Mathematical Induction and Induction in Mathematics

Lance J. Rips[1]

Jennifer Asmuth

Northwestern University

To appear in: A. Feeney and E. Heit (Eds.), *Induction*. Cambridge, England: Cambridge University Press.

Mathematical Induction and Induction in Mathematics

However much we many disparage deduction, it cannot be denied that the laws established by induction are not enough.

Frege (1884/1974, p. 23)

At the yearly proseminar for first-year graduate students at Northwestern, we presented some evidence that reasoning draws on separate cognitive systems for assessing deductive versus inductive arguments (Rips, 2001a, 2001b). In the experiment we described, separate groups of participants evaluated the same set of arguments for deductive validity or inductive strength. For example, one of the validity groups decided whether the conclusions of these arguments necessarily followed from the premises, while one of the strength groups decided how plausible the premises made the conclusions. The results of the study showed that the percentages of “yes” responses (“yes” the argument is deductively valid or “yes” the argument is inductively strong) were differently ordered for the validity and the strength judgments. In some cases, for example, the validity groups judged Argument A to be valid more often than Argument B, but the strength groups judged B inductively strong more often than A. Reversals of this sort suggest that people do not just see arguments as ranging along a single continuum of convincingness or probability but instead employ different methods when engaged in deductive versus inductive reasoning.

Earlier imaging evidence by Goel, Gold, Kapur, and Houle (1997) and by Osherson et al. (1998) had implicated separate brain regions when participants evaluated arguments for validity versus plausibility, and these earlier data had inspired our own experiment. All these results cast doubt on the view that there's a homogeneous "analytic" reasoning system responsible for correctly solving deductive and probabilistic problems.

But an incident that followed the prosem alerted us that not everyone was buying into the our reasoning distinctions. A faculty colleague who had attended the prosem stopped us later and questioned whether deduction and induction were as distinct as we had claimed. Wasn't mathematical induction a counterexample to the separation between these two forms of reasoning? Mathematical induction is essential in many deductive proofs in mathematics, as any high school or college student knows. So how could induction and deduction be isolated subsystems if mathematicians freely used one to support the other?

At the time, we thought we had a perfectly good answer to our colleague's question. Mathematical induction is really nothing like empirical induction but is instead a deductive technique, sanctioned by the definition of the natural numbers (and certain other number systems). It's true that both types of "induction" often attempt to establish the truth of a generalization (*All cyclic groups are Abelian*; *All freshwater turtles hibernate*). So there is a loose analogy between them that might suggest labeling both "induction." But this is really no more than a case of ambiguity or polysemy. Mathematical induction in its usual form requires showing that the generalization holds of a base case (e.g., that the generalization is true for 0) and then showing that if the generalization is true for an arbitrary number k, it is also true for $k + 1$. Hence, by strict deductive reasoning, the generalization must be true for all natural numbers (0, 1, 2, ...). The conclusion of such a proof shares with all deductively valid conclusions the property that it is necessarily true or true in all possible worlds in which the givens are true.

A typical high school example of mathematical induction is the proof that the sum of the first n natural numbers is $\frac{1}{2} n (n + 1)$. Certainly this relationship holds when $n = 0$. If we assume that the relationship holds for the first *k* natural numbers (i.e., the sum of 0 through *k* is ½ *k* (*k* + 1)), then the sum of the first *k* + 1 numbers must be:

$$\sum_{i=0}^{k+1} i = \frac{k(k+1)}{2} + (k+1)$$
$$= \frac{k^2 + 3k + 2}{2}$$
$$= \frac{(k+1)(k+2)}{2}$$

The last expression is also of the form ½ *n* (*n* + 1). So this *sum formula* necessarily holds for all natural numbers.

By contrast, empirical induction leads to conclusions that are not necessarily true but only probably true, even when these conclusions involve mathematical relationships. For example, we might hope to establish inductively that people forget memories at a rate equal to $y = b - m \ln(t)$, where *t* is time since the remembered event and *m* and *b* are constants. To support such a conclusion, we might examine forgetting rates for many people and many event types and show that forgetting follows the proposed function in a convincing range of cases (Rubin & Wenzel, 1996). Even if the function fit the data perfectly, though, no one would suppose that the evidence established the conclusion as necessarily true. It's not difficult to imagine changes in our mental make up or changes in the nature of our everyday experience that would cause us to forget events according to some other functional form. The best we can hope for is a generalization that holds for people who are similar to us in causally important ways (see Heit's chapter in this volume for an overview of empirical induction).

We were able to convince our colleague that this line of thinking was correct. Or perhaps he was just too polite to press the point. In fact, we still believe that this standard answer isn't too far off the mark. But we also think that we dismissed our colleague's objection too quickly. One worry might be that the theoretical distinction between math induction and empirical induction is not as clear as we claimed. And, second, even if the theoretical difference were secure, it wouldn't follow that the psychological counterparts of these operations are distinct. In this chapter we try to give a better answer to the objection by examining ways that induction could play a role in mathematics. In doing so, we won't be presenting further empirical evidence for a dissociation between inductive and deductive reasoning. We will be assuming that the previous evidence we've cited is enough to show a difference between them in clear cases. Our hope is to demonstrate that potentially unclear cases—ones in which people reach generalizations—don't compromise the distinction (see Oaksford & Hahn, this volume, for a more skeptical view of the deduction/induction difference).

We consider first the relationship between math induction and empirical induction to see how deep the similarity goes. Although there are points of resemblance, we argue that there are crucial psychological differences between them. This isn't the end of the story, however. There are other methods of reaching general conclusions in math besides math induction, and it is possible that one of them provides the sort of counterexample that our colleague was seeking. One such method involves reasoning from an arbitrary case—what is called *universal generalization* in logic. Several experiments suggest that people sometimes apply this method incorrectly, and we examine these mistakes for evidence of a continuum between inductive and deductive methods. We claim, however, that although the mistakes show that people use inductive procedures when they should be sticking to deductive ones, they provide no evidence against a qualitative deduction/induction split.

Of course, we *won't* be claiming that there's no role for induction in mathematics. Both students and professional mathematicians use heuristics to discover conjectures and methods that may be helpful in solving problems (see Polya, 1954, for a classic statement on the uses of induction in mathematics). Our claim is simply that these inductive methods can be distinguished from deductive proof, a point on which Polya himself agreed: "There are two kinds of reasoning, as we said: demonstrative reasoning and plausible reasoning. Let me observe that they do not contradict each other; on the contrary they complete each other" (Polya, 1954, p. vi).

Mathematical Induction and Universal Generalization

In their *The Foundations of Mathematics*, Stewart and Tall (1977) provide an example of a proof by induction similar to the one we just gave of the sum formula. They then comment:

> Many people regard this as an 'and so on...' sort of argument in which the truth of the statement is established for $n = 1$; then, having established the general step from $n = k$ to $n = k + 1$, this is applied for $k = 1$ to get us from $n = 1$ to $n = 2$, then used again to go from $n = 2$ to $n = 3$, and so on, as far as we wish to go... The only trouble is that to reach large values of n requires a large number of applications of the general step, and we can never actually cover *all* natural numbers in a single proof of finite length. (p. 145)

This remark echoes Russell (1919, p. 21) in a similar context: "It is this 'and so on' that we must replace by something less vague and indefinite."

Perhaps this feeling of "and so on..." is what was bothering our colleague. If there is no more to a proof by induction than a sort of promissory note that we can keep going, then perhaps such a proof isn't much different from ordinary empirical induction. Just as we can't hope to enumerate all cases that fall under an inductive generalization in science, we can't hope to cover all natural numbers to which a mathematical generalization applies. Mathematical induction may only be able to give us a boost in confidence that the generalization holds in all cases, not an iron-clad proof.

Stewart and Tall (1977), however, offer a solution to show that mathematical induction is a rigorously deductive technique after all. "The way out of this dilemma is to remove the 'and so on...' part from the proof and place it squarely in the actual definition of the natural numbers..." (p. 145).

What follows in Stewart and Tall is an exposition of a variation on the Dedekind-Peano axioms for the natural numbers, with the following induction axiom:

(IAx) If $\boldsymbol{S} \subseteq \mathbb{N}_0$ is such that $0 \in \boldsymbol{S}$; and $n \in \boldsymbol{S} \Rightarrow s(n) \in \boldsymbol{S}$ for all $n \in \mathbb{N}_0$, then $\boldsymbol{S} = \mathbb{N}_0$,

where $\boldsymbol{S}$ is a set, $\mathbb{N}_0$ are the natural numbers, and $s(\cdot)$ is the successor function (i.e., the function that yields $n + 1$ for any natural number n). For example, in the sample proof we gave earlier, $\boldsymbol{S}$ corresponds to the set of all natural numbers n such that the sum of 0 through n is ½ n (n + 1), and the two parts of the proof correspond to the two clauses in (IAx). We first showed that 0 is in this set ($0 \in \boldsymbol{S}$) and then that if n is in the set so is $n + 1$ ($n \in \boldsymbol{S} \Rightarrow s(n) \in \boldsymbol{S}$). The conclusion is therefore that the sum formula is true for all natural numbers ($\boldsymbol{S} = \mathbb{N}_0$). Since (IAx) is an axiom, this conclusion now has a kind of deductive legitimacy.[2]

Worry Number 1: The Justification of (IAx)

Stewart and Tall are right that (IAx) speaks to one sort of doubt about math induction. "And so on..." complaints about our proof of the sum formula could center around the fact that what's been demonstrated is a universal statement of the form: For all natural numbers n, if n satisfies the sum formula so does n + 1. This generalization, together with the fact that 0 satisfies the formula, allows you to conclude by a modus ponens inference that 1 satisfies the formula:

(∀n) Sum-formula(n) ⊃ Sum-formula(n + 1).

Sum-formula(0).

Sum-formula(1).

This conclusion in turn allows you to make another modus ponens inference to 2:

$(\forall n)$ Sum-formula(n) $\supset$ Sum-formula(n + 1).

Sum-formula(1).

Sum-formula(2).

And so on. In general, this *modus ponens strategy* requires *k* modus ponenses to conclude that *k* satisfies the sum formula, and it will be impossible to reach the conclusion for all natural numbers in a finite number of steps. (IAx) sidesteps this worry by reaching the conclusion for all *n* in one step. However, putting the "and so on..." idea "squarely in the actual definition of the natural numbers," as Stewart and Tall suggest, isn't a very comforting solution if this means hiding the "and so on..." pea under another shell. What justifies such an axiom?

There are several ways to proceed. First, we could try to eliminate the "and so on..." doubt by showing that (IAx) is inherently part of people's thinking about the natural numbers and, in this sense, is "squarely in" their definition. According to this idea, any kind of intuitive understanding of these numbers already allows you to conclude in one go that if 0 is in ***S*** and if *k* + 1 is in ***S*** whenever *k* is, then all natural numbers are in ***S***. This is close to Poincaré's (1902/1982) view. Although he acknowledged that there is "a striking analogy with the usual procedures of [empirical] induction," he believed that mathematical induction "is the veritable type of the synthetic *a priori* judgment." By contrast with empirical induction, mathematical induction "imposes itself necessarily, because it is only the affirmation of a property of the mind itself" (pp. 39-40). Thus, any attempt to found math induction on a formal axiom like (IAx) can only replace an idea that is already intuitively clear with one that is more obscure. Goldfarb (1988) points out that the psychological nature of math induction was irrelevant to the goals of those like Frege, Hilbert, and Russell, who were the targets of Poincaré's critique (see Hallett, 1990, for debate about these goals). But perhaps Poincaré's theory is sufficient to settle specifically psychological doubts. Another possibility, one more consistent with Frege's (1884/1974) and Russell's (1919) viewpoint, is that (IAx) is the result of a principle that applies in a maximally domain-general way and is thus a precondition of rationality. We'll look more closely at the source of this generality in just a moment in examining Frege's treatment of (IAx).

We favor a view in which math induction is inherent in the concept of the natural numbers—whether this is specific to the natural-number concept or is inherited from more general principles—and this view preserves a crisp distinction between deductive and inductive reasoning. But can we really rule out the possibility that the "and so on..." doubt is a symptom of a true intermediate case? Maybe (IAx) merely plasters over a gap like that in ordinary induction—the "and so on..." gap that is left over from the modus ponens strategy.

One thing that seems clear is that in actually applying math induction—for example, in our use of it to prove the sum formula—the modus ponens strategy is never performed. In our original proof, for instance, we didn't (explicitly or implicitly) grind through the first few steps of the strategy (from 0 to 1 to 2...) before concluding that the sum formula is true for all *n*. So the way in which the modus ponens strategy gives us reason to believe in math induction must be indirect. Perhaps by reflecting on the modus ponens strategy, we somehow come to see that this strategy is unnecessary and that we can use the short-cut math induction procedure in its place. But if this is true, it can't be the *success* of the modus ponens strategy that convinces us, since this strategy is an infinite process that is never completed. We don't learn math induction from the modus ponens strategy in the usual generalizing or transferring sort of way. Instead, it must be our realizing that the modus ponens strategy could *potentially* yield the right result under some idealized conditions (e.g., infinite time and other resources) that persuades us that math induction will work. If we push further and ask what's responsible for such a realization, however, the answer seems to lead right back to our understanding of the structure of the numbers themselves. At best, the modus ponens strategy plays a mediating role in reminding us of a property of the natural number system but provides no independent justification.

Why is (IAx) inherent in the number concept? One way to see this comes from Frege's (1984/1974) treatment of math induction. (The modernized version that follows is from Quine, 1950, chap. 47. A somewhat similar attempt appears in Dedekind, 1888/1963, especially Theorems 59 and 89.) We can think of the natural numbers as a set that contains 0 and that contains all immediate successors of each of its members. Such a set, however, could contain other objects besides numbers. It could also contain, for example, Frege's hat, since Frege's hat does not have a successor. What we need to define the natural numbers is a minimality rule that says that the natural numbers are the *smallest* set containing 0 and containing all immediate successors of its members. Put slightly differently, the natural numbers are the elements that belong to all such sets, including the set containing only 0 and the successors (of the successors of…) 0. This definition of the natural numbers appears in (1):

(1) $n \in \mathbb{N}_0 \leftrightarrow (\forall u)\ ((0 \in u\ \&\ (\forall k)\ (k \in u \supset s(k) \in u)) \supset n \in u)$,

where $\mathbb{N}_0$ is the set of natural numbers and $s(\cdot)$ is the successor function, as before.

From this definition, mathematical induction follows, just as Stewart and Tall (1977) claimed. For suppose $\boldsymbol{S}$ is a set that fulfills the two conditions on (IAx), as in (2):

(2) $0 \in \boldsymbol{S}\ \&\ (\forall k)\ (k \in \boldsymbol{S} \supset s(k) \in \boldsymbol{S})$.

If n is a natural number, (3) then follows from the definition in (1):

(3) $(\forall u)\ ((0 \in u\ \&\ (\forall k)\ (k \in u \supset s(k) \in u)) \supset n \in u)$,

and in particular (i.e., by instantiating u to $\boldsymbol{S}$):

(4) $(0 \in \boldsymbol{S}\ \&\ (\forall k)\ (k \in \boldsymbol{S} \supset s(k) \in \boldsymbol{S})) \supset n \in \boldsymbol{S}$.

It then follows by modus ponens from (2) and (4) that $n \in \boldsymbol{S}$. Hence, every natural number is an element of $\boldsymbol{S}$.

Frege's method helps to bring out what's important about (IAx) as a part of the natural-number concept. In using (IAx) to prove a theorem, as we did with the sum formula, we take an "insider's view" of the number system in which its familiar properties are already in place. For these purposes, (IAx) is the usual tool we pick up in high school algebra. But to see the role that (IAx) plays in the definition of the natural numbers—the reason that it appears among the Dedekind-Peano axioms—we need to adopt an "outsider's view." From this perspective, (IAx) is a kind of closure principle. It says, in effect, that nothing can be a natural number that isn't either 0 or a successor of (a successor of…) 0. This is the importance of the minimality constraint we've just looked at. For these purposes, we could substitute for (IAx) other rules that have equivalent effects, given the rest of the axioms— rules that may seem less artificial and less similar to empirical induction. An example is the Least-Number Principle (see Kaye, 1991) that we can spell out in a way that can be compared to (IAx):

(LNP) If $\boldsymbol{S} \subseteq \mathbb{N}_0$ and $n \in \boldsymbol{S}$, then there is an $m \in \boldsymbol{S}$ such that for all $k < m$, $k \notin \boldsymbol{S}$.

In other words, if $\boldsymbol{S}$ is a nonempty set of the natural numbers, there's a smallest number in $\boldsymbol{S}$. For other number systems, other types of math induction may be required (see, e.g., Knuth, 1974).

Worry Number 2: The Justification of (UG)

But there could be another hang up about proofs like the example we gave for the sum formula. As part of the proof of this formula, we showed that for any particular k, if k satisfies the sum formula then so does $k + 1$. The second sort of doubt is whether this is enough to prove that the same is true for *all* natural numbers. Clearly, (IAx) is completely powerless to quell this second kind of doubt. The axiom requires us to show that $n \in \boldsymbol{S}$ entails $s(n) \in \boldsymbol{S}$ for *all* natural numbers n, and this is precisely where the second sort of doubt arises.

This kind of doubt surfaces in some psychological treatments of math induction. For example, in one of the very few empirical studies on this topic, Smith (2002, p. 6) cites a definition similar to (IAx) and then goes on to remark, "The definition secures a generalization (universalization). The reference in the premises is to a property true of *any* (some particular) number. But the reference in the conclusion is to *any* number whatsoever, any number at all. Note that the term *any* is ambiguous here since it covers an inference from some-to-all." The word *any* does not, in fact, appear in the definition that Smith cites. The relevant clause is "...and if it is established that [the property] is true of $n + 1$ provided it is true of n, it will be true of all whole numbers." But what Smith seems to mean is that the proof is a demonstration for some particular case, k and $k + 1$, whereas the conclusion is that the property is true for all natural numbers n. Similarly, in discussing earlier work by Inhelder and Piaget, Smith states:

> [mathematical induction] runs from what is shown to be true of a specified number n (*n'importe quel nombre*) and what in that case is also shown to be true of its successor ($n + 1$) to a conclusion about any number whatsoever (*nombre quelconque*). It is just such an inference which is at issue in mathematical induction and it was Inhelder and Piaget's (1963) contention that their evidence showed that children were capable of this type of reasoning... (p. 23)

The inference that Smith is pointing to, however, is not a feature of math induction per se but of the universal generalization that feeds it. Universal generalization, roughly speaking, is the principle that if a predicate can be proved true for an arbitrary member of a domain, it must be true for all members. A formal statement of this rule can be found in most textbooks on predicate logic. We have adapted slightly the following definition from Lemmon (1965):

> (UG) Let $P(e)$ be a well-formed formula containing the arbitrary name e, and x be a variable not occurring in $P(e)$; let $P(x)$ be the propositional function in x which results from replacing all and only occurrences of e in $P(e)$ by x. Then, given $P(e)$, UG permits us to draw the conclusion $(\forall x)P(x)$, provided that e occurs in no assumption on which $P(e)$ rests.

For example, when we proved that the sum formula holds for $k + 1$ if it holds for k, we used no special properties of k except for those common to all natural numbers. In this sense, k functioned as an arbitrary name. Hence (by universal generalization), for all natural numbers n, the sum formula holds of $n + 1$ if it holds for *n*. This generalization is needed to satisfy the second condition of (IAx), but it is also needed to supply the major premises for the modus ponens strategy and, of course, for much other mathematical reasoning. The confusion between (IAx) and (UG) is an understandable one. We could rephrase (IAx) in a way that refers to an arbitrary case: If ***S*** is a subset of the natural numbers, if 0 is an element of ***S***, and if whenever an arbitrary natural number *k* is in ***S*** so is *k* + 1, then ***S*** includes all natural numbers. This rephrasing is valid, but it combines (IAx) and (UG) in a way that obscures their separate contributions.

(UG), like (IAx), may seem to lend deductive legitimacy to what looks at first like empirical induction. Unlike the case of (IAx), though, there's not much temptation to see this rule as sanctioning a kind of induction-by-enumeration. Informal use of mathematical induction seemed to fall short because it required an infinite number of modus ponens inferences, and (IAx) was introduced to avoid this. But the informal use of universal generalization in mathematics doesn't have the same kind of "and so on..." gap. Once we've shown that *P* is true for an arbitrary element *e*, we needn't apply this result to get *P*(0), *P*(1), *P*(2), and so on. In general, informal universal generalization (e.g., in the proof of the sum formula) doesn't raise the same kind of sophisticated issues about the nature of proof that informal use of math induction does (although it does take some sophistication to formulate (UG) correctly for a logic system, as the definition from Lemmon, 1965, suggests).

Instead, the worry about (UG) stems from its similarity to an inductive heuristic in which you generalize from a specific number. (Kahneman & Tversky's, 1972, representativeness heuristic is a related example in the context of probability estimation.) In the case of the sum formula, for instance, it's clearly not mathematically kosher to proceed like this:

Let's pick an arbitrary number, say, 7.

Simple calculations show that 0 + 1 + 2 + 3 + 4 + 5 + 6 + 7 = 28 and ½*7*(7 + 1) = 28.

Hence, $\sum i = ½\, n\, (n + 1)$.

An example of this sort may increase your confidence that the sum formula is true, but it doesn't produce a deductively valid proof. However, there's an intuitive kinship between this pseudo-proof and the earlier correct proof of the sum formula. The phrase *arbitrary number* captures this kinship, since it can mean either a specific number that we select in some haphazard way or it can mean a placeholder that could potentially take any number as a value. Sliding between these two meanings yields an inductive, example-based argument, on one hand, and a deductively valid (UG)-based proof, on the other. For the sum formula, it's the difference between the argument from seven that we just gave and the correct argument with *k*.

(UG) is pervasive in mathematics, in part because most algebraic manipulations depend on it. Variables like *k* turn up everywhere, not just in the context of mathematical induction. Thus, confusion about the nature of variables is apt to have wide-spread consequences and blur the distinction between inductive arguments and deductive proof. We consider the empirical evidence for this confusion in the next section, including actual pseudo-proofs from students that are quite similar to our argument from seven. We then return to the issue of whether this evidence shows that the difference between deduction and induction in math is merely one of degree.

Induction in Mathematics

In Smith's (2002) study of children's arithmetic, the key question to participants was: "If you add any number at all [of toys] to one pot and the same number to the other, would there be the same in each, or would there be more in one than in the other?" A correct answer to this question seems to depend on the children realizing that for all numbers *x* and quantities *m* and *n*, $x + m = x + n$ if and only if *m* = *n*. In the experiment, *m* and *n* are the initial quantities of toys in the pots (*m* = *n* = 0 in one condition of the study, and *m* = 1 and *n* = 0 in another). It is certainly possible to prove this principle using (IAx), but it is far from clear that this is the way children were approaching the problem. Reasoning along the lines of (UG) would mean going from "for arbitrary *k*, (*m* + *k* = *n* + *k*) if and only if (*m* = *n*)" to the above conclusion for all *x*. But again it is unclear that this is how children went about answering this question. Among the children's justifications that Smith quotes on this problem, the only one that directly suggests (UG) is "I'm putting like a number in the orange pot and a number in the green pot, and it's got to be the same."

There is plenty of reason, however, to suspect that children in this experiment were using specific examples to justify their decisions. Some simply cited a specific number, for example, "If you added a million in there, there would be a million, and if you added a million in there, there would be a million," or "it would be just a million and six in there and a million and six in there." Others gave a more complete justification but also based on a specific number of items: "If we started putting 0 in there and 1 in there, and then adding 5 and 5, and so it wouldn't be the same—this would be 5 and 6." Instead of the (UG) strategy of thinking about an arbitrary addition, these children seem to generalize directly from a particular number, such as 5 or 1,000,006, to all numbers. The numbers may be selected arbitrarily but they aren't capable of denoting any natural number in the way (UG) requires. It is possible, of course, that these children were using a correct general strategy and simply citing specific cases to aid in their explanations, but the experiments we are about to review suggest than even college students evaluate "proof by specific examples" as a valid method.

Variations on this kind of instance-based strategy are probably common in mathematics. One traditional way of thinking about this is that people use these strategies as heuristics to suggest the right generalization before they apply more rigorous mathematical methods to produce a genuine proof. This is similar to the distinction from Polya (1954) that we mentioned earlier (see, also, Thagard's chapter on abduction in this volume). There might be a "context of discovery" in which such informal methods are appropriate, and a "context of justification" where only proof-based methods such as (IAx) and (UG) apply. We were once given the advice that you should only try to prove a theorem when you were 95% sure it was right. This advice separates neatly the informal, confidence-based aspect of mathematics from its formal, proof-based side. So does the usual classroom advice to experiment with different values to find a formula before you try to prove it with (IAx).

But whatever validity the discovery/justification division has in mathematics, it is clear that students do not always honor it. Even when asked to prove a proposition (so that justification is required), students often produce what look to be instance-based strategies. We give some examples here from experiments on college students and then try to draw some implications about the relation between induction and deduction in math contexts.

Algebra

Asking college student to prove simple propositions in number theory often results in "proof by multiple examples," such as the following one (Eliaser, 2000):

> Show that every odd multiple of 5 when squared is not divisible by 2.
>
> Proof:
>
> > Odd multiples of 5: 5, 15, 25...
> >
> > $5^2 = 25 \rightarrow$ not divisible by 2.
> >
> > $15^2 = 225 \rightarrow$ not divisible by 2.
> >
> > Therefore shown.

In the experiment from which we took this excerpt, Northwestern University undergraduates had to prove a group of simple statements, half of which had a universal form, like the one above, requiring an algebraic proof. Although nearly all these students were able to use algebra (as they demonstrated in a post test), only 41% initially employed it in proving universal propositions. Instead, 33% used multiple examples (as in the sample "proof") and 17% used single examples.

Of course, students may not always see the right proof strategy immediately and may fall back on examples instead. For instance, the students may not remember or may not know how to set up a problem that requires representing the parity of a number (i.e., that all even numbers have the form $2x$ and all odds ones $2x + 1$). But this doesn't seem to be the whole story behind the students' dismal performance. In a second study, Eliaser (2000) gave potential proofs to a new group of undergraduates and asked them to rate the quality of the proof on a four-point scale from "shows no support" to "completely proves." The "proofs" included single-example, multiple-example, and algebraic versions. But despite the fact that students didn't have to produce their own proofs and only had to evaluate the proofs of others, they nevertheless rated multiple-example pseudo-proofs higher than either single-example or (correct) algebraic proofs, for universal claims, such as the one about odd multiples of 5.

Depressing performance of a similar kind comes from a study of students who were training to be elementary math teachers (Martin & Harel, 1989). These students rated proofs and pseudo-proofs for two number-theoretic statements (e.g., If *a* divides *b* and *b* divides *c*, then *a* divides *c*). The pseudo-proofs included a typical example (e.g., "12 divides 36, 36 divides 360, 12 divides 360") and an example with a big, random-looking number:

> Let's pick any three numbers, taking care that the first divides the second, and the second divides the third: 49 divides 98, and 98 divides 1176. Does 49 divide 1176? (*Computation shown to left.*) The answer is yes.

The pseudo-proofs also included an instantiation of the more general, correct proof:

> Take 4, 8, and 24. 4 divides 8, which means that there must exist a number, in this case 2, such that 2 x 4 = 8. 8 divides 24, which means there must exist a number, in this case 3, such that 3 x 8 = 24. Now substitute for 8 in the previous equation, and we

get 3 x (2 x 4) = 24. So we found a number (3 x 2), such that (3 x 2) x 4 = 24.

Therefore, 4 divides 24.

Students rated these pseudo-proofs, along with other pseudo- and correct proofs on a scale from 1 (not considered a proof) to 4 (a mathematical proof). If we take a rating of 4 to be an endorsement of the argument as a proof, then 28% endorsed the correct proof, 22% the instantiated pseudo-proof, 39% the specific example, and 38% the big random-looking example.

Geometry

Students' use of example-based proof strategies isn't limited to algebra. It's a common classroom warning in high school geometry that students shouldn't draw general conclusions from the specific cases that they see in the diagrams accompanying a problem. Diagrams represent one way in which the givens of a problem could be true, not all possible ways. Despite these warnings, students do generate answers to test questions based on single diagrams.

Evidence on this point comes from a study by Koedinger and Anderson (1991). This study presented students with yes/no questions about geometric relationships, accompanied by one of two diagrams. For example, one problem asked, "If $\angle BCD = \angle ACD$, must $\overline{AB} \perp \overline{CD}$?" The two possible diagrams for this problem appear as Figures 1a and 1b. For half the problems the correct answer was "yes," as in this example. The remaining problems had "no" as the correct answer, for instance, "If $\overline{AC} = \overline{CB}$, must $\overline{AB} \perp \overline{CD}$?" The two possible diagrams for this problem are Figures 1c and 1d. For each of these two problems, the given information and the conclusion were true of one diagram (Figures 1a and 1c) and were false of the other (Figures 1b and 1d). Individual participants saw only one of the possible diagrams per problem.

Please insert Figure 1 about here.

The results showed that students were usually able to get the right answer: They were correct on about 68% of trials. For problems where the correct answer was "yes" (so that a proof was possible), the students performed more accurately when the corresponding proof was easy than when it was difficult, suggesting that they were attempting such a proof. Nevertheless, they tended to be swayed by the diagrams: They made more "yes" responses in all conditions when the diagram pictured the given information and the conclusion (as in Figures 1a and 1c) than when it didn't (Figures 1b and 1d). Koedinger and Anderson take this pattern of results to mean that the students were attempting to find a proof, but were using the diagram as a backup inductive strategy if no proof was forthcoming (a "modified misinterpreted necessity model" in the jargon of reasoning research; see Evans, Newstead, & Byrne, 1993).

Further evidence that students rely on diagrams for inductive support comes from one of our own studies of non-Euclidean geometry. In this experiment, we gave participants a short introduction to hyperbolic geometry, which differs from Euclidean geometry by the following *hyperbolic axiom*: For any given line ℓ and point *p* not on ℓ, there are multiple lines through *p* parallel to ℓ. The participants then solved a number of problems in the new geometry while thinking aloud. Figure 2 illustrates one such problem. To avoid confusing participants about the meaning of *parallel*, we substituted the nonsense term *cordian*, which we defined according to the hyperbolic axiom. For the problem in Figure 2, we told participants that lines ℓ and *m* are cordian, that $\angle$BAC and $\angle$BDC were 90° angles, and asked them to decide whether triangle ABC was congruent to triangle DCB.[3]

Please insert Figure 2 about here.

Diagrams for non-Euclidean problems are apt to be misleading. Although there are several conventions for drawing such diagrams (such as the one adopted in Figure 2; see Greenberg, 1993), they can nevertheless suggest Euclidean relationships that don't hold in the non-Euclidean domain.

Participants quickly recognize this difficulty, but they still fall back on the appearance of a diagram. The following transcript is from a participant working on the problem in Figure 2:

> Must be true, ABC is congruent to DCB. If ∠BAC is 90, must they be congruent? It does not mention a line will diverge, so how can I judge how far apart these lines will be at any given point, since they're curving? Angle ABD is split, maybe in half, if you dissect it, it appears as though it's dissected... I have no information about the angles ACD or ABD. I must assume that there can be less than 360° [in a quadrilateral]. That the angle B could be equal to angle C. Angle B equals angle ACD. Therefore, if those angles are equal, and the line joining them is drawn, angle ABC is equal to angle DBC. Okay. The dotted line will be line *q*. Line segment *q* bisects angle ABD.... Angle, angle, angle...theorem states that ABD is congruent with—ABC is congruent with DCB. Must be true? Yes.

Although this participant seems initially unsure of the angle relationships, he or she is gradually seduced by the appearance that ∠ABD and ∠ACD are equal and that the dashed line bisects both these angles.

Implications

These examples show that people sometimes use inductive strategies where deductive methods are necessary. More surprisingly, people sometimes evaluate the inductive methods more favorably than the correct deductive ones. We can explain the first of these findings on the assumption that people resort to inductive methods when they run out of deductive ones—either because no proof is possible or because they can't find one—as Koedinger and Anderson (1991) propose. But the evaluation results make it likely that this isn't the whole story. Why do people find inductive pseudo-proofs better than true proofs when they're able to assess them directly?

The answer may lie in the relative complexity of the true proofs versus the inductive lures. Examples are likely to be easier to understand than proofs, especially if the proofs are lengthy. Similarly, there's a difference in concreteness between proofs with individual numbers and those with arbitrary names (i.e., mathematical variables) in algebra. Geometry diagrams also have a concreteness advantage over the proofs that refer to them. Even among correct deductive proofs, we evaluate the simpler, more elegant ones as superior to their more complex competitors, and it is possible that students' incorrect evaluations are a more extreme version of the same tendency.

Explanations along these lines do no harm to the thesis that deductive and inductive reasoning make use of qualitatively different mechanisms. These explanations take the induction/deduction difference for granted and account for students' errors on the basis of factors, such as complexity or abstractness, that swamp deductive methods or make them seem less attractive. We know of no experimental results that can eliminate these explanations. But it's also interesting to consider the possibility of a more continuous gradient between inductive and deductive strategies. One proposal along these lines might be based on the idea that people search for increasingly nonobvious counterexamples to an argument or theorem. If they are unable to find any counterexamples, either because there are none or because the individuals can no longer hold enough examples in working memory, they declare the argument valid. In the case of algebra, we can imagine students first choosing a specific number that the problem suggests and that meets the theorem's givens. For example, in the problem mentioned earlier of showing that the squares of odd multiples of 5 are not divisible by 2, the participant began by verifying that 5^2 is not divisible by 2. If the first instance does not yield a counterexample, the students try another (e.g., $(3*5)^2$) until they either discover a counterexample or run out of working-memory resources. In the case of geometry, students might first verify that the conclusion is true of the diagram that appears with the problem. They may then consider new diagrams of their own in the same search for counterexamples.

In the present context, one difficulty with this line of thinking, however, is that it doesn't account for Koedinger and Anderson's (1991) evidence that students also attempt to apply a separate deductive strategy and resort to diagrams only when the first strategy fails. The counterexample hypothesis predicts instead that students should begin by considering the given diagram, look for other diagrams that might falsify the conclusion, and end by accepting the theorem as valid if it correctly applies to all the diagrams they have examined. In addition, there is a theoretical problem. In the case of a universal proposition—such as the sum formula that applies to all natural numbers—there's no endpoint at which you can stop and be sure that no counterexamples will be found. There will always be additional numbers that have features other than the ones you've so far considered. Rather than provide a continuum between inductive and deductive reasoning, this strategy never provides more than weak inductive support. Of course, the fact that a pure counterexample strategy is inadequate as an explanation of the data doesn't mean that there aren't other methods intermediate between induction and deduction that may be more reasonable. But we know of no convincing cases of this type.

Another way to put this point is to say that there is no number or diagram that is "completely arbitrary" in a way that could prove a universal proposition. "Each number has its own peculiarities. To what extent a given particular number can represent all others, and at what point its own special character comes into play, cannot be laid down generally in advance" (Frege, 1884/1974, p. 20). The sort of arbitrariness that does prove such a proposition—and that (UG) demands—is a matter of abstractness, not randomness or atypicality. Arbitrariness in the latter sense drives you to more obscure cases as potential counterexamples. But the sort of arbitrariness that is relevant to proving a universal proposition pulls in the opposite direction of inclusiveness—an instance that could potentially be any number.

Summary and Conclusion

We now have a more complicated answer to give to our colleague. Our original response was that mathematical induction is a deductive technique with a confusing name. When people use math induction to prove a theorem, they aren't using induction and they aren't sacrificing deductive validity; the theorem is still necessarily true, just as with other standard proof techniques. Although there is debate about the role of probabilistic proofs in mathematics (see Fallis, 1997)—methods that truly are inductive—math induction isn't such a method.

But while we still think this is correct, math induction does loosely resemble empirical induction. From a formal point of view, there is a tension between math induction and the usual constraint that proofs be finite. Math induction seems to rely on an infinite process of iterating through the natural numbers—the modus ponens strategy that we discussed earlier. From a psychological point of view, of course, no such infinite process is remotely possible. It's this gap that creates the similarity with empirical induction, since in both cases the conclusion of the argument seems beyond what the premises afford. An induction axiom like (IAx) makes up for this gap in formal contexts, but why is math induction a deductively correct proof tool if it can't be backed up by the infinite iteration it seems to require?

We've argued elsewhere (Rips, Bloomfield, & Asmuth, 2005) that math induction is central to knowledge of mathematics: It seems unlikely that people could have correct concepts of number and other key math objects without a grip on induction. Basic rules of arithmetic, such as the associative and commutative laws of addition and multiplication, are naturally proved via math induction. If this is right, then it's a crucial issue how people come to terms with it. We've tried to argue here that the analogy we've just noted—enumeration : conclusion by empirical induction :: modus ponens strategy : conclusion by math induction—is misleading. Math induction doesn't get its backing from the modus ponens strategy. Rather the modus ponens strategy plays at most a propaedeutic role, revealing an abstract property of the natural number system that is responsible for the correctness of math induction. It is still possible, of course, that math induction picks up quasi-inductive support from its fruitfulness in proving further theorems, but in this respect it doesn't differ from any other math axiom.

Although we think that math induction doesn't threaten the distinction between deductive and inductive reasoning, there is a related issue about generalization in math that might. Math proofs often proceed by selecting an "arbitrary instance" from a domain, showing that some property is true of this instance, and then generalizing to all the domain's members. For this universal generalization to work, the instance in question must be an abstraction or stand-in (an "arbitrary name" or variable) for all relevant individuals, and there is no real concern that such a strategy is not properly deductive. However, there's psychological evidence that students don't always recognize the difference between such an abstraction and an arbitrarily selected exemplar. Sometimes, in fact, students use exemplars in their proofs (and evaluate positively proofs that contain exemplars) that don't even look arbitrary but are simply convenient, perhaps because the exemplars lend themselves to concrete arguments that are easy to understand. In these cases, students are using an inductive strategy, since the exemplar can at most increase their confidence in the to-be-proved proposition.

It's no news, of course, that people make mistakes in math. And it's also no news that ordinary induction has a role to play in math, especially in the context of discovery. The question here is whether these inductive intrusions provide evidence that the deduction/induction split is psychologically untenable. Does the use of arbitrary and not-so-arbitrary instances show that people have a single type of reasoning mechanism that delivers conclusions that are quantitatively stronger or weaker, but not qualitatively inductive versus deductive? We've considered one way in which this might be the case. Perhaps people look for counterexamples, continuing their search through increasingly arbitrary (i.e., haphazard or atypical) cases until they've found such a counterexample or have run out of steam. The longer the search, the more secure the conclusion. We've seen, however, that this procedure doesn't extend to deductively valid proofs; no matter how obscure the instance, it will still have an infinite number of properties that prohibit you from generalizing from it. It is possible to contend that this procedure is nevertheless all that people have at their disposal—that they can never ascend from their search for examples and counterexamples to a deductively adequate method. But although the evidence on proof evaluation paints a fairly bleak picture of students' ability to recognize genuine math proofs, the existence of such proofs shows they are not completely out of reach.

References

Dedekind, R. (1963). The nature and meaning of numbers (W. W. Beman, Trans.). In *Essays on the theory of numbers* (pp. 31-115). New York: Dover. (Original work published 1888)

Eliaser, N. M. (2000). What constitutes a mathematical proof? *Dissertation Abstracts International, 60 (12),* 6390B. (UMI No. AAT 9953274)

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning*. Hillsdale, NJ: Erlbaum.

Fallis, D. (1997). The epistemic status of probabilistic proofs. *Journal of Philosophy, 94*, 165-186.

Frege, G. (1974). *The foundations of arithmetic* (J. L. Austin, Trans.). Oxford, England: Blackwell. (Original work published 1884)

Goldfarb, W. (1988). Poincaré against the logicists. In W. Aspray & P. Kitcher (Eds.), *History and philosophy of modern mathematics* (pp. 61-81). Minneapolis: University of Minnesota Press.

Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *NeuroReport, 8,* 1305-1310.

Grattan-Guinness, I. (1997). *The Norton history of the mathematical sciences*. New York: Norton.

Greenberg, M. J. (1993). *Euclidean and non-Euclidean geometries* (3rd ed.). New York: Freeman.

Hallett, M. (1990). Review of *History and Philosophy of Modern Mathematics*. *Journal of Symbolic Logic*, *55,* 1315-1319.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3,* 430-454.

Kaye, R. (1991). *Models of Peano arithmetic.* Oxford, England: Oxford University Press.

Kline, M. (1972). *Mathematical thought from ancient to modern times*. Oxford, England: Oxford University Press.

Knuth, D. E. (1974). *Surreal numbers.* Reading, MA: Addison-Wesley.

Koedinger, K. R., & Anderson, J. R. (1991). Interaction of deductive and inductive reasoning strategies in geometry novices. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, 780-784.

Lemmon, E. J. (1965). *Beginning logic*. London: Nelson.

Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education, 20,* 41-51.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., & Fazio, F. (1998). Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia*, *36*, 369-376.

Poincaré, H. (1982). *Science and hypothesis* (G. B. Halsted, Trans.). In *The Foundations of Science*. Washington, D. C.: University Press of America. (Original work published 1902)

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

Quine, W. V. (1950). *Methods of logic* (4th ed.). Cambridge, MA: Harvard University Press.

Rips, L. J. (2001a). Two kinds of reasoning. *Psychological Science*, *12,* 129-134.

Rips, L. J. (2001b). Reasoning imperialism. In R. Elio (Ed.), *Common sense, reasoning, and rationality* (pp. 215-235). Oxford, England: Oxford University Press.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2005). *The psychology of mathematical objects: Current research and theory*. Manuscript submitted for publication.

Ruben, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting. *Psychological Review*, *103*, 734-760.

Russell, B. (1919). *Introduction to mathematical philosophy.* New York: Dover.

Smith, L. (2002). *Reasoning by mathematical induction in children's arithmetic.* Amsterdam: Pergamon.

Stewart, I., & Tall, D. (1977). *The foundations of mathematics*. Oxford, England: Oxford University Press.

Footnotes

1. This paper was written while the first author was a Fulbright fellow at the Catholic University of Leuven, Belgium. We thank the Fulbright Foundation and the University for their support. We also thank Stella Christie, Mike Oaksford, Claes Strannegård, and the editors of this volume for comments and discussion of an earlier draft. Please send correspondence about this paper to Lance Rips, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208. Email: rips@northwestern.edu.

2. Use of math induction as a proof technique was, of course, well-established before the introduction of (IAx) as a formal axiom. Medieval Arabic and Hebrew sources use implicit versions of math induction to prove theorems for finite series and for combinations and permutations (Grattan-Guinness, 1997). According to Kline (1972, p. 272), "the method was recognized explicitly by Maurolycus in his *Arithmetica* of 1575 and was used by him to prove, for example, that $1 + 3 + 5 + ... + (2n - 1) = n^2$."

3. If the two triangles ABC and DCB are congruent, corresponding angles of these triangles must also be congruent. In Figure 2, ∠BAC and ∠CDB are given as congruent, but what about the pair ∠DCB and ∠ABC and the pair ∠BCA and ∠CBD? In Euclidean geometry, the congruence of each of these pairs follows from the fact that alternate interior angles formed by a transversal and two parallel lines are congruent. In hyperbolic geometry, however, the alternate interior angles formed by a transversal and two parallel ("cordian") lines are not congruent. For example, in Figure 2, ∠ABC and ∠DCB are both the alternate interior angles formed by the cut of transversal BC across parallel lines ℓ and *m*. But because these corresponding angles cannot be proved congruent, neither can the two triangles.

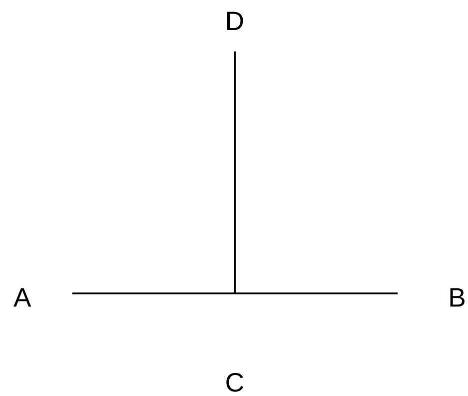
Figures

Figure 1. Stimulus problems from a study of geometry problem solving (adapted from Koedinger & Anderson, 1991, Figure 2).

Figure 2. Stimulus problem for a study of non-Euclidean geometry.

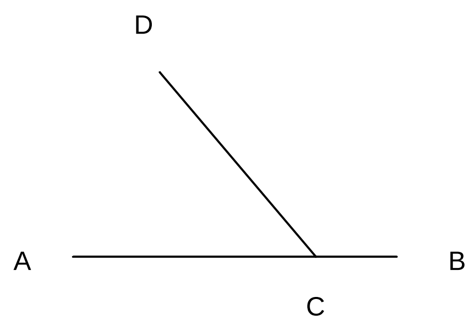
a.

If ∠BCD =∠ACD,
must AB⊥ CD?



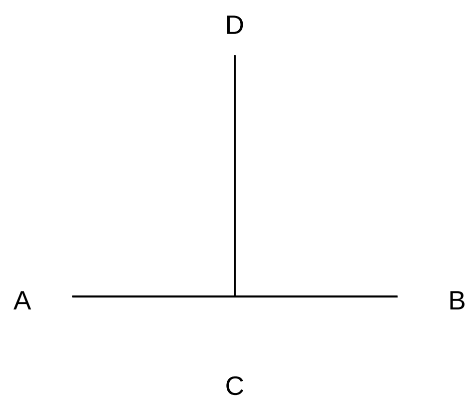
b.

If ∠BCD =∠ACD,
must AB⊥ CD?



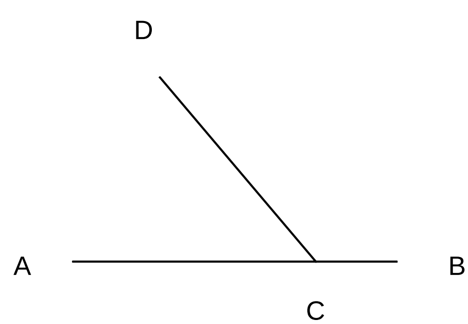
c.

If AC = CB ,
must AB⊥ CD?



d.

If AC = CB ,
must AB⊥ CD?



Given: Line ℓ cordian [parallel] to line m.

∠BAC and ∠BDC are 90° angles.

Must the following be true?

Triangle ABC is congruent to triangle DCB.