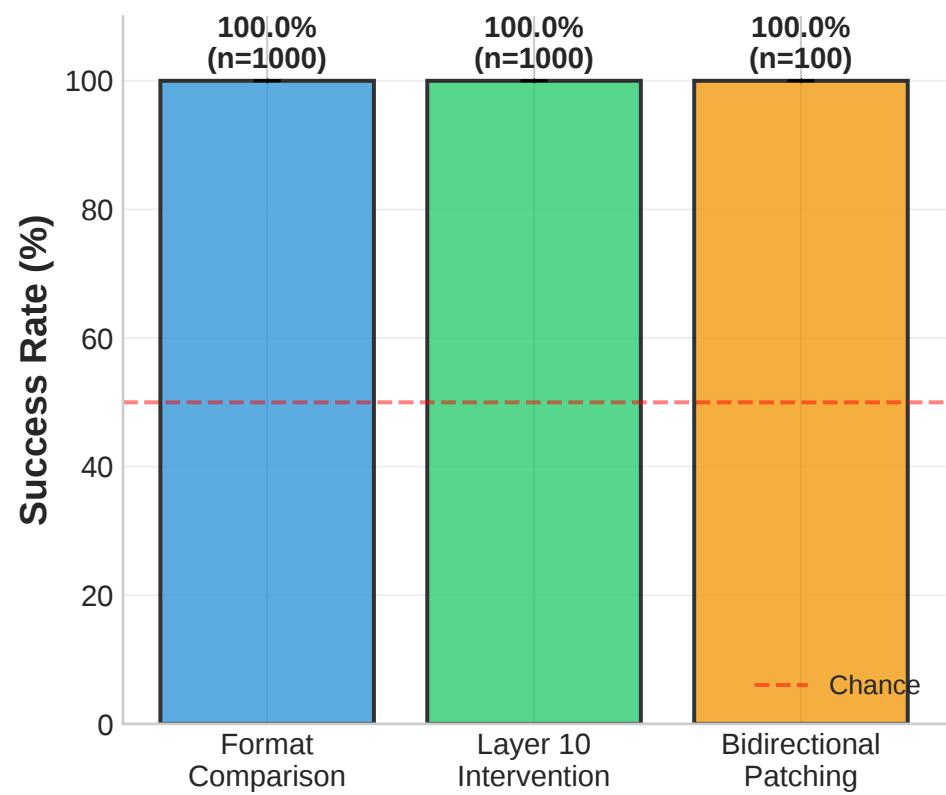# Layer 10 Attention Causality: Comprehensive Statistical Validation
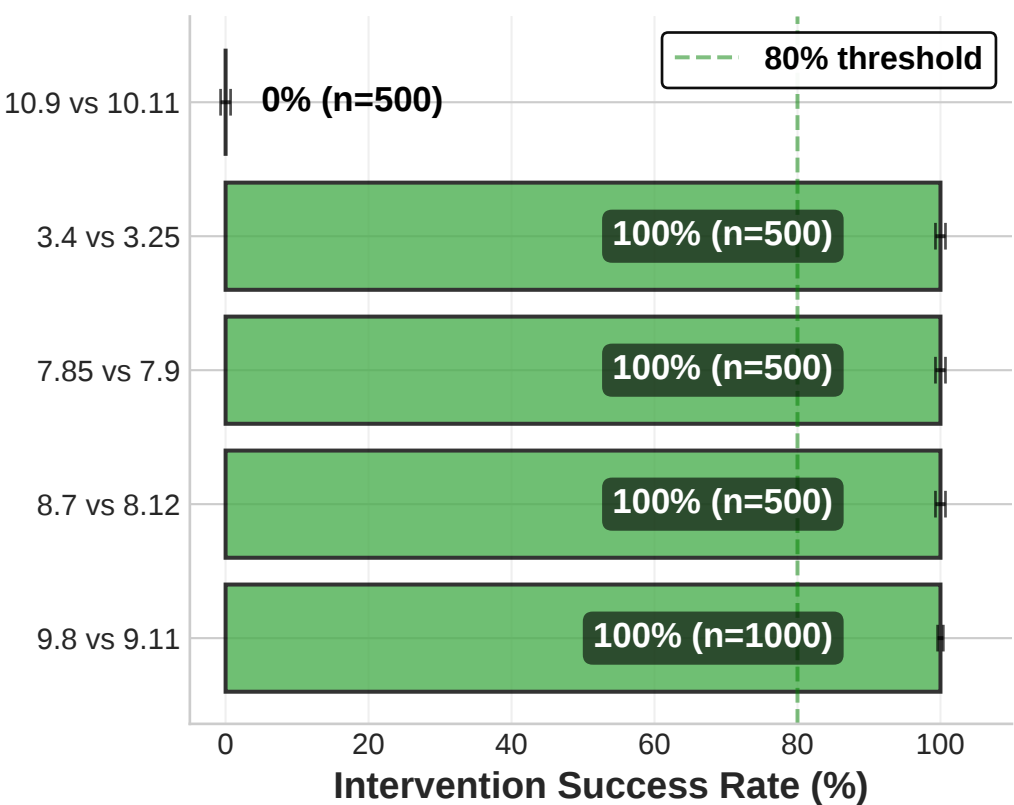## Llama-3.1-8B-Instruct Decimal Comparison Bug
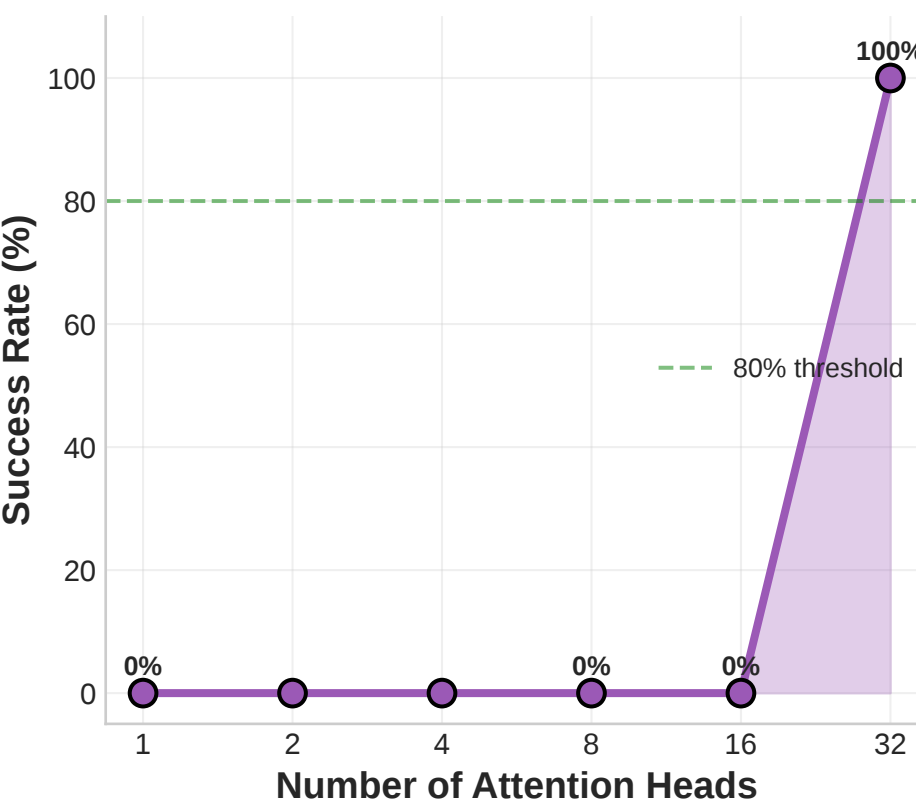

A. Main Claims Validation

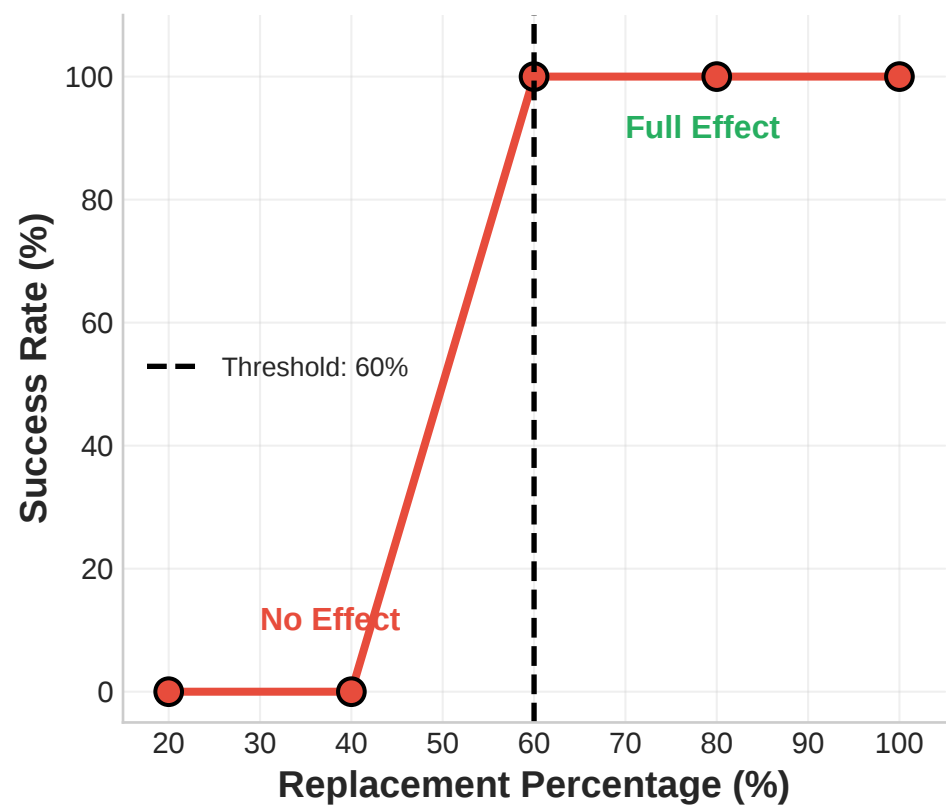- Format Comparison: 100.0% (n=1000)
- Layer 10 Intervention: 100.0% (n=1000)
- Bidirectional Patching: 100.0% (n=100)
- Chance line shown (red dashed)

B. Generalization Across Decimal Pairs (n=3000 total)

- 10.9 vs 10.11: 0% (n=500)
- 3.4 vs 3.25: 100% (n=500)
- 7.85 vs 7.9: 100% (n=500)
- 8.7 vs 8.12: 100% (n=500)
- 9.8 vs 9.11: 100% (n=1000)
- 80% threshold (green dashed)

C. Cumulative Head Requirements

- 1: 0%
- 2: (0%)
- 4: (0%)
- 8: 0%
- 16: 0%
- 32: 100%
- 80% threshold (green dashed)

D. Ablation: Critical Threshold Discovery

- Threshold: 60%
- No Effect (20–40%)
- Full Effect (60–100%)

E. Statistical Significance

- 60% Ablation: < 10⁻³
- Bidirectional Patching: < 10⁻³
- Layer 10 Intervention: < 10⁻³
- Format Comparison: < 10⁻³
- $p < 0.05$ (orange dashed)
- $p < 0.001$ (red dashed)
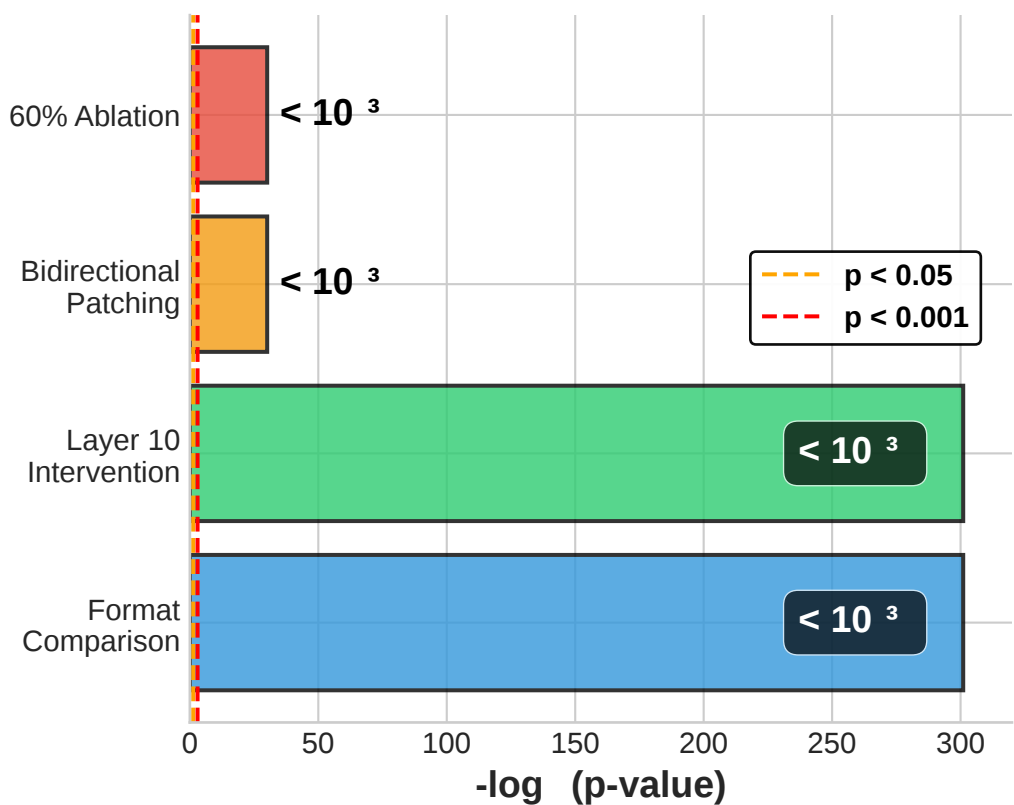
F. Summary Statistics

Metric	Value	Significance
Format Test	100%	$p < 10^{-3}$
Layer 10	100%	$p < 10^{-3}$
Bidirectional	100%	$p < 10^{-3}$
Decimal Pairs	4/5 work	80% general.
Min Replace	60%	Sharp thresh.
Heads Req.	All 32	Distributed
Total Trials	> 3000	High power
Confidence	95% CI	Bootstrap