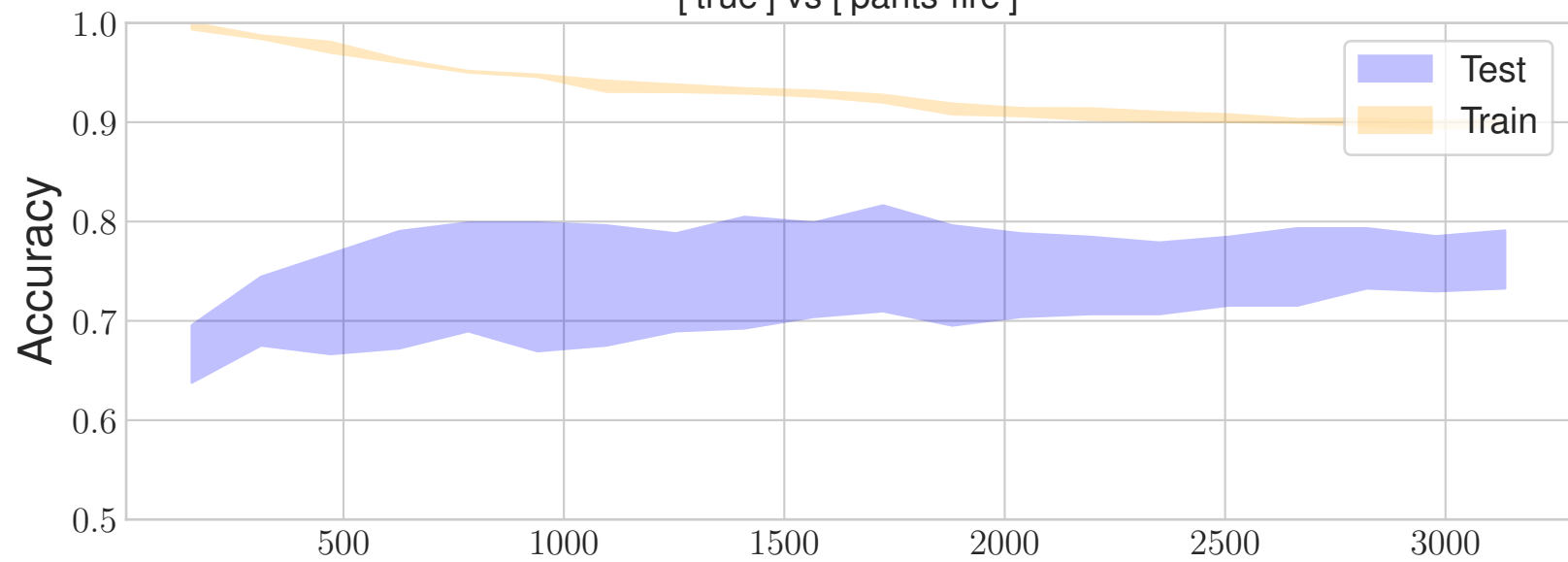
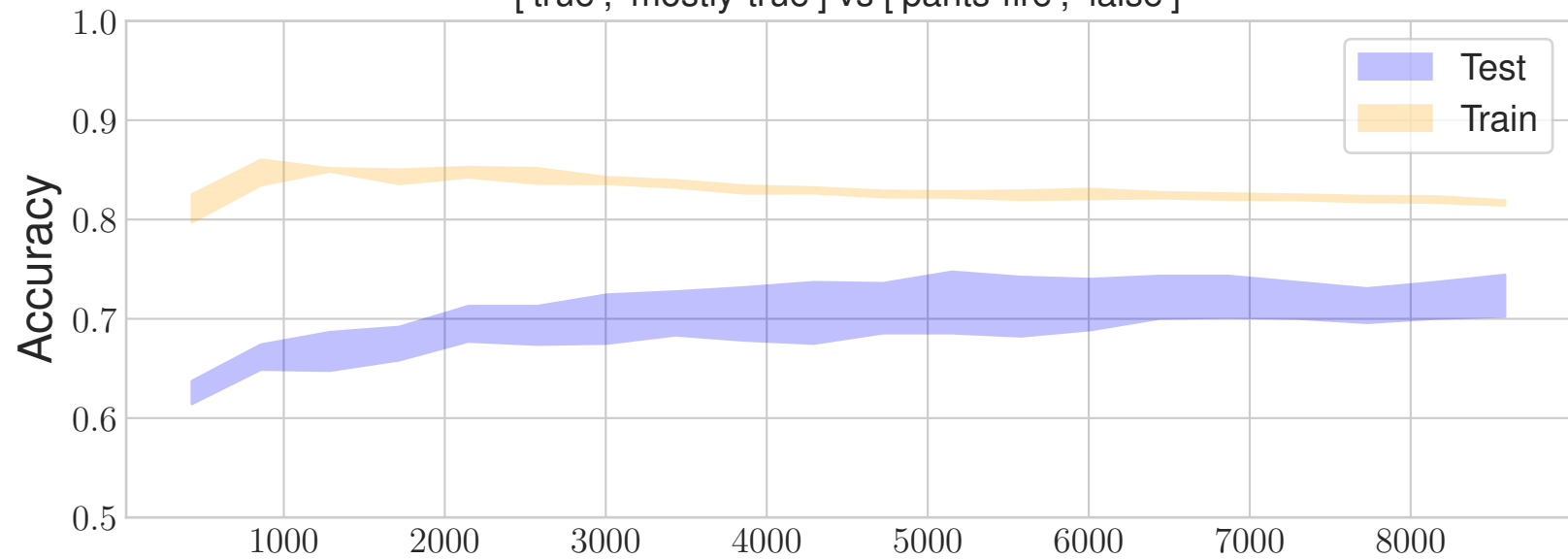


Range of performance (min, max)

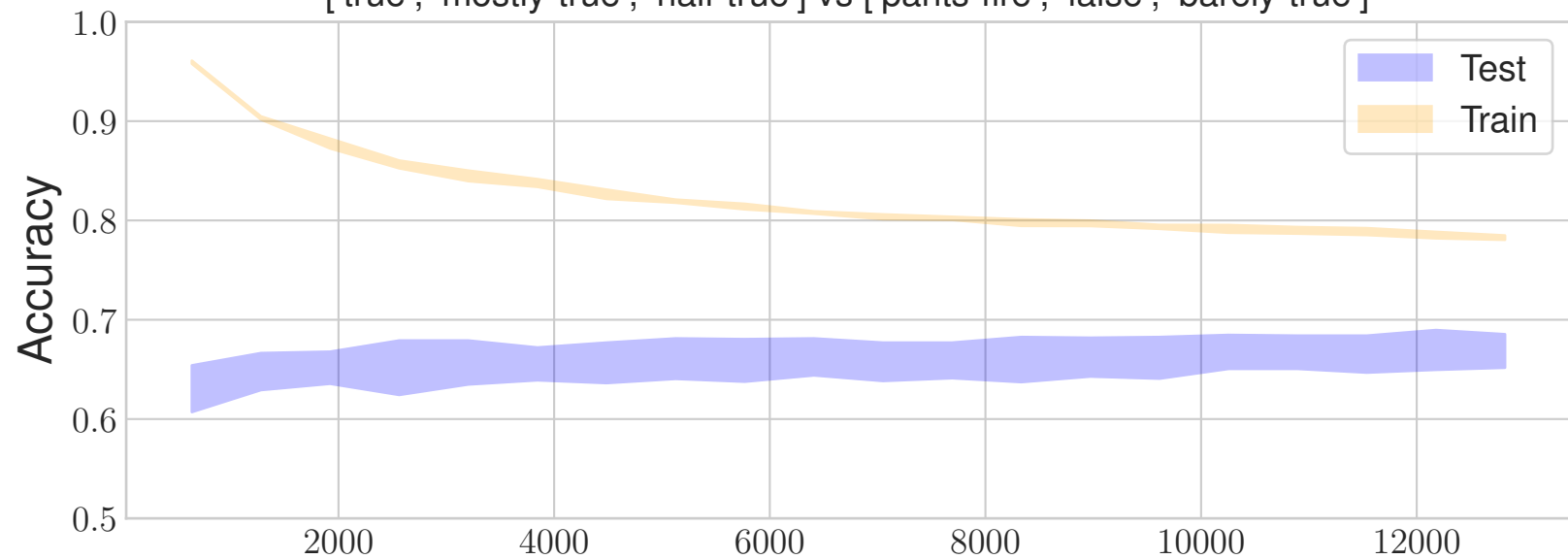
['true'] vs ['pants-fire']



['true', 'mostly-true'] vs ['pants-fire', 'false']



['true', 'mostly-true', 'half-true'] vs ['pants-fire', 'false', 'barely-true']



Train sizes