Title
When does a ʻvisual proof by inductionʼ serve a proof-like function in mathematics?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Relaford-Doyle, Josephine
Núñez, Rafael

Publication Date
2017

Peer reviewed

When does a ‘visual proof by induction’ serve a proof-like function in mathematics?

Josephine Relaford-Doyle (jorelafo@ucsd.edu)
Rafael Núñez (rnunez@ucsd.edu)
Department of Cognitive Science, University of California – San Diego
9500 Gilman Drive, La Jolla, CA 92093 USA

Abstract

A proof by mathematical induction demonstrates that a general theorem is necessarily true for all natural numbers. It has been suggested that some theorems may also be proven by a ‘visual proof by induction’ (Brown, 2010), despite the fact that the image only displays particular cases of the general theorem. In this study we examine the nature of the conclusions drawn from a visual proof by induction. We find that, while most university-educated viewers demonstrate a willingness to generalize the statement to nearby cases not depicted in the image, only viewers who have been trained in formal proof strategies show significantly higher resistance to the suggestion of large-magnitude counterexamples to the theorem. We conclude that for most university-educated adults without proof-training the image serves as the basis of a standard inductive generalization and does not provide the degree of certainty required for mathematical proof.

Keywords: mathematical reasoning; proof; mathematical induction; visual proof; induction; generalization

Introduction

Mathematics has been defined as “the science which draws necessary conclusions” (Peirce, 1881). To this end, proofs are indispensible to formal mathematics. A mathematical proof uses deductive logic to establish the truth of a general theorem – for instance, that a property holds for *all* triangles, or *all* natural numbers. Relative to the logical system being used, the conclusion of the proof is *certain*; if the premises are true, then the conclusion is necessarily true as well. The certainty of results obtained through formal, deductive proof is a defining feature of mathematics.

Mathematical induction, despite what its name suggests, is a well-established *deductive* proof method that can be used to prove that a theorem holds for all natural numbers. It has been suggested that some general theorems that can be formally proved using mathematical induction may also be proved using specially designed images known as ‘visual proofs’ (Brown, 2010). The claim that a ‘visual proof by induction’ can prove a general theorem is an interesting one, since any image is necessarily finite and thus can only display a particular set of cases of the theorem. Case-based argumentation falls under the umbrella of inductive reasoning, which does not provide certain conclusions and is not accepted in formal mathematical justification. However, visual proofs contain additional structure that could be leveraged to demonstrate that a theorem necessarily holds in *all* cases, even those not depicted in the image. Thus, it is possible that a visual proof, despite displaying only a finite number of cases, could serve a proof-like function for some viewers.

Although the status of visual proofs is at the center of a debate in the philosophy of mathematics (see, e.g., Brown, 2010; Doyle, Kutler, Miller, & Schueller, 2014; Folina, 1999), they have been largely ignored within Cognitive Science and little is known about the nature of reasoning with these images. How do viewers reason with a visual proof by induction? Do they consider the conclusions to be certain, as in mathematical induction, or only likely, as in standard inductive reasoning?

Induction in Mathematics, Mathematical Induction, and Visual Proofs

The distinction between certain, necessary conclusions and probable or likely conclusions is of central concern in mathematics. Proofs – deductive arguments which provide certain conclusions – are exalted. The writing of proofs, however, comprises only a small part of mathematical practice, and it is widely acknowledged that inductive reasoning plays an important role in mathematics (see Polya, 1954 for an account of induction in mathematics). A commonly held view is that inductive reasoning is an essential part of mathematical *discovery*, while deduction is required for formal mathematical *justification* (i.e., proof).

Consider the expressions in Figure 1(a). One might notice a pattern in these examples, namely, that when one adds consecutive odd numbers starting at 1, the resulting sum seems to be the square of the number of terms being added. We might guess that this pattern holds for other numbers; for example, we might predict that the sum of the first 8 odd numbers is 64. However, while these six examples allowed us to discover a *possible* relationship, the examples themselves do not *prove* that the general theorem is true for all natural numbers. Without a formal proof, any conjecture we have is uncertain and remains open to the possibility of counterexamples. A formal proof of our theorem using mathematical induction is given in Figure 1(b).

Figure 1(c) shows a visual proof of the same theorem (from Brown, 1997). In the image, consecutive odd numbers of dots are arranged in layers, beginning with 1 in the lower left-hand corner. When the dots in the first n layers are considered together the resulting array forms a square, and so the total number of dots in the array is given by n^2. While the image displays only the first six cases of the general theorem, a viewer might be inclined to guess that the pattern will continue to hold as more layers are added, and therefore be convinced that the general theorem is true. Indeed, images such as these have been described as “rapidly and deeply convincing” (Doyle et al., 2014).

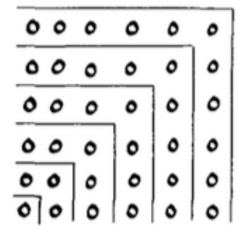
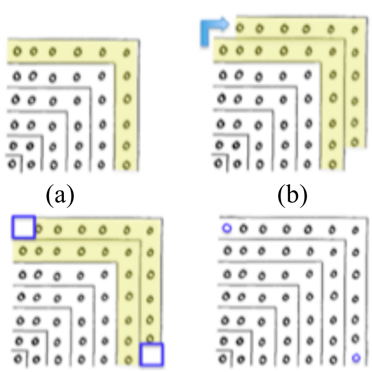
Theorem: The sum of the first *n* odd numbers is equal to n^2.		
(a) Six cases $1 = 1^2$ $1 + 3 = 4 = 2^2$ $1 + 3 + 5 = 9 = 3^2$ $1 + 3 + 5 + 7 = 16 = 4^2$ $1 + 3 + 5 + 7 + 9 = 25 = 5^2$ $1 + 3 + 5 + 7 + 9 + 11 = 36 = 6^2$	(b) Proof by mathematical induction Theorem: $1 + 3 + \ldots + (2n - 1) = n^2$ Base case: $n = 1 \rightarrow 1 = 1^2$ Inductive step: Assume $1 + 3 + \ldots + (2k - 1) = k^2$, for some fixed number k. Adding the next odd number $2k + 1$ to both sides of the equation, we have: $1 + 3 + \ldots + (2k - 1) + (2k + 1) = k^2 + (2k + 1)$ Re-writing the last odd term and factoring the right side gives us: $1 + 3 + \ldots + (2k - 1) + [2(k+1) - 1] = (k + 1)^2$, QED	(c) Visual Proof 

Figure 1: Varying forms of evidence for a general theorem. (a) Specific cases suggest, but do not prove, the general theorem. (b) Formal proof of the general theorem by mathematical induction. (c) Visual proof (from Brown, 1997).

It is unclear, however, the exact nature of the conclusions drawn from the visual proof. As many have pointed out, any image is necessarily finite and thus can only display particular cases of a general theorem (Doyle et al., 2014; Rips & Asmuth, 2007). This would suggest that the image in Figure 1(c), like the cases presented in 1(a), would serve as the basis for a standard inductive generalization – the image might convince the viewer that the theorem *likely* holds for all natural numbers, but cannot provide certainty. On the other hand, the image contains structure that is not available in the same cases presented numerically, and which could be exploited in order to demonstrate that the property would *necessarily* continue to hold for values not depicted in the image. Such an argument would need to demonstrate that the square shape is preserved if and only if the next layer contains the next consecutive odd number of dots. For example, if we start with an n x n square, the $(n + 1)^{st}$ layer could be constructed by copying the n^{th} layer and translating the copy up one unit and right one unit (Figure 2a, b). This results in two vacant positions that must be filled in order to maintain the square shape (Figure 2c, d). Thus, every layer must contain exactly two more dots than its predecessor. Since the difference between any two consecutive odd numbers is 2, we can conclude that the new layer must contain the next consecutive odd number of dots.



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

Figure 2: Rigorous image-based argument for the general theorem

Though not a traditional deductive proof, it could be argued that an argument such as this does establish the truth of the general theorem for all natural numbers, and that this conclusion meets the level of certainty required for mathematical proof. It is unknown how accessible such arguments are to viewers, and, more generally, how closely the conclusions drawn from the visual proof resemble the conclusions drawn from a formal proof. In this study we seek to assess the extent to which a visual proof may serve a proof-like function. Specifically, we ask two key questions: (1) Given the visual proof, do viewers generalize the theorem to cases not depicted in the image? (2) If so, is that conclusion considered certain, as in mathematical induction, or only likely, as in standard inductive reasoning?

Finally, to properly address these questions we consider *who* is viewing the image and *in what context* the image is viewed. In this study we compare two groups of viewers, one drawn from the general population of university students and one drawn from a group that has received university-level training in formal proofs, including mathematical induction. Additionally, to address the key theoretical difference in mathematics between discovery (in which inductive reasoning is acceptable) and justification (in which it is not), we manipulate the context in which the image is viewed by varying the amount of information provided to the viewer.

Method

Participants

Two groups were drawn from distinct populations. The first group (n = 25) was recruited through the university subject pool. None of these participants had taken a university-level course on mathematical proofs, and so we refer to this group as "proof-untrained" (PU). The second group (n = 24) was recruited through the mathematics department and consisted of individuals who had received at least a B- in "Mathematical Reasoning", a university-level mathematics course on formal proof strategies including mathematical induction. We refer to this group as "proof-trained" (PT). PT participants had taken significantly more university-level math classes than had PU participants (mean PT = 6.67, mean PU = 2.8, t = 6.58, df = 44.63, $p < 0.001$). No PU participants had taken the "Math Reasoning" course on proofs (but four of the 25 indicated familiarity with mathematical induction).

Materials

Each participant received one of three tasks. All three tasks included the visual proof in Figure 1(c), which was designed to prove the target statement "The sum of the first *n* odd numbers is equal to n^2", but varied in the amount of information that was provided to the reader.

Condition A – *Justification*: The participant was given the full target statement and the visual proof. They were asked to explain how the picture shows that the statement is true.

Condition B – *Supported Discovery*: The participant was given the visual proof and a fill-in-the-blank version of the target statement ("The sum of the first *n* odd numbers is equal to ______."). They were asked to fill in the blank and explain how they got their answer.

Condition C – *Full Discovery:* The participant was given the visual proof and told that a mathematician drew the picture while trying to prove a statement about the sum of odd numbers. They were asked to guess what the mathematician was trying to prove and explain their answer.

Additionally, participants completed a background questionnaire in which they provided the number and names of university math courses they had taken and their level of familiarity with mathematical induction.

Procedure

Each participant received one of the three task sheets, and the researcher explained that their ultimate task was to create a short tutorial video in which they would explain their response as clearly as possible to potential third-party viewers. A camera was set up directly above the participant's workspace, recording their writing, speech, and manual gestures (Figure 3). Before filming their video each participant was given as much time as they needed to think and plan their response. During this time participants had access to pencils, highlighters, and blank paper, and were free to add any markings to the sheet that might be helpful in explaining their response. The participant indicated they were ready to start their tutorial video by placing a sign under the camera, and then filmed their explanation. The planning and filming were entirely self-paced and occurred without the researcher present.

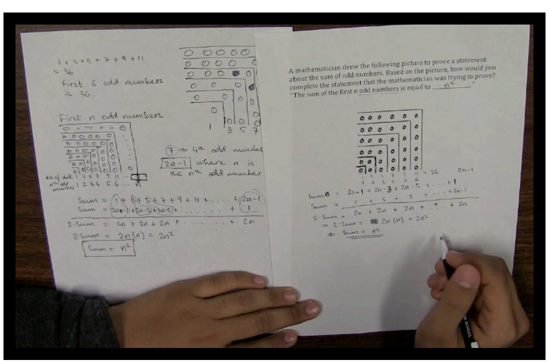


Figure 3: Screenshot from video footage of a participant's workspace

When the participant indicated that they had finished their tutorial video, the researcher returned to the room and conducted a semi-structured interview with the participant. To assess whether the participant had generalized the target statement to cases not depicted in the image, any participant who demonstrated understanding of the target statement was asked two questions: "Do you think the statement is true in all cases?" (Q1) and "What would be the sum of the first 8 odd numbers?" (Q2). If the participant indicated generalization to nearby cases (by answering "yes" and "64", respectively), the researcher raised the possibility that large-magnitude counterexamples to the statement may exist and asked the participant what they thought about that suggestion. Any participant who resisted the suggestion of counterexamples was asked how they would argue against such a possibility. After the interview all participants completed the background questionnaire.

Analysis

Two coders scored each video for six distinct outcomes. The participant's tutorial video received three scores:

(a) Mathematical Statement: Rated whether the participant demonstrated understanding of the target statement.

(b) Explanation Strategy: Rated whether the participant gave a *case-based* explanation (using the image to show particular cases of the statement), or a *pattern-based* explanation (describing a general pattern in the image).

(c) Relevant Features of Image: Reflected which features of the image the participant identified as relevant, including odd numbers in layers, square shape, possibility of pattern extension, and necessity of pattern extension.

An additional three scores were given based on the interview portion of the study.

(d) Generalization: Rated whether the participant demonstrated generalization of the target statement to nearby cases as assessed by questions Q1 and Q2.

(e) Resistance to Counterexamples: Rated the participant's resistance to the possibility of large-magnitude counterexamples, ranging from no resistance (0) to complete rejection (5).

(f) Image-Based Argument: Rated whether the participant provided a rigorous image-based argument comparable to the argument represented in Figure 2.

For nominal criteria (a)-(d) and (f), the two coders showed 96.2% agreement (Cohen's Kappa=0.96). Criterion (e) was was scored on 1-5 scale and also showed high reliability between coders (Krippendorff's alpha = 0.894).

Results

The two groups differed with respect to the ability to demonstrate understanding of the target statement across conditions. While proof-trained (PT) participants systematically demonstrated such understanding, proof-untrained (PU) participants' ability to do so varied significantly across conditions (Fisher Exact test, $p = 0.012$; Figure 4). All PU participants showed understanding of the target statement when it was provided (Condition A), but only 6/8 participants provided the target response of n^2 in Condition B, and 3/9 participants generated the complete target statement in Condition C. PT participants, on the other hand, did not significantly vary in their likelihood to demonstrate understanding of the target statement across conditions (Fisher Exact test, $p = 0.3$; Figure 4).

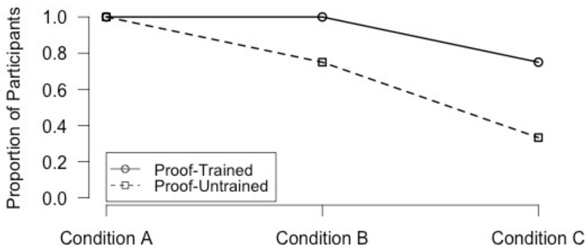


Figure 4: Proportion of participants who demonstrated understanding of target statement.

Of the 17 total PU participants across all conditions who demonstrated understanding of the target statement, 10 (59%) used case-based explanations in their tutorial video (Figure 5); notably, none of these 10 participants referred to the square shape as a relevant feature of the image during their explanation. Explanation strategy varied significantly across conditions, with participants in Condition A showing a stronger preference for case-based strategies, while participants in Condition C were more likely to generate pattern-based explanations (Fisher Exact test, $p = 0.026$) and more likely to mention the square shape ($p = 0.026$). As we found no other significant effects of task context, in the following analysis we group participants across all conditions who demonstrated understanding of the target statement, keeping PU and PT groups separate.

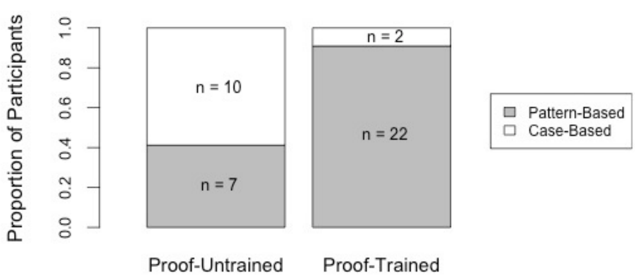


Figure 5: Explanation Strategy: PU participants were significantly more likely than PT participants to give case-based (rather than pattern-based) explanations.

Proof-Untrained (PU) Group In the interview, 16 of the 17 (94.1%) PU participants who demonstrated understanding of the target statement indicated a willingness to generalize the statement to nearby cases (Figure 6a). Only 5 (31%) of these participants indicated a high degree of resistance to large-magnitude counterexamples (characterized by a resistance score of 4 or higher), and only one stated that counterexamples were impossible (Figure 6b). Notably, questionnaire responses revealed that three of the five PU participants who showed high resistance were familiar with mathematical induction. When asked for an argument against counterexamples, only two PU participants were able to generate a rigorous argument based on the image. PU participants had taken significantly fewer university-level math courses than had PT participants; however, number of math courses was not significantly correlated to any study outcomes for the PU group.

Proof-Trained (PT) Group Across all conditions there were 22 PT participants who demonstrated understanding of the target statement. These participants were significantly more likely than PU participants to provide pattern-based explanations (Fisher Exact test, $p = 0.026$), with only 2/22 (9%) relying on case-based strategies (Figure 5). PT participants were significantly more likely than PU participants to mention the square shape as a relevant feature of the image (21/22; Fisher Exact test $p < 0.001$). In the interview, all 22 PT participants who demonstrated understanding of the target statement indicated a willingness to generalize the statement to nearby cases not depicted in the image. The likelihood to generalize did not differ between PT and PU participants (Figure 6a); however, PT individuals were significantly more likely to indicate a high degree of resistance to the suggestion of large-magnitude counterexamples (17/22; Fisher Exact test, $p = 0.008$; Figure 6b). When considering all participants who demonstrated understanding of the target statement, PT participants were significantly more likely to provide a rigorous image-based argument against counterexamples than PU participants (8/22; Fisher Exact test, $p = 0.035$).

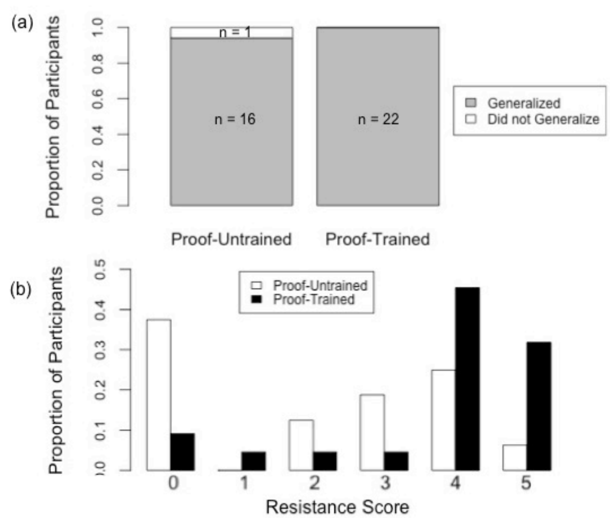


Figure 6: (a) Participants in both groups generalized the target statement to nearby cases. (b) However, PT participants showed significantly higher resistance to large-magnitude counterexamples than did PU participants.

Discussion

The present study explored the conditions in which a visual proof by induction may serve a proof-like function, characterized by generalization to all natural numbers and a belief that this conclusion is necessarily true. Our findings reveal significantly different outcomes for the proof-trained and proof-untrained participants. Specifically, while both groups demonstrated a willingness to generalize to nearby cases, the PU participants showed relatively low resistance to the suggestion that large-magnitude counterexamples may exist. This suggests that for these viewers the visual proof serves as the basis for a standard inductive generalization, and does not provide certainty. Further evidence for this analysis comes from the observation that PU participants were significantly more likely to provide case-based explanations, using the image to demonstrate one or more particular cases of the general theorem. PT participants, on the other hand, showed higher resistance to counterexamples and were more likely than PU participants to provide a rigorous argument using the image. Thus, it seems that the image can serve a proof-like function for viewers who have been trained in formal proof methods. The significant differences between the PT and PU groups contradict claims that visual proofs by induction are equally convincing to all viewers regardless of their knowledge of mathematical induction (Brown, 2010), or that interpreting the image as a proof requires only “basic secondary school knowledge of mathematics” (Jamnik, 2001).

We were surprised to find that the PU participants – highly educated adults enrolled at a prestigious university – often overlooked key features of the visual proof. Less than 60% of the PU participants who demonstrated understanding of the target statement mentioned the square shape as a relevant feature of the image. Furthermore, many participants who re-drew the image during their explanation did so in a way that violated the row-column structure of the square array (Figure 7), indicating that they were truly unaware of its importance. However, failure to notice the relevance of the square shape does not explain the PU group’s low resistance to counterexamples, as mentioning the square shape was not significantly related to high resistance (Fisher Exact test, $p = 0.59$) within this group.

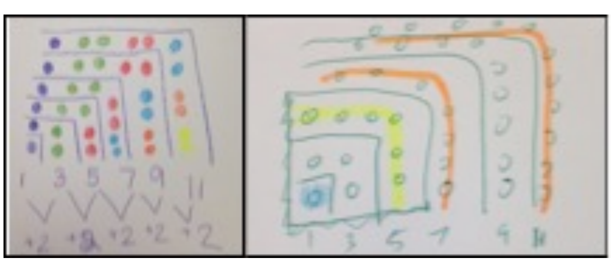


Figure 7: Work of PU participants who re-drew the image in a way that violated the row-column structure and square shape of the array

There were 5 PU participants who expressed a high degree of resistance to the suggestion of large-magnitude counterexamples, two of whom were unfamiliar with mathematical induction. We cannot conclude, however, that the image was serving a proof-like function for these viewers. Prior research has shown that adults do not reliably distinguish between inductive and deductive mathematical arguments and often accept case-based arguments as valid proofs of statements about infinite sets (Eliaser, 2000; Martin & Harel, 1989). Thus, even if the image functions as a basis for a standard inductive generalization, we would nonetheless expect to see a group of participants who find it highly convincing.

PT participants were more likely than PU participants to show high resistance to counterexamples, and subsequently more likely to provide a rigorous image-based argument for the general theorem. What accounts for these differences? One possibility is that PT participants had been exposed to significantly more university-level mathematics than PU participants. However, the number of university math classes taken prior to participation in the study was not related to any outcome for either group. This suggests that the differences between the two groups cannot be explained simply based on differing amounts of exposure to general mathematics. Instead, it seems that training in proof-writing – a specific and highly technical mathematical practice – may make viewers more likely to draw certain conclusions from the image.

Based on our data, exposure to proof-writing could make certain conclusions more likely in at least three ways (not mutually exclusive). First, it could be that some aspect of the task reminds PT participants of the specific proof-method of mathematical induction (indeed, 75% of PT participants mentioned mathematical induction at least once during their video and/or interview). These viewers might then recognize that they could use mathematical induction to prove the target statement, and perhaps even complete the proof (as did 25% of our PT participants). Thus, knowledge of the formal proof could provide an alternate means of acquiring certainty about the conclusion; once achieving this certainty, participants may be more likely to attempt to generate an alternate argument based on the image. However, it cannot be the case that knowledge of mathematical induction is a necessary condition for such an argument, as we observed one participant who was not familiar with mathematical induction produce a rigorous image-based justification of the general theorem.

Second, in addition to gaining familiarity with mathematical induction, training in proof-writing would also expose individuals to a set of general mathematical norms which may lead these participants to demonstrate a higher degree of certainty. All participants who demonstrated understanding of the target statement were asked if they believed the statement to be “true in all cases”; however, the two groups likely interpreted this question differently. For PT participants, “all” (when used in a mathematical context) is a technical term, which by definition implies the impossibility of counterexamples. PU participants may have been operating with an everyday use of "all", in which the term is considered synonymous to "generally" or "usually" (e.g., "All Californians love the beach."). In this light, the two groups' differing responses to the suggestion of counterexamples may be revealing of their different conceptualizations of the term "all".

Finally, in the practice of writing proofs one learns standard ways of representing general mathematical objects, and these representations may be useful in interpreting visual proofs. We observed that PT participants often invoked the fact that odd numbers are of the form ($2n + 1$) to explain how they knew that the layers (symmetric legs extending from a single corner dot) would always contain an odd number of dots. Fewer PU participants offered this argument, perhaps because they were not familiar with algebraic representations of parity. Future studies are necessary to determine whether knowledge of algebraic representations of parity allow viewers to exploit structure available in an image.

We also explored how conclusions drawn from the image differed between contexts of justification (Condition A) and discovery (Conditions B/C). We observed only three effects of task context. First, PU participants' ability to demonstrate understanding of the target statement varied significantly between conditions. Specifically, PU participants – while perfectly capable of understanding the target statement when it was provided – were highly unlikely to "discover" the full target statement based only on the image (with only 25% able to do so in Condition C). These results suggest that, even for most highly-educated viewers, the image must be accompanied by the statement it is intended to prove (or at least a substantial hint, as in Condition B). Next, we observed that PU participants in contexts of full discovery (Condition C) who generated the target statement were more likely to provide pattern-based justifications and more likely to mention the square shape than PU participants who were given the full target statement. This is not surprising, since the target statement was unknown to these participants at the outset of the task and was only discovered if the participant noticed a pattern in the image.

We find it interesting that these three results were the only effects of task context for either group. All participants who demonstrated understanding of the target statement were likely to generalize it to nearby cases, regardless of the justification/discovery context in which they had seen the visual proof. Subsequent resistance to large-magnitude counterexamples – relatively low for PU participants, and high for PT participants – did not vary significantly between task contexts. This suggests that certainty of the conclusion has more to do with the viewer's exposure to mathematical proof-writing than with the justification/discovery context in which the image is viewed. The lack of any effect of task context for PT participants suggests that the sharp distinction between justification and discovery, of such theoretical importance in mathematics, is less prevalent in advanced mathematical practice.

Conclusion

In this study we investigated the reasoning underlying a visual proof by induction and the nature of the conclusions drawn from the image. A visual proof by induction displays a particular set of cases of a general theorem, yet it also contains structure that could be used to construct a rigorous argument that the theorem is necessarily true for all natural numbers. We found that, while most viewers are willing to generalize the theorem to nearby cases not displayed in the image, viewers who have been exposed to formal proof methods (including mathematical induction) show significantly higher resistance to the suggestion that large-magnitude counterexamples to the theorem are possible, and are significantly more likely to provide a rigorous image-based argument against counterexamples. For participants without proof-training, conclusions drawn from a visual proof resemble a standard inductive generalization and do not display the level of certainty associated with mathematical proof. These results are consistent between contexts of justification and discovery, indicating that the certainty of conclusions drawn from a visual proof by induction are primarily dependent on the viewer's exposure to proof-writing, rather than the viewing context.

Acknowledgments

We are grateful to Lillian Chang and Emmanuel Gygi for their assistance on this project.

References

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

Brown, J. R. (2010). *Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures*. Routledge.

Doyle, T., Kutler, L., Miller, R., & Schueller, A. (2014). Proofs without words and beyond. *Convergence*. Retrieved from http://www.maa.org/press/periodicals/convergence/proofs-without-words-and-beyond.

Eliaser, N. M. (1998). *What constitutes a mathematical proof?* (Doctoral dissertation, Northwestern University).

Folina, J. (1999). Pictures, proofs, and 'mathematical practice': Reply to James Robert Brown. *The British Journal for the Philosophy of Science*, *50*(3).

Jamnik, M. (2001). *Mathematical reasoning with diagrams*. University of Chicago Press.

Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education, 20*(1).

Peirce, B. (1881). Linear associative algebra. *American Journal of Mathematics*, *4*(1).

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

Rips, L., & Asmuth, J. (2007). Mathematical Induction and Induction in Mathematics. In A. Feeney & E. Heit (Eds.), *Inductive Reasoning: Experimental, Developmental, and Computational Approaches*. Cambridge University Press.