

# What inductive explanations could not be

John Dougherty

the date of receipt and acceptance should be inserted later

**Abstract** Marc Lange argues that proofs by mathematical induction are generally not explanatory because inductive explanation is irreparably circular. He supports this circularity claim by presenting two putative inductive explanantia that are one another's explananda. On pain of circularity, at most one of this pair may be a true explanation. But because there are no relevant differences between the two explanantia on offer, neither has the explanatory high ground. Thus, neither is an explanation. I argue that there is no important asymmetry between the two cases because they are two presentations of the same explanation. The circularity argument requires a problematic notion of identity of proofs. I argue for a criterion of proof individuation that identifies the two proofs Lange offers. This criterion can be expressed in two equivalent ways: one uses the language of homotopy type theory, and the second assigns algebraic representatives to proofs. Though I will concentrate on one example, a criterion of proof identity has much broader consequences: any investigation into mathematical practice must make use of some proof-individuation principle.

**Keywords** mathematical explanation, proof theory, homotopy type theory, mathematical induction

## 1 Introduction

Lange (2009) argues that proofs by mathematical induction are generally not explanatory on the grounds that inductive explanation is irreparably circular. He supports this circularity claim by presenting two putative inductive explanantia that are one another's explananda. On pain of circularity, at most one of this pair may be a true explanation. But because there are no relevant differences between the two explanantia on offer, neither has the explanatory

high ground. Thus, neither is an explanation. Since any inductive explanatory attempt has this kind of evil twin, no inductive proof is explanatory. The claimed virtue of Lange’s argument is that it avoids “making any controversial presuppositions about what mathematical explanations would be” (2009, 203), unlike most other approaches to this question, which rely on intuitions about particular cases and the nature of the natural numbers.

One kind of defense of induction’s explanatory value tries to break the symmetry between the two cases (Baker 2010). Perhaps one of the explanations will be longer, or perhaps one will be more complex, disjunctive, unnatural, or whatever. And perhaps one of these differences is relevant when it comes to explanatory power. I do not think that this approach can work here, because there are no differences in explanatory power. Lange’s argument fails for a different reason. There is no important asymmetry between the two cases because they are two presentations of the same explanation. That is, Lange’s argument requires the supposition that the two explanations on offer are different. This supposition is controversial, though the controversy has not, to my knowledge, been raised in the literature on mathematical explanation. In what follows, I aim to drum up this controversy by arguing for a criterion of proof identity that identifies the two proofs Lange offers. Though I will concentrate on one example, a criterion of proof identity has much broader consequences: any investigation into mathematical practice must make use of some proof-individuation principle.

## 2 The need for a theory of proof identity

When presented with the question of what makes a proof explanatory, there is a natural strategy to pursue. Suppose that an explanatory proof is a proof that has some  $X$  factor. Collect a stack of proofs. From these, select the ones that are intuitively explanatory. If we are lucky, there will be some characteristic shared by all and only the explanatory proofs. We can then take this to be the  $X$  factor. Sadly, we are spectacularly unlucky. Intuitions on the explanatory power of particular proofs diverge wildly. On the one hand, inductive proofs are sometimes taken “to be *paradigms of non-explanatory proofs*” (Hafner and Mancosu 2005, 237). On the other, it is sometimes said that induction “more than any other feature, best characterizes the natural numbers,” and that this is why, compared to other approaches, proof by induction “is in many ways better—it is more explanatory” (Brown 1997, 177). This battle of intuitions has reached a standoff.

The great strength of Lange’s argument is that it sidesteps this problem entirely. Leave explanatory virtue unanalyzed. Whatever else an explanatory proof is, it is an explanation, and so it must partake of the same properties as other explanations. In particular, it cannot be circular—no mathematical fact can explain itself. But suppose, says Lange, that some proof by mathematical induction were explanatory. Then it would explain for some predicate  $P$  why it is that for all  $n \in \mathbb{N}$ ,  $P(n)$ . And it would do so by the following method:

**$\mathbb{N}$  induction:** to prove that  $P(n)$  for all  $n \in \mathbb{N}$ , prove:

- $P(0)$
- for all  $n \in \mathbb{N}$ : if  $P(n)$ , then  $P(n + 1)$

But there is another proof method nearby, which would also establish that  $P(n)$  for all  $n \in \mathbb{N}$ :

**$\mathbb{N}$  5-induction:** to prove that  $P(n)$  for all  $n \in \mathbb{N}$ , prove:

- $P(5)$
- for all  $n \geq 5$ , if  $P(n)$  then  $P(n + 1)$
- for all  $0 \leq n < 5$ , if  $P(n + 1)$  then  $P(n)$

While induction on the naturals starts at 0 and goes up, 5-induction starts at 5, then goes both up and down. They are apparently different methods of proof, but none of the differences are explanatorily-relevant, so they are equally explanatory. But they cannot both be explanatory, because this leads to circularity. To see why, note that through standard induction,  $P(0)$  helps to explain why  $P(5)$  holds. Through 5-induction,  $P(5)$  helps to explain why  $P(0)$  holds. Thus, by transitivity of explanation,  $P(0)$  helps to explain why  $P(0)$  holds, an absurdity.

This circularity argument assumes that induction and 5-induction are different proof methods. That is, it tacitly relies on some principle of individuation of mathematical proofs. And perhaps taking these proofs to be different has some initial plausibility. But this first thought weakens under scrutiny. First, syntactic differences don't suffice to distinguish proofs. Simply rearranging the order of the premises doesn't produce a new proof. Nor, for that matter, does changing the order of subproofs, as long as they are logically independent. Just imagine submitting such a reshuffled proof to a mathematics journal as a piece of novel scholarship! These syntactic features are aspects of the presentation, not the proof. Presenting a proof in written form requires writing it down in a certain order, but ordinal facts about this presentation are properties of the proof itself just to the extent that they track inferential relationships.

This is not to say that syntax is the only difference between induction and 5-induction. It is just to say that if there is an important difference between the two proof methods, it must be a difference in inferential structure, not in presentation. Furthermore, it seems that not just any difference in inferential structure will do. Suppose that we have some proof that  $P$  from some set of premises. Our premises stand in some inferential relationship with our conclusion, and these relationships are most likely mediated by other inferential relationships between intermediate conclusions. We can modify this inferential network by appending “if  $P$ , then  $P$ ; therefore  $P$ ” to the end of our proof, introducing one more inferential relationship. This results in a valid proof, and the inferential network is different in some sense, but this modification does not give us a new proof, and certainly not in a sense that's relevant to explanation. And this is just one of many ways that a proof might be trivially “modified” without actually producing a new proof—a similarly empty

instance of modus ponens might be introduced, or a tautology inserted, at any step along the way.

It seems to me that the differences between some proof by induction and its 5-inductive partner are differences of this kind. In both cases we show that  $P$  holds of some small number or numbers, and then we show that  $P$  also holds for an arbitrarily large number because it holds for its predecessor. The difference between a proof by induction and a proof by 5-induction is a matter of where we turn our attention when describing the inferential structure of the proof, not a matter of the inferential structure itself. But as noted above, intuitions on these questions diverge wildly, and someone moved by Lange's argument is not likely to share this one. So this appeal to intuition needs to be backed up with more general theoretical considerations. The next two sections are two ways of spelling out a more systematic story about identity of proofs that agrees with the intuition that induction and 5-induction coincide.

The intuitive picture behind this story proceeds from the observation above that some rearrangements of a proof's presentation are clearly changes in presentation alone, not in the proof itself. Changing the order of the premises, or changing the names of variables, or translating the proof into French does not change the mathematical content. In these cases it's easy to see that such manipulations change nothing about the mathematics, because they make no reference to the mathematical facts involved or the relationships between them. So too, I claim, for the manipulations that turn a proof by induction into a proof by 5-induction, and vice versa. A proof by induction can be turned into a proof by 5-induction via a recipe expressed just in terms of the two proof steps of induction and the three proof steps of 5-induction. This translation can be effected regardless of the conclusion of the proofs or the internal details of the proof steps. And it really is a translation, not a mere correspondence: the justifications of  $P(5)$  in each proof, for example, or the proof that  $P(n) \rightarrow P(n+1)$  for arbitrarily high  $n$ , are the same before and after the translation. This systematicity suggests that the underlying inferential networks are the same, and that an appropriate principle of individuation would—by virtue of correctly capturing the inferential network—make the translation plain.

For a very simple example, consider the statement that every natural number has a number less than or equal to it. We could prove this by induction in two intuitively different ways. We could, for example, prove it by noting that 0 is equal to itself, and  $n+1$  is greater than 0 for all  $n$ . Alternatively, we could prove it by noting that 0 is equal to itself, and  $n+1$  is greater than  $n$  for all  $n$ . Either way, the translation into a proof by 5-induction is the same: the fact that 0 is equal to itself appears in both translations, and the witness to  $P(5)$ —which is 0 in the first proof and 4 in the second—appears in the corresponding proof by 5-induction as the justification for  $P(5)$ , as well. Clearly it doesn't matter to our translation what the specific justification for  $P(5)$  is in a proof by induction; we just take the justification for  $P(5)$ , whatever it may be, and present it as the justification in the base case of our proof. So, intuitively, such a translation turns one description of the justifications for  $P(n)$  and their arrangement into another description of the same array

of justifications. We have some reassurance that this translation hasn't forgotten anything about the inferential structure of the proof: it produces two intuitively different proofs by  $\omega$ -induction from two intuitively different proofs by induction. And indeed, we can translate any proof by  $\omega$ -induction back the other direction, preserving intuitive differences in the same way.

We might make this intuitive story precise in two ways. Either way, the strategy will be to find some mathematical representation of the network of inferential relationships that connects the premises, intermediate conclusions, and conclusion. This allows us to individuate proofs by individuating their mathematical representatives, which is a much more familiar task. On the first approach, induction principles and the proofs they produce are treated syntactically. This is the point of view found in the theory behind automated proof software, which may be used to check and construct mathematical proofs. In this theory, induction principles like induction,  $\omega$ -induction, and more are associated with particular types of objects, and proofs about these objects just are proofs that involve the corresponding induction principle. That is, an object of an inductive type is picked out by its place in a network of inferences that is described by the corresponding induction principle. The theory of these types can provide us with a criterion of identity of types, and we can distinguish or identify induction principles on the basis of their corresponding types. Alternatively, we might try to describe this network more directly. Given an induction principle, we can construct an algebraic object that characterizes the schematic shape of the inferential relationships that the principle describes. For example, any proof by induction must involve some proof that the property of interest holds for 0, as well as a proof that the property holds for  $n+1$  when it holds for  $n$ , and these must be appropriately related. Once we have an algebraic characterization, we can then distinguish proofs by induction by distinguishing their corresponding algebras.

### 3 The syntactic approach

For the syntactic approach I will adopt a version of homotopy type theory (UFP, 2013), including at least higher inductive types and univalence. Homotopy type theory is a recent interpretation of Martin-Löf type theory into abstract homotopy theory. As a type theory, it allows us to implement the broad-strokes syntactic picture of proof identity in the previous paragraph. Indeed, this is the same broad-strokes picture Kreisel (1971, p. 111) draws on the basis of Martin-Löf's work. As a type theory, homotopy type theory is proof relevant. Any mathematical statement has a corresponding type in the theory, and a proof of this statement is a term of that type. The same statement may have more than one proof, and as a result there may be more than one term of the corresponding type. For example, consider the claim "for all  $n \in \mathbb{N}$ , there exists some  $m \in \mathbb{N}$  such that  $m \leq n$ ". Giving a proof of this means deriving a term of type  $\forall n.P(n)$ , where  $P$  is a predicate corresponding to the existential claim. By induction, it suffices to give a term of type  $P(0)$

and one of type  $\forall n.(P(n) \rightarrow P(n+1))$ . Two terms will differ just in case they involve different terms of type  $P(0)$  or different terms for the induction step. For example, we might produce 0 for all  $n$  in the induction step, or we might produce  $n+1$ . Because these are different terms, they produce different terms of type  $\forall n.P(n)$  after applying induction. So if we adopt term identity as a criterion of proof identity, then these proofs are different, in accord with the above plausibility argument. Furthermore, trivial modifications like appending “if  $\forall n.P(n)$ , then  $\forall n.P(n)$ ; therefore  $\forall n.P(n)$ ” will not change the term produced, so this appendix does not give a new proof. Inserting tautological detours along the way will also have no effect.

Proof relevance of this kind allows homotopy type theory to avoid some worries faced by the naive criterion. But it does not completely address the problem of different syntactic presentations of the same proof. For example, there is no way to compare a proof that starts with the base case to a proof that starts with the induction step, even though this ordering doesn’t reflect anything about the inferential structure of the proof. This is where the homotopical side of homotopy type theory steps in. The univalence axiom, introduced by Voevodsky, gives a way to prove that two types are equal, and thus a way to compare their terms (UFP, 2013, §2.10). Without adding the univalence axiom to Martin-Löf type theory, there is in general no way to prove that two types are equal. So if we are going to say that two types give different syntactic presentations of the same mathematical statement, we need univalence.

Roughly, univalence says that to prove that two inductive types are equal, we must show that their induction principles describe the same inferential structure. It must be possible to give a canonical, systematic translation rule between proofs using one induction principle and proofs using the other. This rule must be canonical in the sense that there must be only one possible translation, given the rule. That is, the translator must not be required to make any arbitrary choices. The systematicity requirement means that the translation must be independent of the proofs given as input, for example the base case and induction step. Moreover, it must really be a translation, not just a correlation, in that it must also correlate proper parts of proofs it relates, in a way compatible with the correlations between proofs themselves. Induction principles are meant to describe a network of inferential relations in which the various subproofs stand, and these relations are independent of the internal structure of the subproofs. A systematic translation between proofs by induction and shuffled proofs by induction is easy: we just provide the two subproofs in a different order. So the two types are equal, as we would expect.

This syntactic criterion judges induction and 5-induction to be the same induction principle. This follows straightforwardly from the definition of the natural numbers (UFP, 2013, §1.9), any adequate characterization of 5-induction as a higher inductive type (UFP, 2013, Ch. 6), and univalence.<sup>1</sup> To show that

<sup>1</sup> 5-induction must be given by a higher inductive type because it has  $5 = S^5(0)$  as a generator, as well as equalities that quotient out the higher path structure.

these induction principles produce the same proofs, it suffices to show that they correspond to equal inductive types. To show this, we must give a canonical, systematic translation from one induction principle to the other. So suppose that we have a proof by induction. To give a proof of  $P(5)$ , we can apply the induction step five times to our proof of  $P(0)$ . To give a proof of  $P(n)$  for all  $n \geq 5$ , note that this is just the induction step with a restricted domain of quantification. Finally, for each  $0 \leq n < 5$ , we can apply the induction step  $n$  times to  $P(0)$  to prove  $P(n)$ , from which follows the conditional “if  $P(n+1)$ , then  $P(n)$ ”. Feeding these three pieces into the 5-induction principle, we have produced a proof of  $\forall n.P(n)$ . We can tell a similar story if we start with a proof by 5-induction; using  $P(5)$  and the downward induction step gives a proof of  $P(0)$ , and taking all three subproofs together gives the standard induction step. Most importantly, this translation goes two ways: if I start with a proof by induction, translate it to a proof by 5-induction, and then translate it back, I end up with the same proof I started with, and vice versa. So if I start with distinct proofs by induction I will produce distinct proofs by 5-induction. If we take these induction rules to describe a particular inferential structure, then this translation rule shows that they give superficially different descriptions of the exact same inferential structure. That is, nothing is lost in translation.

So, to return to the motivating case, this syntactic principle of proof individuation avoids the problems of the naive account, and it adequately characterizes the inferential structure of a proof by induction. Because it judges induction and 5-induction to describe the same network of inferences, there is no circularity nearby. One could, of course, use induction to conclude that  $\forall n.P(n)$ , and then argue, again by induction, that this implies  $\forall n.P(n)$ . But this is not some pathological feature of induction, because this can be done with any explanation of a universal generalization. If induction has some endemic explanatory defect, it must be found in whatever  $X$  factor it is that endows proofs with explanatory power. And this is just the very difficult question that Lange’s argument aimed to avoid. So if we adopt this syntactic principle, then Lange’s argument is blocked.

#### 4 The semantic approach

Alternatively, we might try to characterize the inferential network of an inductive proof more directly, with more familiar mathematical objects. The main difficulty here is turning an inductive principle, which is like a recipe, into a mathematical object, which seems much more like a freestanding thing. A principle of induction tells you how to do something; it is instructions for some sort of behavior that will result in a proof. A mathematical object, on the other hand, is inert. It does not tell you how to make things. If we are to encode an inductive principle as a mathematical object, we must have some way of recovering the principle from that object. So we need some way to turn an inert mathematical object into a recipe for doing.

The key feature that allows this is the fact that induction acts “freely”. That is, it takes some number of subproofs and extends them to a larger domain in the only possible natural way. So an inductive principle is a recipe, but it is like a recipe for a stew. A stew recipe is a list of some vegetables, some meat, and some liquid. If I give you such a list, there’s really only one natural thing you can do with the listed ingredients: put them in a pot and heat them. Indeed, stews can be characterized by this fact. That is, a stew just is whatever food you end up with after heating vegetables, meat, and liquid in a pot.

Slightly less fancifully, we might try to characterize an induction principle by giving some minimal algebraic structure that will support the principle’s application. Natural-numbers induction can be applied in many contexts. For a concrete example, suppose we would like to show that a binary tree of height  $h$  has at most  $2^{h+1} - 1$  nodes. A natural strategy is to run an inductive argument on the height  $h$  of the tree. If we show that this fact is true for a tree of height 0, and that it is true for height  $h + 1$  whenever it is true for height  $h$ , then by induction it is true for any height  $h$ . We can similarly induct on lengths of lists, number of lines in a natural deduction, and so on. We could even run induction on subsets of the natural numbers. The minimal structure that will support induction is a triple  $(X, x_0, s)$  of a set  $X$ , a distinguished element  $x_0 \in X$ , and a function  $s : X \rightarrow X$ . Whenever we have a triple like this, we can apply induction. We can do this when  $X$  is the natural numbers,  $x_0$  is 0, and  $s$  is the successor function, and we can do this when  $X$  is the set of equivalence classes of binary trees by height,  $x_0$  is the set of trees of height zero, and  $s$  increments the height by 1.

Define an  $\mathbb{N}$ -algebra to be some triple like this. Any time that we have this structure, we can apply induction. Similarly, any time we have a list of vegetables and meat, we can make a stew. What we would like to do is go the other way. That is, we would like some  $\mathbb{N}$ -algebra such that any proof about it will be a proof by induction. This algebra would then be the perfect stand-in for the structure of proof by induction: it would represent all and only the structure required for the proof technique. If we had such a representative  $(X, x_0, s)$ , then any proof by induction must have an inferential structure comparable to a proof about  $(X, x_0, s)$ . There must be some domain of quantification corresponding to  $X$ , and there must be some base case corresponding to  $x_0$ , and there must be some “successor” function corresponding to  $s$ . Moreover, this correspondence must preserve the inferential structure of the proof. We should be able to label elements of our domain of quantification with elements of  $X$ , and the successor function should take the element labeled by  $x$  to the element labeled by  $s(x)$ . This would give a homomorphism of inferential structures, which is what we seek to capture. And if we have a similar algebraic stand-in for some other proof method, we can turn the question of comparing the inferential structure of the proofs into the question of comparing the algebraic structure of the algebras.

So a representative of an induction principle, if it exists, is a universal “labeling algebra” for structures to which that induction principle applies. But



what makes an induction principle an *induction* principle, and not some other method of proof, is the fact that it acts “freely”. This is why a proof by induction is more like a recipe for stew than for something fussy like, say, a macaron. Mathematically, this freedom is witnessed by the existence of this sort of universal labeling algebra.<sup>2</sup> So we may assign to each induction principle a particular algebraic object: its “universal labeler”. For example, induction over the natural numbers is assigned the triple  $(\mathbb{N}, 0, S)$ , where  $S$  is the successor function. You might think of this triple as equipping the target domain of the proof,  $\mathbb{N}$ , with an algebraic structure  $(0, S)$  representing the structure of proofs by induction. Other kinds of proofs about the naturals amount to other kinds of structures that can be put on the set of natural numbers. Likewise, some other  $\mathbb{N}$ -algebra  $(X, x_0, s)$  can be thought of as equipping the set  $X$  with an algebraic structure  $(x_0, s)$  representing proofs by induction.

We’ve reduced the problem of distinguishing between proofs to a problem of distinguishing between the associated universal labelers. Again, we must be cautious. We ought to distinguish between labelers only insofar as they disagree about the inferential structure of the induction principles they represent. For example, one might take  $\mathbb{N}$  to be the set of Zermelo ordinals, where 0 is represented by  $\emptyset$ , and the successor function sends some set  $n$  to  $\{n\}$ . Or one might take  $\mathbb{N}$  to be the set of von Neumann ordinals, where 0 is represented by  $\emptyset$  and the successor function sends  $n$  to  $n \cup \{n\}$ . These choices give isomorphic  $\mathbb{N}$ -algebras, meaning that they are identical with respect to their labeling ability. In other words, there is a unique way to label von Neumann ordinals with Zermelo ordinals in a way that respects the successor function, and likewise for the other direction. So two induction principles with isomorphic representatives differ merely in relabeling, not in inferential structure.

But isomorphism is still too strict a criterion. For suppose that we preferred to prove the induction step first. Then our universal labeler would be of the form  $(X, s, x_0)$ , with  $s : X \rightarrow X$  the successor function and  $x_0$  the distinguished element. And this triple could never be isomorphic to  $(\mathbb{N}, 0, S)$ , because they are algebras of different signature. The second entry in  $(X, s, x_0)$  is a function on the first entry, and the second entry of  $(\mathbb{N}, 0, S)$  is an element of the first entry. So isomorphism simply can’t be defined between these two objects. Nevertheless, they must represent the same induction principle, if syntactic reshuffling really makes no difference.

To solve this, we may again appeal to the existence of a canonical, systematic correspondence between substructures that generates a translation. Given an  $\mathbb{N}$ -algebra, we can produce an algebra for the shuffled induction principle by using the same set but taking the third entry for the second and the second for the third. This same translation principle will also allow us to turn a shuffled induction algebra into an  $\mathbb{N}$ -algebra. And no choices ever need be made in this translation, so when following this rule we never add or subtract any information. Moreover, when this sort of translation exists, it always sends universal

<sup>2</sup> More prosaically, every induction principle has an associated (homotopy) initial object in the category of algebras associated with that induction principle. See Sojakova (2015) for a precise statement and proof of this fact.

labelers to universal labelers, so  $(\mathbb{N}, 0, S)$  is matched up with  $(\mathbb{N}, S, 0)$ . If we can reconstruct a network of inferences from the triple  $(\mathbb{N}, 0, S)$ —and we can—then we can reconstruct the same network from  $(\mathbb{N}, S, 0)$  by first translating it by our rule to  $(\mathbb{N}, 0, S)$ . So we should not distinguish between induction principles that differ only up to this kind of translation.<sup>3</sup>

This individuation principle also judges induction and 5-induction to be the same principle. Induction may be represented by the triple  $(\mathbb{N}, 0, S)$  of a set, a distinguished zero element, and a successor function. 5-induction appears to be more complicated, as it is represented by a quadruple  $(\mathbb{N}', 0, 5, S)$  of a set  $\mathbb{N}$ , two distinguished elements 0 and 5, and a successor function  $S$  such that 5 is equal to  $S^5(0)$ . These could not be isomorphic, since they are tuples of different length, but as algebras they describe the same structure. Given a triple  $(\mathbb{N}, 0, S)$  we can produce a quadruple  $(\mathbb{N}, 0, S^5(0), S)$ , and given a quadruple we can produce a triple by forgetting about the second distinguished element. That is, the only difference between induction and 5-induction is that 5-induction gives a special name to the element  $S^5(0)$ . But we should not care about the names we give to places in the network of inferences, because these are facts about the representation, not the structure of the proof itself. It is really no different from reordering  $S$  and 0.

So we have a second principle of proof individuation that also avoids the problems of the naive account, and also adequately characterizes the inferential structure of proof by induction. And it comes to the same verdict as the first principle: induction and 5-induction describe the same network of inferences from different points of view. The circularity argument is undercut because there is no circle after all. So again, if induction systematically fails to provide explanations, then this is a fact about the extra explanatory factor, not the structure of inductive proofs.

## 5 Conclusion

Two principles of proof individuation may seem an embarrassment of riches. The success of Lange's argument turns on which principle we adopt, so we must pick the correct one. I have offered two serious options in addition to our naive point of departure, so we now seem to be faced with the task of sorting out the right individuation principle from all the available options. This is a valuable task, and its consequences are far broader than the present question of induction's explanatory power. Mathematicians assess proofs with respect to many other values, as well—beauty, purity, depth, and more. Insofar as these are virtues of proofs and not presentations, an account of any one of these must only appeal to properties that are invariant under the right notion of proof identity. Or, contrapositively, any aspect of a proof that appears only in some of its presentations cannot be relevant to its explanatory power, beauty,

---

<sup>3</sup> To put this all more precisely, we should not distinguish between induction principles if their associated categories of algebras are equivalent. Equivalence preserves initiality, so a “universal labeler” for one principle is equivalently a “universal labeler” for the other.

depth, and so on. So a criterion of proof individuation is a valuable tool in the development of these accounts, offering at least some constraint on the wildly divergent intuitions about explanatory power. By the same token, a criterion of proof individuation helps our intuitions about these other features to regiment our intuitions about explanatory power. When we evaluate criteria like the two I have presented, we can appeal to many more considerations than explanation. This gives us many reasons on which to make our choice—and my opponent a wider set of cases to wield against my criteria.

But even without looking to these other considerations, you might worry that neither of my criteria gets things right, since they conflict with intuitions about proof identity itself. You might worry that these criteria identify proofs that they should not. For example, most any proof of a statement beginning “for all natural numbers  $n...$ ” is identical to some proof by induction or other, by these criteria—this is part of why induction and  $\omega$ -induction deliver identical proofs of statements about the naturals. This seems counterintuitive, for it seems that there are some proofs about the natural numbers that are not proofs by induction. And if this worry is right, then what I’ve said above isn’t very reassuring when it comes to the circularity worry. Since my solution relies on identifying the two proofs that generated the circularity, any reason to think that my criteria are too weak is also a reason to worry that the right criteria won’t defuse the circularity.

To some extent this is just the clash of intuitions about induction with which we began. On the one hand are the intuitions that induction is paradigmatically non-explanatory; on the other, the thought that induction best characterizes the naturals and gives the most explanatory proofs. According to my criteria proofs about the naturals are always inductive, so they give the *only* proofs. This seems right to me. Proofs about the naturals generally involve induction because they are about the natural numbers, what it is to be a natural number is to be zero or a successor of a natural number. It’s not just that induction best characterizes the naturals, it defines them. So if you’d like to prove something about the naturals, you’ll have to say what you’re talking about, and this means referring to induction. The approaches I’ve sketched above are ways of spelling out the thought that induction best characterizes the natural numbers.

Because this is just where we began, I don’t expect that these considerations will do much to assuage worries about this verdict. My purpose in offering the criteria above is, in part, to draw a connection between this debate and any other where it matters when two proofs are the same or different. With luck, we will find further resources to adjudicate the present dispute. But this will only succeed if we use the right criterion from above. However, we need not choose; they come down to the same thing. As shown by Awodey et al. (2012) and Sojakova (2015), the inductive types of the syntactic approach may be equivalently characterized by the universal labeling algebras of the semantic approach. So the approaches agree on how to individuate inferential networks produced by inductive principles—we need not choose between them. We have one good criterion of proof individuation, and with it in hand there is no threat

of circularity in inductive proofs. What's more, this criterion gives a univocal and well-defined tool we can bring to bear on questions of beauty, purity, depth, and other virtues of proofs.

**Acknowledgements** I am grateful to Craig Callender, Casey McCoy, Sebastian Speitel, and Adam Streed for helpful conversations, criticisms, and comments. Thanks, too, to the audience of the 2015 Philosophy of Logic, Math, and Physics conference at Western University. This work was performed under a collaborative agreement between the University of Illinois at Chicago and the University of Geneva and made possible by grant number (10) 56314 from the John Templeton Foundation and its contents are solely the responsibility of the authors and do not necessarily represent the official views of the John Templeton Foundation.

## References

- Awodey, S., Gambino, N., and Sojakova, K. (2012). Inductive types in homotopy type theory. In *Logic in Computer Science (LICS 2012)*, pages 95–104. IEEE Computer Society.
- Baker, A. (2010). Mathematical induction and explanation. *Analysis*, 70(4):681–689.
- Brown, J. R. (1997). Proofs and pictures. *British Journal for the Philosophy of Science*, 48:161–180.
- Hafner, J. and Mancosu, P. (2005). The varieties of mathematical explanation. In Mancosu, P., Jørgensen, K. F., and Pedersen, S. A., editors, *Visualization, Explanation and Reasoning Styles in Mathematics*, pages 215–250. Springer, Dordrecht.
- Kreisel, G. (1971). A survey of proof theory II. In Renstad, J. E., editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 109–170. North-Holland, Amsterdam.
- Lange, M. (2009). Why proofs by mathematical induction are generally not explanatory. *Analysis*, 69(2):203–211.
- Sojakova, K. (2015). Higher inductive types as homotopy-initial algebras. In Rajamani, S. and Walker, D., editors, *Symposium on Principles of Programming Languages (POPL 2015)*, pages 31–42. Association for Computing Machinery.
- The Univalent Foundations Program (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. <http://homotopytypetheory.org/book>, Institute for Advanced Study.