# A summary of:

# "Expert and Novice Approaches to Reading Mathematical Proofs"

Matthew Inglis and Lara Alcock
Journal for Research in Mathematics Education, vol 43, 358-390

# Background

- Who are these people?

- **Why isn't the title "Expert and novice approaches to validating mathematical proofs"?**
  - Because most studies on reading proofs have actually focused on validating proofs.

# Part 1 – Giant Literature Review

- Review lit. on how experts and novices judge proof validity
  - Novices do poorly when judging the validity of proofs, often no better than chance (Selden & Selden 2003; Alcock & Weber 2005).
    - Note: Also true for judging the validity of single statements (ex. Math 220)
  - Past studies (Selden and Selden 2003) and general belief is that, in contrast, experts exhibit "uniform agreement" about the validity of proofs.
    - Recently, there is some (tentative) disagreement on this, i.e. even experts may not uniformly agree whether a proof is valid or invalid (Weber 2008)

Research Question 1: Do research mathematicians typically agree on the validity of purported proofs?

# Part 1 – Giant Literature Review

- Discuss theoretical ideas for why student difficulties occur
  - Students may not understand what constitutes a proof (Coe & Ruthven, 1994; Harel & Sowder, 1998)
    - Recent research (Weber, 2010) suggests that "high-achieving" students did not have this misconception, and it is more likely that these students are unskilled at validating deductive arguments
  - Students devote attention to surface features, rather than attending to the underlying logical structures of proofs (Selden & Selden, 2003).

Research Question 2: When validating proofs, do experts and novices attend to different parts of purported proofs to different degrees?

# Part 1 – Giant Literature Review

- Review the literature on how successful validation may be done
  - Two broad strategies based on introspection by experts (Weber & Mejia-Ramos, 2011):
    - “zooming-in”: filling in gaps between successive statements
      - Warrant: a justification that allows the reader to conclude that the statement follows from a subset of the previous lines and known axioms
      - Warrants may be implicit: the reader needs to
        - (a) decide when a warrant is required,
        - (b) infer the implicit warrant intended by the author and
        - (c) evaluate the validity of the warrant.

# Part 1 – Giant Literature Review

- Review the literature on how successful validation may be done
  - Two broad strategies based on introspection by experts (Weber & Mejia-Ramos, 2011):
    - "zooming-out": decompose a proof into methodological moves (cohesive strings of logical derivations that form chunks of the whole argument) and evaluate whether these moves fit together to imply the theorem.

Research Question 3: Is there evidence of these two distinct strategies for proof validation? If so, are there expert/novice differences in frequency or sophistication of use of these strategies?

# Part 1 – Giant Literature Review

- Last: Discuss methodological difficulties in studying reading/validating proofs
  - Previously there have been two major approaches to studying proof validation:
  - 1. introspective reports by experts:
  - 2. Think-aloud protocols:
    - Problems with *reactivity*: verbalizing thoughts can alter behaviour (for better or worse)
    - Problems with *veridicality*: the accuracy of self-reports (may omit crucial components, or may report elements that did not actually occur)

  - Motivate their use of eye-tracking technology

# What did they actually do?

- 18 undergrads (50% male), completed 2 semesters of proof-based calculus and linear algebra
- 12 research mathematicians (83% male)
- Read proofs on computer screens while eye-tracking technology (embedded in screen) tracked the lines they focused on.
- 3 phases:
  - 4 “student-generated” proofs of the same theorem (Proofs 1-4 in appendix), judge validity, estimate confidence in response
  - “Break”: read two short nonmathematical passages from newpapers.
  - 2 proofs “submitted to a math journal” (Proofs 5-6 in appendix), judge validity, estimate confidence in response (?)

# Research Question 1:

- Do research mathematicians typically agree on the validity of purported proofs?
- Got 12 experts (research mathematicians) to judge the validity of 6 proofs in number theory.
  - Uniform agreement on 3 of them (proofs 1, 4, and 6 – all invalid)
  - Disagreement on remaining 3.

  - Conclusion: Experts do not agree uniformly on what constitutes a valid proof. It may depend instead on the social context in which the proof is read, or other factors.

# Research Question 1:



Table 1
*Responses to the Six Arguments, Showing Frequencies of Valid and Invalid Responses, for Each Group*

| | P1* | P2 | P3 | P4* | P5 | P6* |
|---|---|---|---|---|---|---|
| | | | Mathematicians | | | |
| Valid | 0 | 7 | 5 | 0 | 6 | 0 |
| Invalid | 12 | 5 | 7 | 12 | 6 | 12 |
| | | | Undergraduates | | | |
| Valid | 9 | 11 | 4 | 11 | 11 | 8 |
| Invalid | 9 | 7 | 14 | 7 | 7 | 10 |

*Note*. * indicates that the difference between the responses of the two groups reached significance (Fisher's Exact Test).

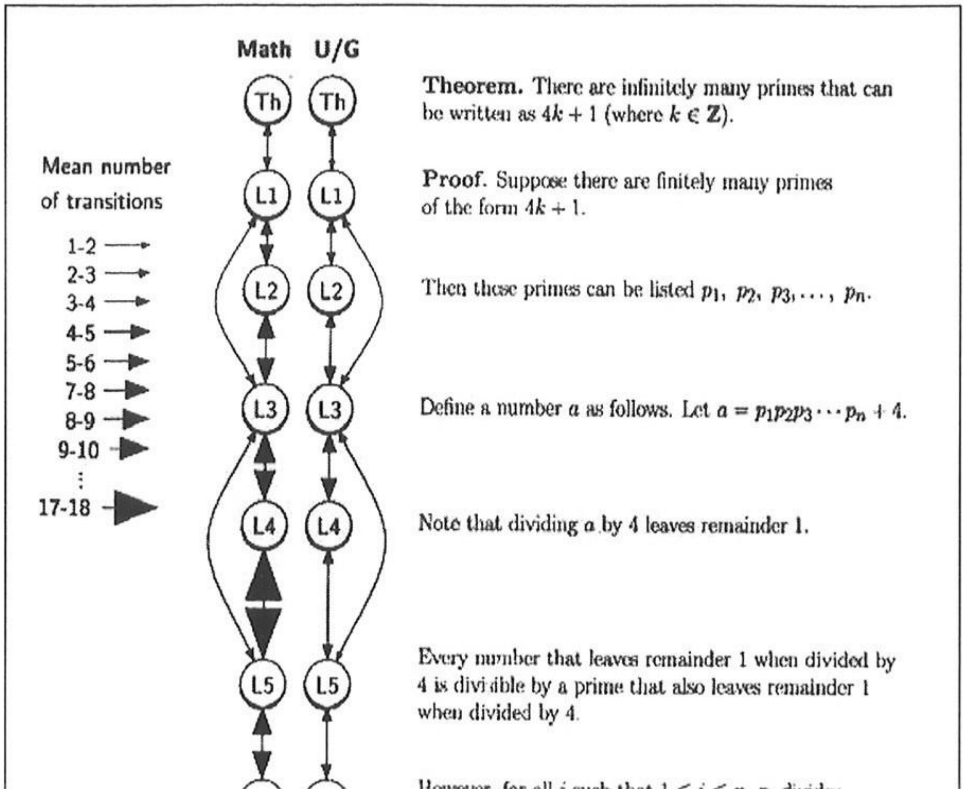
# Research Question/Goal 2:

- Do experts and novices attend to different parts of purported proofs to different degrees? ex. Do novices spend proportionately longer focusing on surface features?

- Fig. 1. → Yes. Undergraduates spent 55% of time on formulae, experts spent 45% of time.

- Fig. 3 & 4 → Experts made approx. 50% more between-line saccades than undergraduates. (moved back and forth between consecutive lines more).

- Fig. 4. → Experts had more transitions between lines that required a warrant (ex. L3-L4, L4-L5)
- (on avg. experts made 14.9 more warrant-seeking sequences when a warrant was required, undergrads made only 5.0 more when required)


Math U/G
Mean number of transitions
1-2
2-3
3-4
4-5
5-6
7-8
8-9
9-10
17-18
Th
L1
L2
L3
L4
L5
**Theorem.** There are infinitely many primes that can be written as $4k+1$ (where $k \in \mathbb{Z}$).
**Proof.** Suppose there are finitely many primes of the form $4k+1$.
Then these primes can be listed $p_1, p_2, p_3, \ldots, p_n$.
Define a number $a$ as follows. Let $a = p_1p_2p_3 \cdots p_n + 4$.
Note that dividing $a$ by 4 leaves remainder 1.
Every number that leaves remainder 1 when divided by 4 is divisible by a prime that also leaves remainder 1 when divided by 4.

# Research Question/Goal 3:

- Is there evidence for two distinct strategies (zooming out, vs. zooming-in) for proof validation?  If so, are there expert/novice differences in the frequency or sophistication of use?

- Examined “first-fixation time” for each line

- Found no significant difference between expert and undegraduate times. (fig. 2)

# Research Implications

- Having experts reflect on their practices may not lead to a valid understanding of that practice
  - The validity of proofs is not uniformly accepted/rejected
  - Experts do not “zoom-out” as self-reported

  - → to teach students expert behaviour, experiments need to be done to determine what is “expert behaviour”, even on concepts or issues that appear to be “uniformly” understood or accepted.

# Educational Implications

- We know that students' validation procedures differ from those of experts in the following ways:
  - Undergraduate students do not reliably distinguish invalid proofs, even in those cases where experts agreed uniformly.
  - Undergrads spend proportionately more time on formulae (surface features).
  - Undergrads "zoom-in" less than experts – experts devote more effort to inferring implicit between-line warrants.

- What can we do to change this?
  - 1. Self-explanation (ex. Chi et al., 1994)
    - Develop materials that encourage students to (a) decide when a warrant is required (b) infer an appropriate warrant and (c) evaluate the inferred warrant
  - 2. Write proofs differently to aid validation attempts.
    - Ex. Reduce symbolism, formulas, etc.
    - This needs to be done carefully, to train students up to read genuine proofs.