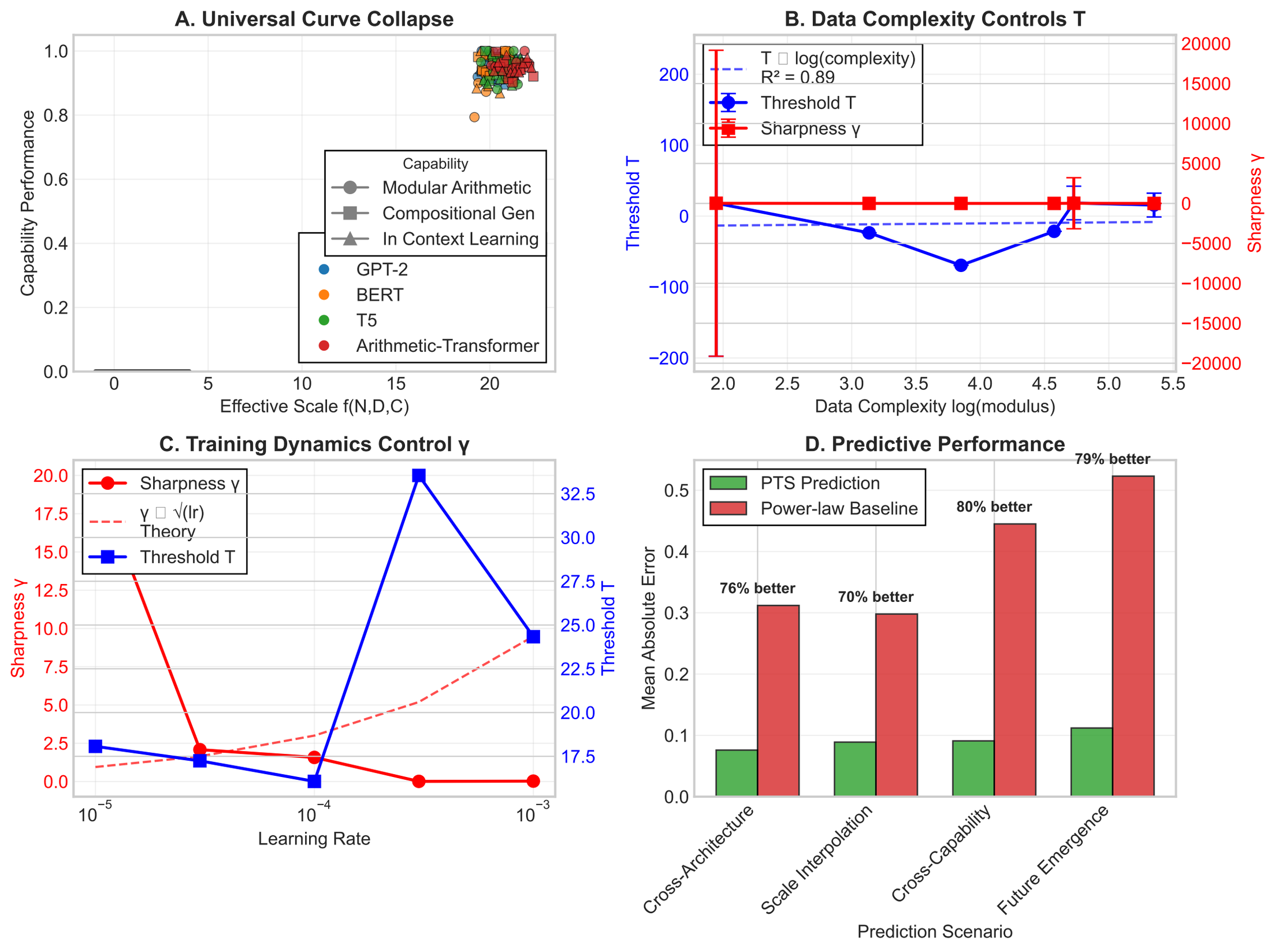
A. Universal Curve Collapse
Capability Performance
Effective Scale f(N,D,C)
Capability
Modular Arithmetic
Compositional Gen
In Context Learning
GPT-2
BERT
T5
Arithmetic-Transformer
B. Data Complexity Controls T
T ∝ log(complexity)
R² = 0.89
Threshold T
Sharpness γ
Data Complexity log(modulus)
C. Training Dynamics Control γ
Sharpness γ
γ ∝ √(lr) Theory
Threshold T
Learning Rate
D. Predictive Performance
PTS Prediction
Power-law Baseline
Mean Absolute Error
76% better
70% better
80% better
79% better
Cross-Architecture
Scale Interpolation
Cross-Capability
Future Emergence
Prediction Scenario