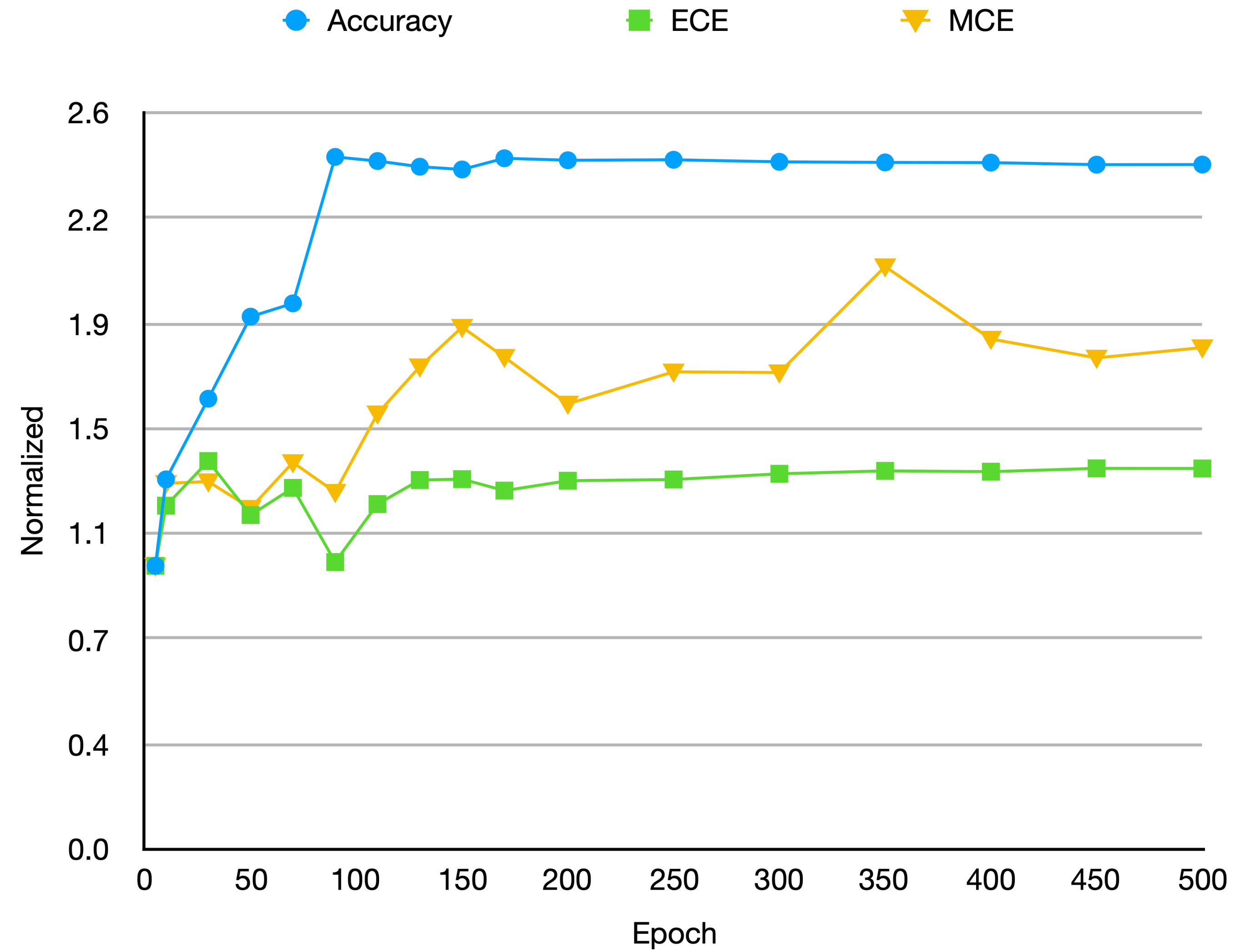


Epoch vs Calibration

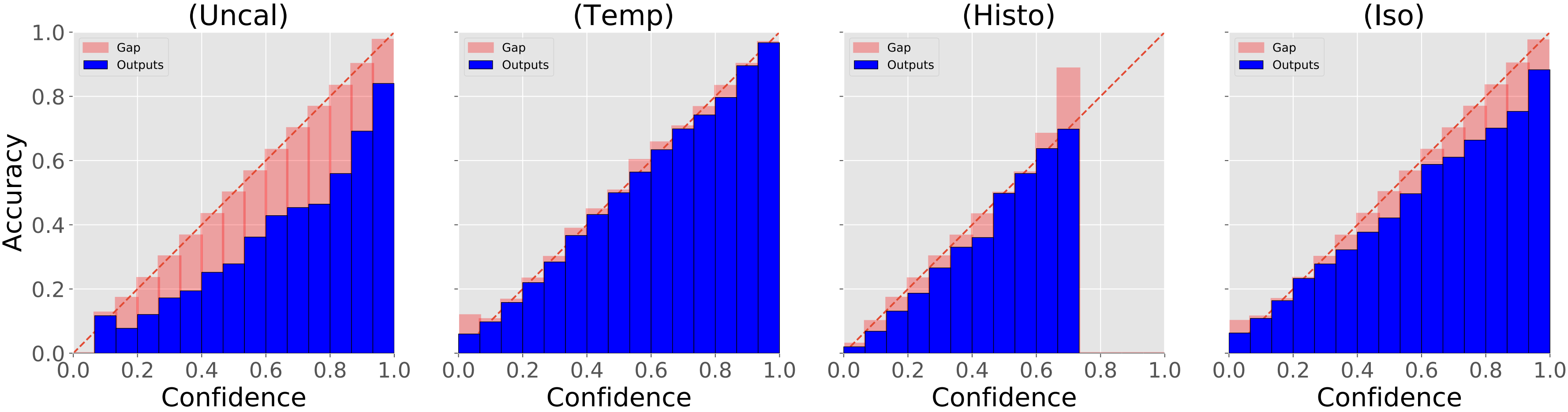
Jie Lyu

Introduction

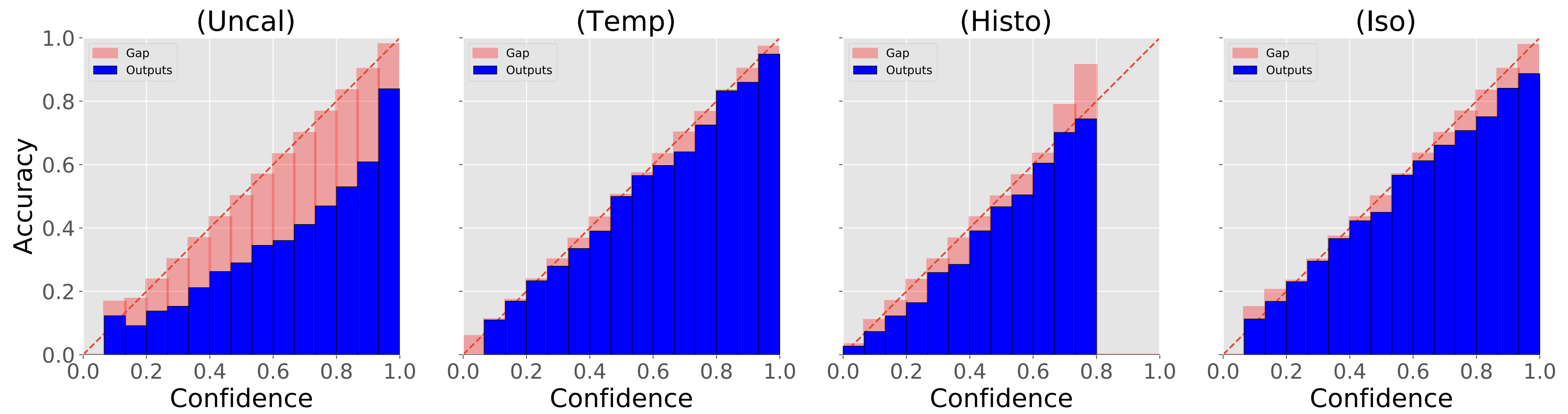
- **Question:** can training a model for too long cause it to be less calibrated?
- **Method:** trained ResNet-110 on CIFAR-100 for 500 epochs, which is more than necessary, then calculated the ECE, MCE and plot reliability diagram for different model checkpoints
- **Calibration:** methods such as temperature scaling, histogram binning and isotonic regression are applied to calibrate each checkpoint



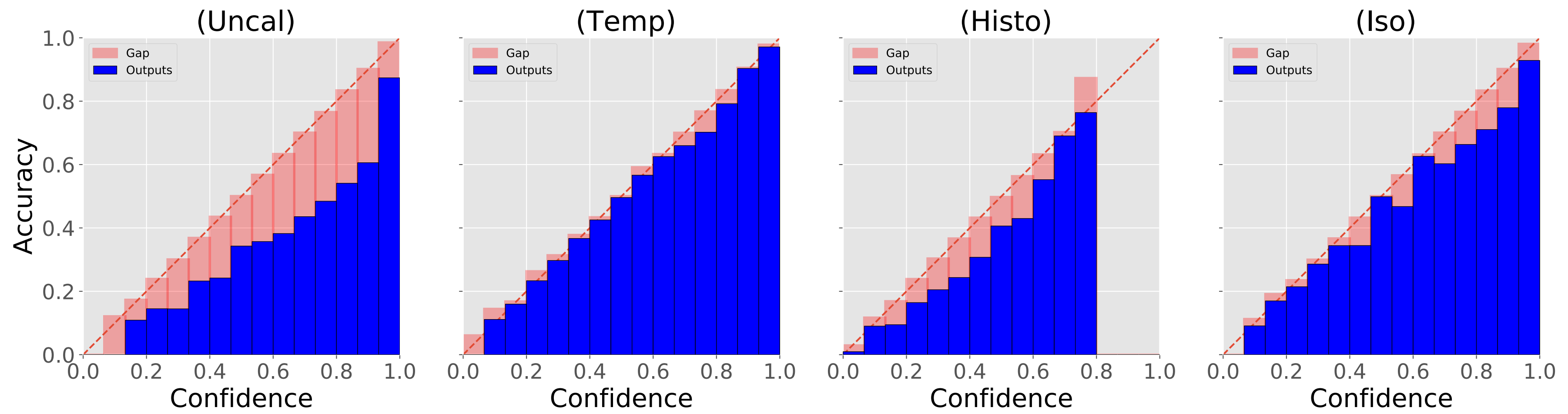
epoch_10



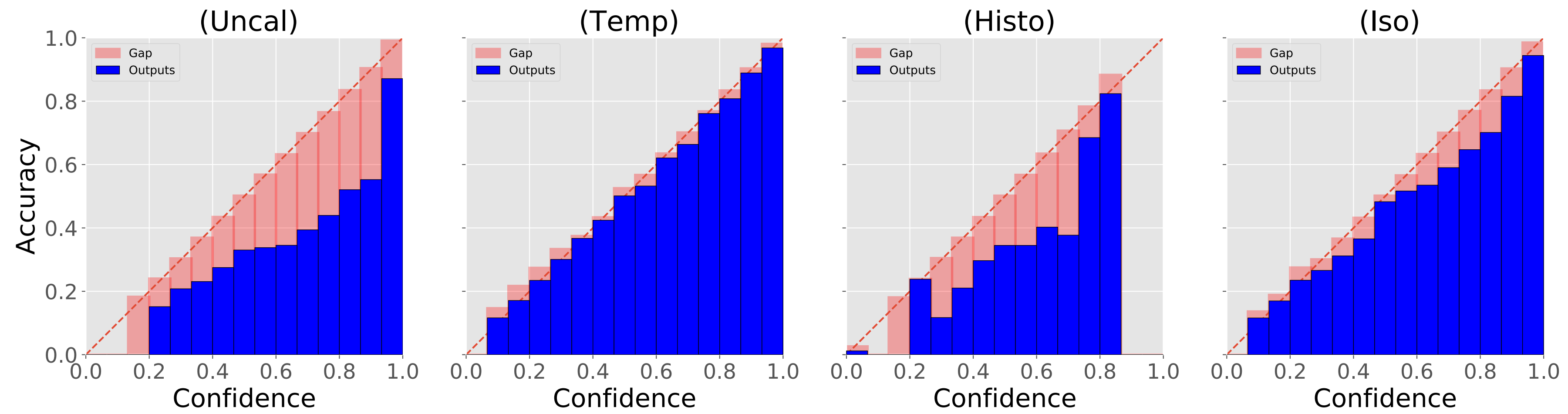
epoch_30



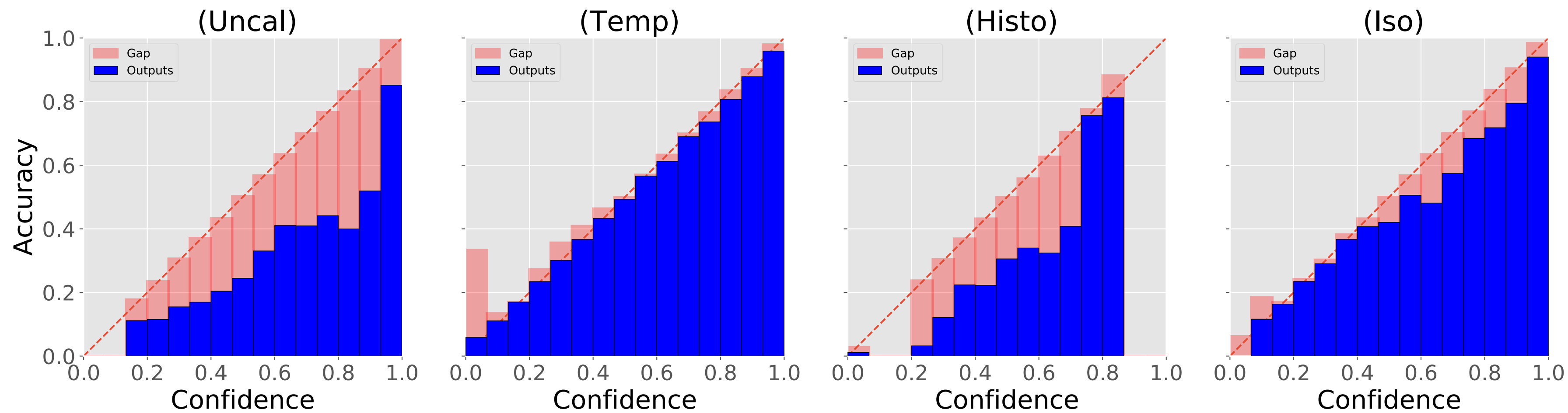
epoch_60



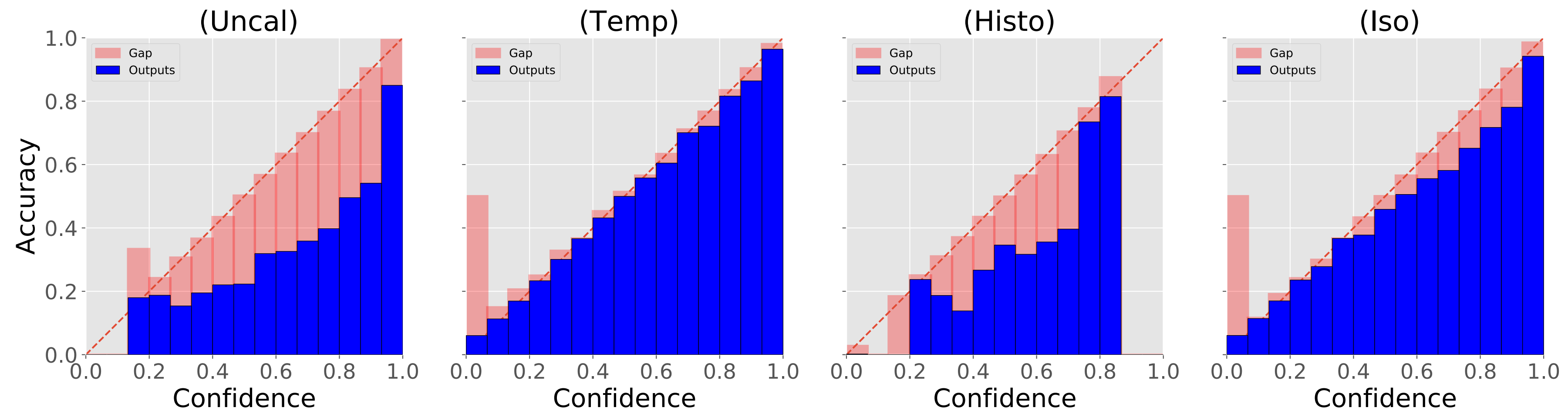
epoch_100



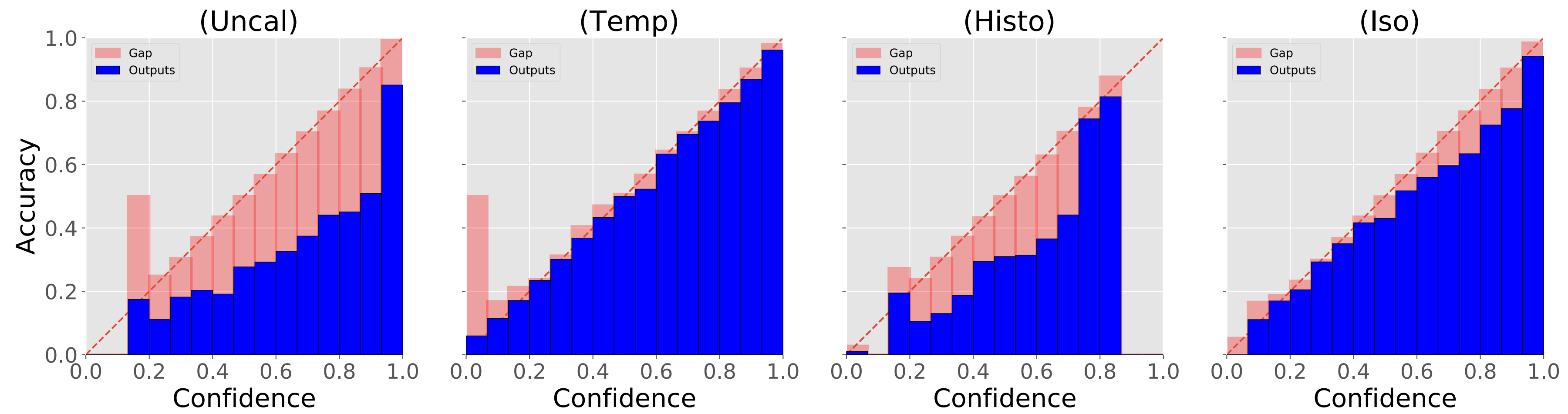
epoch_150



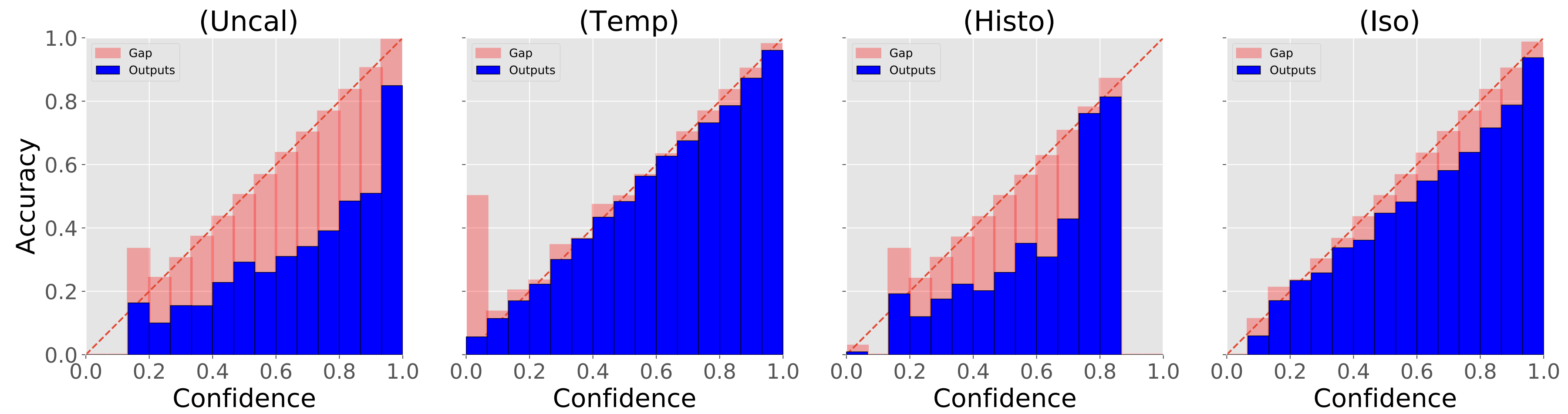
epoch_200



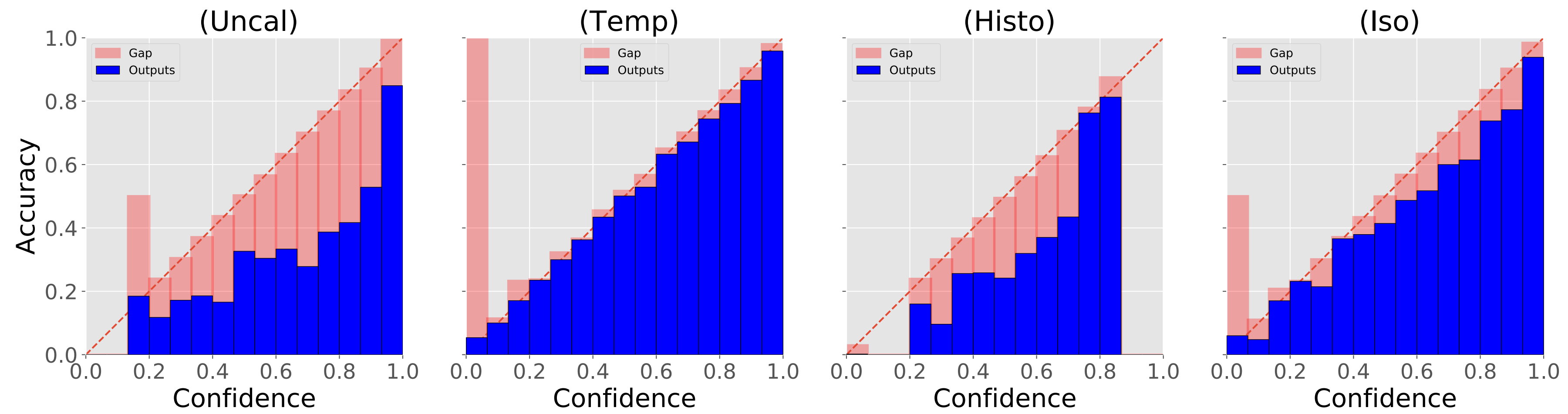
epoch_250



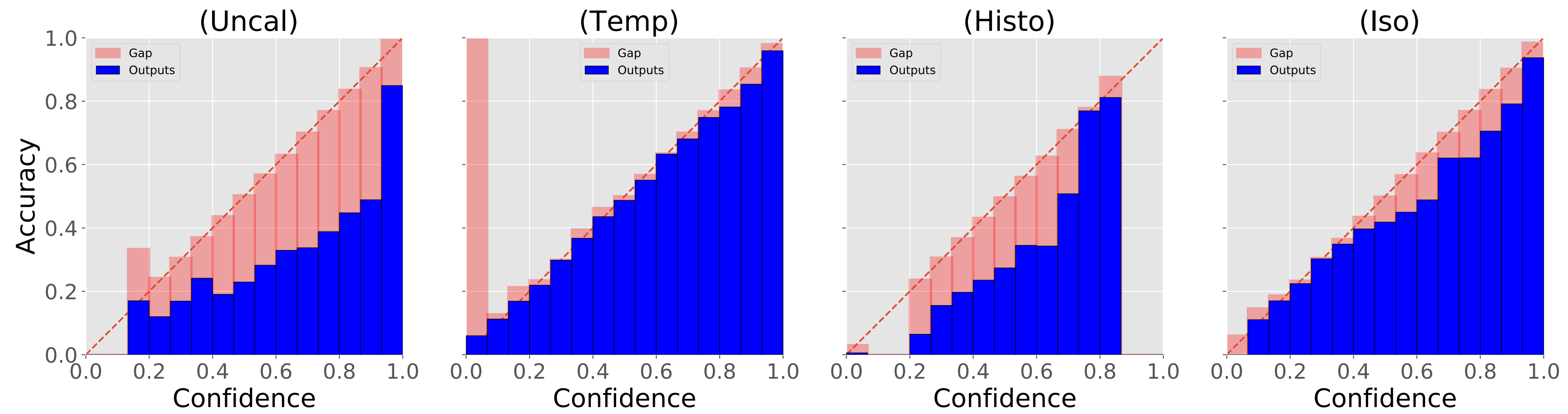
epoch_300



epoch_400



epoch_500



Conclusion

- Long period of training did make ResNet less calibrated
- While ECE showed little increase, MCE kept going up as training goes
- Temperature scaling was the most effective calibration method
- <https://github.com/jielyugt/calibration>