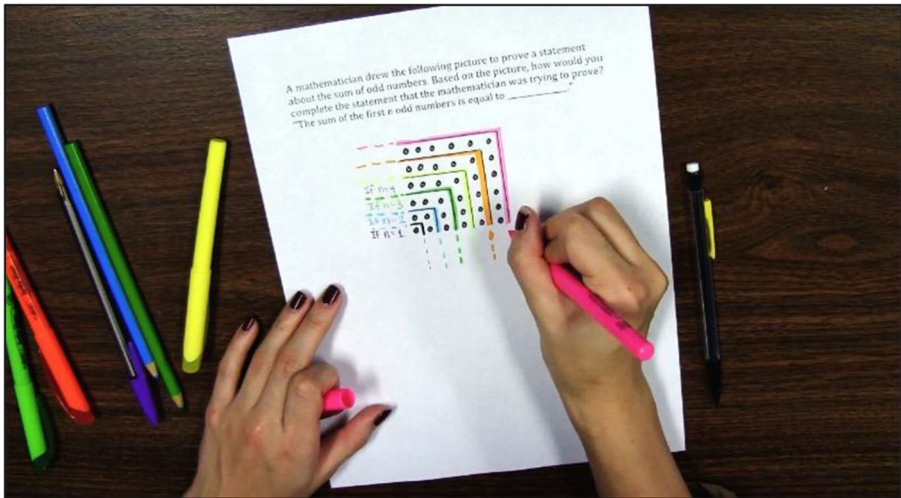
Fig. 3 Participant workspace

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. 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.

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 min, SD = 5.2 min; Untrained M = 9.31 min, SD = 5.6 min; $t(37) = -1.03$, $p = 0.31$), or the duration of their tutorial videos (Trained M = 6.26 min, SD = 5.49; Untrained M = 4.12 min, SD = 2.16 min; $t(37) = 1.49$, $p = 0.15$). 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-

Table 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."*

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; Fig. 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).

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 (Fig. 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
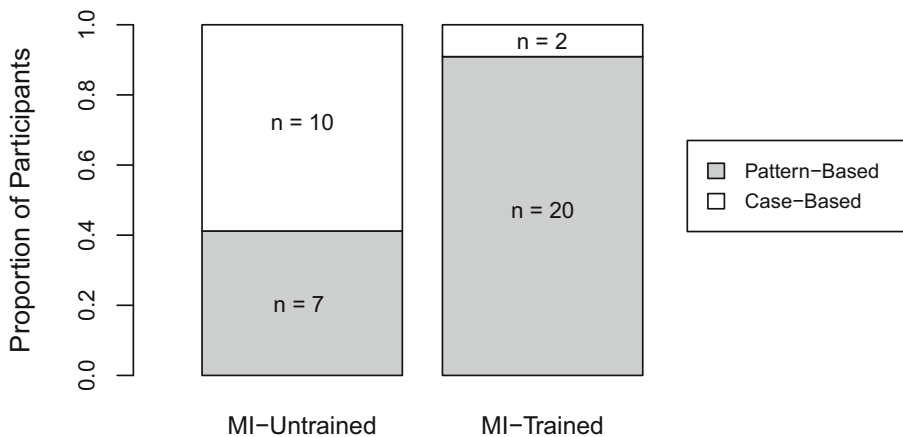
Fig. 4 MI-Trained participants overwhelmingly preferred pattern-based explanations, while the majority of MI-Untrained participants provided case-based explanations

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.

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 (Fig. 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
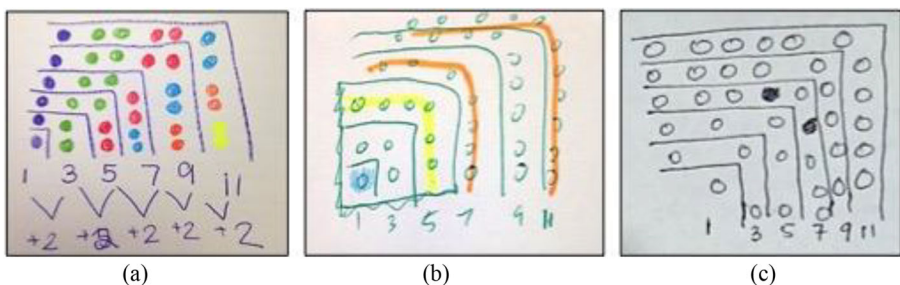

Fig. 5 In their tutorial videos some MI-Untrained participants produced images that violated the essential row-column structure of the original figure

high resistance to large-magnitude counterexamples (Fisher Exact Test, $p = 0.007$, Odds Ratio = 7.02; 95% CI 1.4, 41.2; Fig. 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.

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 counter-examples 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
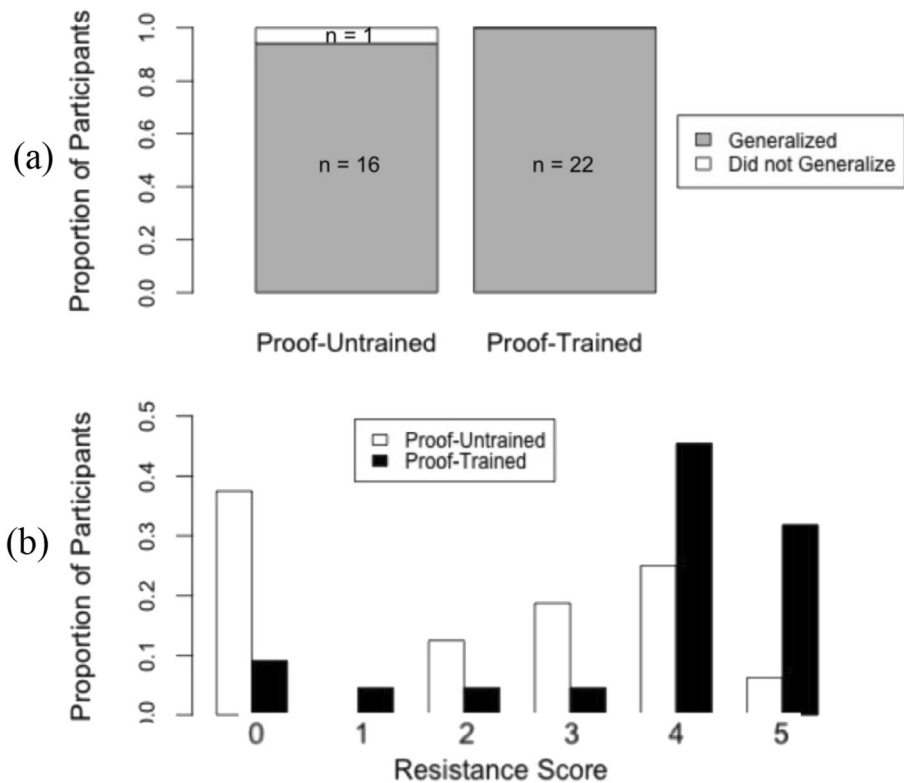

Fig. 6 All participants were willing to generalize the theorem to nearby cases (a). However, MI-Untrained participants were significantly less likely to show a high degree of doubt that large-magnitude counterexamples to the statement are possible (b). Figures from Relaford-Doyle and Núñez (2017)

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 and 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 and 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 males peers (Tariq et al. 2013; Peters 2013; Felder et al. 1995); 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) × 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 (Smith 2003; Baroody et al. 2013). 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 and 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 and 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 and Dreyfus 1991; Inglis and 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 (Schoenfeld 1985; Lesh et al. 1987; Pape and Tchoshanov 2001). 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; Fig. 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.

Compliance with Ethical Standards

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*. New York: Paper presented at the Annual Meeting of the American Educational Research Association.

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 ed.). 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). 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: 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 Educsational 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.