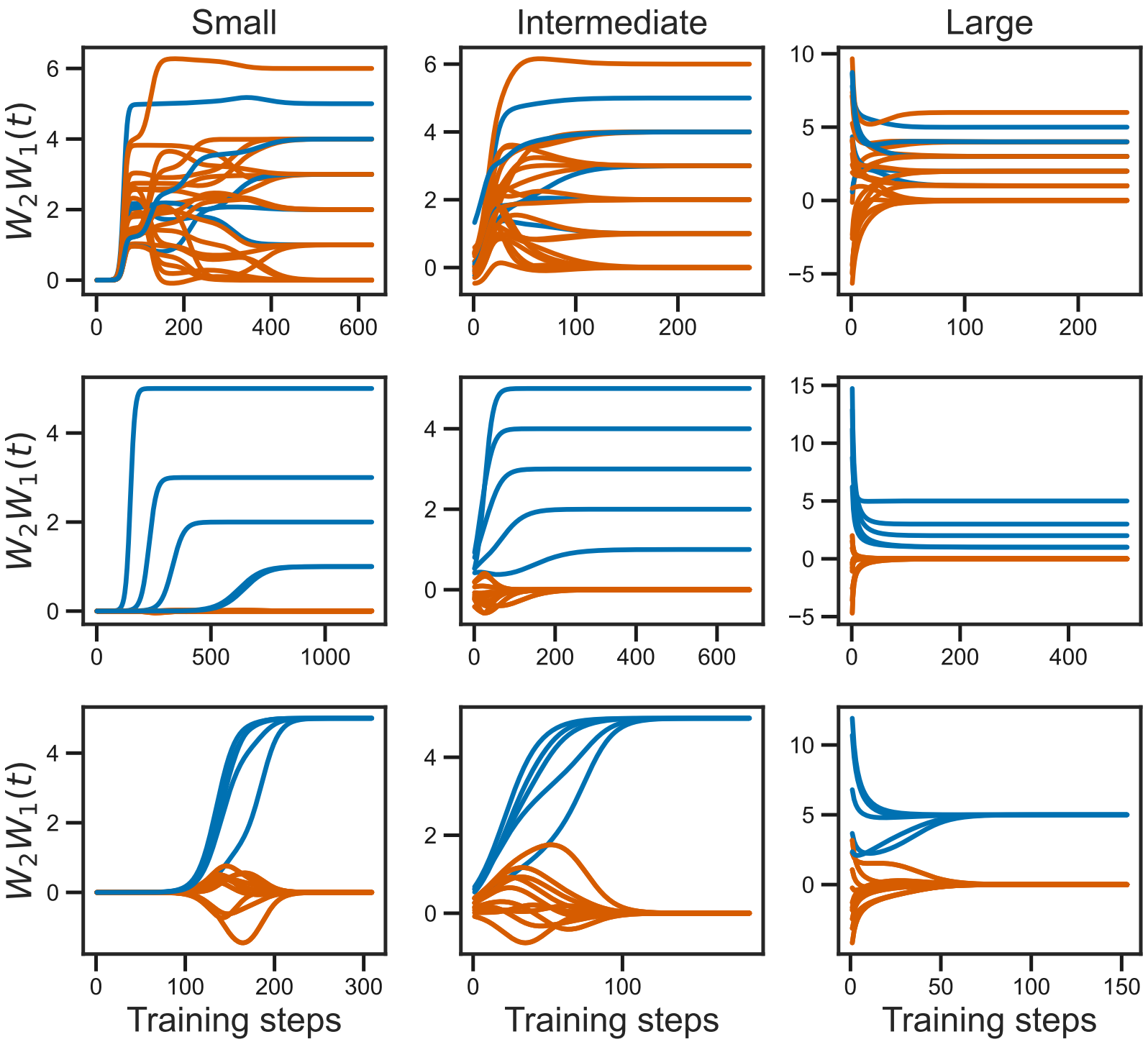
Small
Intermediate
Large
$W_2W_1(t)$
Training steps