Mathematical proof representations and assessment with Parsons problems

Chris Sangwin (& Sal Mercuri)

March 2026

What is a Parson's problem?

Prove the following theorem by dragging sentences and arranging them in the correct order.

Theorem: If $3 \cdot 2^{172} + 1$ is a perfect square, then $3 \cdot 2^{172} + 173$ is not a perfect square.

Construct your solution here:	*Drag from here:*
Assume that $3 \cdot 2^{172} + 1$ is a perfect square.	Since $k^2 < 3 \cdot 2^{172} + 173 < (k+1)^2$ it is strictly between two successive squares k^2 and $(k+1)^2$, it cannot be a perfect square.
Since $3 \cdot 2^{172} + 1 > 2^{172} = (2^{86})^2 > 172^2$, we have $k > 172$.	We have $k^2 = 3 \cdot 2^{172} + 1 < 3 \cdot 2^{172} + 173$.
There is a positive integer k such that $3 \cdot 2^{172} + 1 = k^2$.	
Also, $3 \cdot 2^{172} + 173 = (3 \cdot 2^{172} + 1) + 172 < k^2 + k$. Further, $k^2 + k < (k+1)^2$.	

Check

Online assessment of proof...

- When writing STACK we don't know (exactly) what we want...
- We learn a lot...
- Validity, correctness, and aesthetics...
- Writing problems for a format is an art.
- Tension between general tools *vs* bespoke question...

Issues: reliability, consistency, efficiency.
vs
Flexibility.

On the nature of proof

Standard proof types:

- Calculation/equivalence reasoning. (The Cartesian pattern)
- Reformulate to an equivalent:
  - using a definition
  - splitting “if any only if” into two proofs
  - re-writing an implication as the contrapositive
- Definition chase.
- Exhaustive cases.
- Proof by induction. (The recursive pattern)
- The pattern of two loci (Geometry)
- Proof by contradiction.
- Counterexample.
- Proof by generalizing. (Rarely found)

Nested structure

Theorem

$n \in \mathbb{Z}$ is odd if and only if n^2 is odd.

Proof.

Assume that n is odd.
Then there exists an $m \in \mathbb{Z}$ such that $n = 2m + 1$.

$$n^2 = (2m+1)^2 = 2(2m^2 + 2m) + 1.$$

Define $M = 2m^2 + 2m \in \mathbb{Z}$ then $n^2 = 2M + 1$. Hence n^2 is odd.
We reformulate "n^2 is odd $\Rightarrow$ n is odd " as the contrapositive.
Assume that n is not odd. Then n is even, and so there exists an $m \in \mathbb{Z}$ such that $n = 2m$.

$$n^2 = (2m)^2 = 2(2m^2).$$

Define $M = 2m^2 \in \mathbb{Z}$ then $n^2 = 2M$. Hence n^2 is even. □

1. Reformulate “if and only if”
   - $A \Rightarrow B$
   - $B \Rightarrow A$
2. $A \Rightarrow B$ is a straightforward “definition chase”.

$$n = 2m + 1 \Rightarrow n^2 = 2M + 1$$

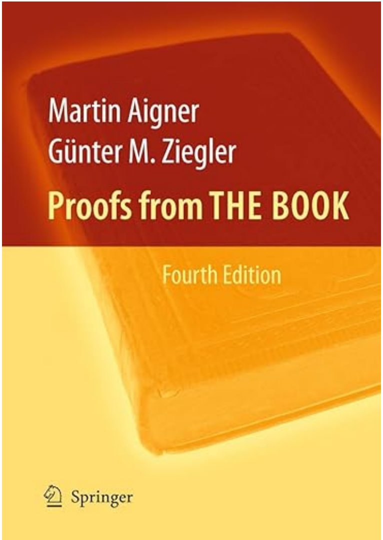
3. $B \Rightarrow A$ is reformulated as the contrapositive.
   1. Every integer is either odd or even, (but not both!). “not odd” $\rightarrow$ “even”.
   2. “definition chase” using the definition of even.

$$n = 2m \Rightarrow n^2 = 2M$$

Advice to students: write simple and boring proofs!

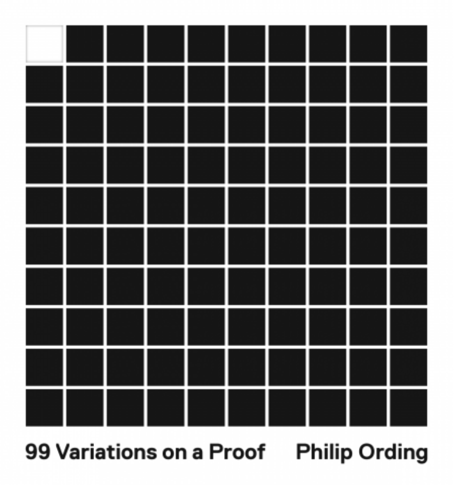
- Structured proofs are easier to check
- Structured proofs are easier to discuss
- Structure aids pattern matching
- Reduce load

An example of expert reversal.


Martin Aigner
Günter M. Ziegler
Proofs from THE BOOK
Fourth Edition
Springer

Ording....

"... a deep and thoughtful examination of the nature of mathematical arguments, of mathematical style, and of proof itself."



Tools in STACK to *represent* proof structure

- We are not building a theorem prover.
- Proofs are made up of *strings*.

Proof blocks

`proof_iff(A, B)`

- `proof_iff` represents “if and only if”
- `A` and `B` are proofs.
  Or tags to lines in a proof.
- STACK has tools to make
  `proof_iff(A, B)`≡`proof_iff(B, A)`
  ... which is why we need CAS
- Other blocks for different proofs.
- Goal: let teachers write structured proofs.
- Future: Let students write structured proofs themselves....

Damerau-Levenshtein distance

- A metric for measuring the difference between two strings.
- Edit distance:
  insertions, deletions, transition or substitutions.
- Already in STACK.

Use the Damerau-Levenshtein distance between two (flattened) proofs.

Example

$n \in \mathbb{Z}$ is odd if and only if n^2 is odd.

Example on STACK-demo: question authoring run-through.

✖ Incorrect answer.

Assume that $3 \cdot 2^{172} + 1$ is a perfect square.

Since $3 \cdot 2^{172} + 1 > 2^{172} = (2^{86})^2 > 172^2$, we have $k > 172$.

There is a positive integer k such that $3 \cdot 2^{172} + 1 = k^2$.

Also, $3 \cdot 2^{172} + 173 = (3 \cdot 2^{172} + 1) + 172 < k^2 + k$. Further, $k^2 + k < (k+1)^2$.

We have $k^2 = 3 \cdot 2^{172} + 1 < 3 \cdot 2^{172} + 173$.

Since $k^2 < 3 \cdot 2^{172} + 173 < (k+1)^2$ it is strictly between two successive squares k^2 and $(k+1)^2$, it cannot be a perfect square.

Algorithm

Student's & teacher's proofs are *lists*.

1. Establish closest correct proof
   `proof_iff(A, B)`≡`proof_iff(B, A)`
2. Provide feedback on any edits needed
   "insert here", "swap lines", ...

Students' proof: interleaving

Assume n is odd.
There exists j such that $n = 2j - 1$.
$[\cdots]$
Assume m is even.
There exists k such that $m = 2k$.
$[\cdots]$

Student interleaved their blocks:

Assume n is odd.
Assume m is even.
$[\cdots]$

Poulson 2023

Drag from here:

6 Since $3 \cdot 2^{172} + 173$ is strictly between two successive perfect squares k^2 and $(k+1)^2$, it cannot be a perfect square.

Construct your solution here:

1 Assume that $3 \cdot 2^{172} + 1$ is a perfect square.

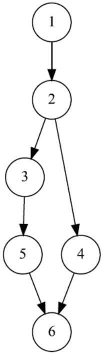
3 Since $3 \cdot 2^{172} + 1 > 2^{172} = (2^{86})^2 > (172)^2$, we have $k > 172$.

4 We have, $k^2 = 3 \cdot 2^{172} + 1 < 3 \cdot 2^{172} + 173$.

5 Also, $3 \cdot 2^{172} + 173 = (3 \cdot 2^{172} + 1) + 172 < k^2 + k$. Further $k^2 + k < (k+1)^2$.

2 There is a positive integer k such that $3 \cdot 2^{172} + 1 = k^2$.

7 Since 172 is not a perfect square, $3 \cdot 2^{172} + 173 = (3 \cdot 2^{172} + 1) + 172$ cannot be a perfect square.



Theorem: If $3 \cdot 2^{172} + 1$ is a perfect square, then $3 \cdot 2^{172} + 173$ is not a perfect square.

Assume that $3 \cdot 2^{172} + 1$ is a perfect square.
There is a positive integer k such that $3 \cdot 2^{172} + 1 = k^2$.

Since $3 \cdot 2^{172} + 1 > 2^{172} = (2^{86})^2 > 172^2$, we have $k > 172$.
Also, $3 \cdot 2^{172} + 173 = (3 \cdot 2^{172} + 1) + 172 < k^2 + k$. Further, $k^2 + k < (k+1)^2$.

We have $k^2 = 3 \cdot 2^{172} + 1 < 3 \cdot 2^{172} + 173$.

Since $k^2 < 3 \cdot 2^{172} + 173 < (k+1)^2$ it is strictly between two successive squares k^2 and $(k+1)^2$, it cannot be a perfect square.

Theorem: If $3 \cdot 2^{172} + 1$ is a perfect square, then $3 \cdot 2^{172} + 173$ is not a perfect square.

Assume that $3 \cdot 2^{172} + 1$ is a perfect square.
There is a positive integer k such that $3 \cdot 2^{172} + 1 = k^2$.

Since $3 \cdot 2^{172} + 1 > 2^{172} = (2^{86})^2 > 172^2$, we have $k > 172$.
Also, $3 \cdot 2^{172} + 173 = (3 \cdot 2^{172} + 1) + 172 < k^2 + k$. Further, $k^2 + k < (k+1)^2$.

We have $k^2 = 3 \cdot 2^{172} + 1 < 3 \cdot 2^{172} + 173$.

Since $k^2 < 3 \cdot 2^{172} + 173 < (k+1)^2$ it is strictly between two successive squares k^2 and $(k+1)^2$, it cannot be a perfect square.

Do students interleave and do they notice?

1. To what extent do students interleave steps in a proof?
2. When asked to compare do students identify interleaving as the difference?
3. To what extent do students consider this difference as important?

1. Students were asked to prove Poulson's problem
2. Students were asked to compare two versions of proofs
    - Interleaved
    - non-interleaved

Participants were in year 1 of a mathematics degree.

Results

36 attempts

- The most popular correct answer (19 students) was the non-interleaved: (**3**, **4**, **5**).
- The only other correct response (10 students) was the interleaved version.

Q2a. Which of these proofs is correct?

- Proof 1 only: 5 (8%)
- Proof 2 only: 4 (6%)
- Both are correct: 55 (86%)
- Neither are correct: 0 (0%)

Q2b. Which of these proofs do you prefer?

- Proof 1: 21 (32%)
- Proof 2: 33 (52%)
- No preference: 10 (16%)

Q3. Please describe any differences between the two proofs above?

Proof 2 has a much better flow, going from one step to another, following a logical path - defines k^2, shows k^2 is less than the claimed perfect square, then shows how $k > 172$ and then uses that to show that the claimed square cannot be a square. Whereas proof one seems to jump around a bit.

This difference doesn't actually matter to the logic of the proof because both of the statements are self-evident, and so have an interchangeable order. However, Proof 2 feels nicer to read, since line 4 in this proof is more closely tied to line 5.

Our experience of Parsons problems

Raises questions about:

1. the nature of the structure of traditional mathematical proof itself;
2. the criteria used to assess the correctness of a proof;
3. what feedback might help students appreciate structure and correctness.

1. The discipline of creating a proof-tree is rather helpful to teachers;
2. dependency graphs are necessary for some, but not all, proofs.

Conclusion

Technical innovations:

- Write proof as a tree
- Automatic feedback via Damerau-Levenshtein distance.
- Need tools to create dependency graphs
- Completely bespoke feedback

Educational conclusions

- These items are not easy for students!
- Force students to *read* carefully.
- Some “spotting” - might change interaction with proof.
- (Still need to let students create trees)

Interesting alternative activity.