

Figure S1: Plots of the non-dominated policies obtained by PBMORL versus MOMPO, META-MORL, MORL-Adaptation and MORAL with different preferences on each objective.

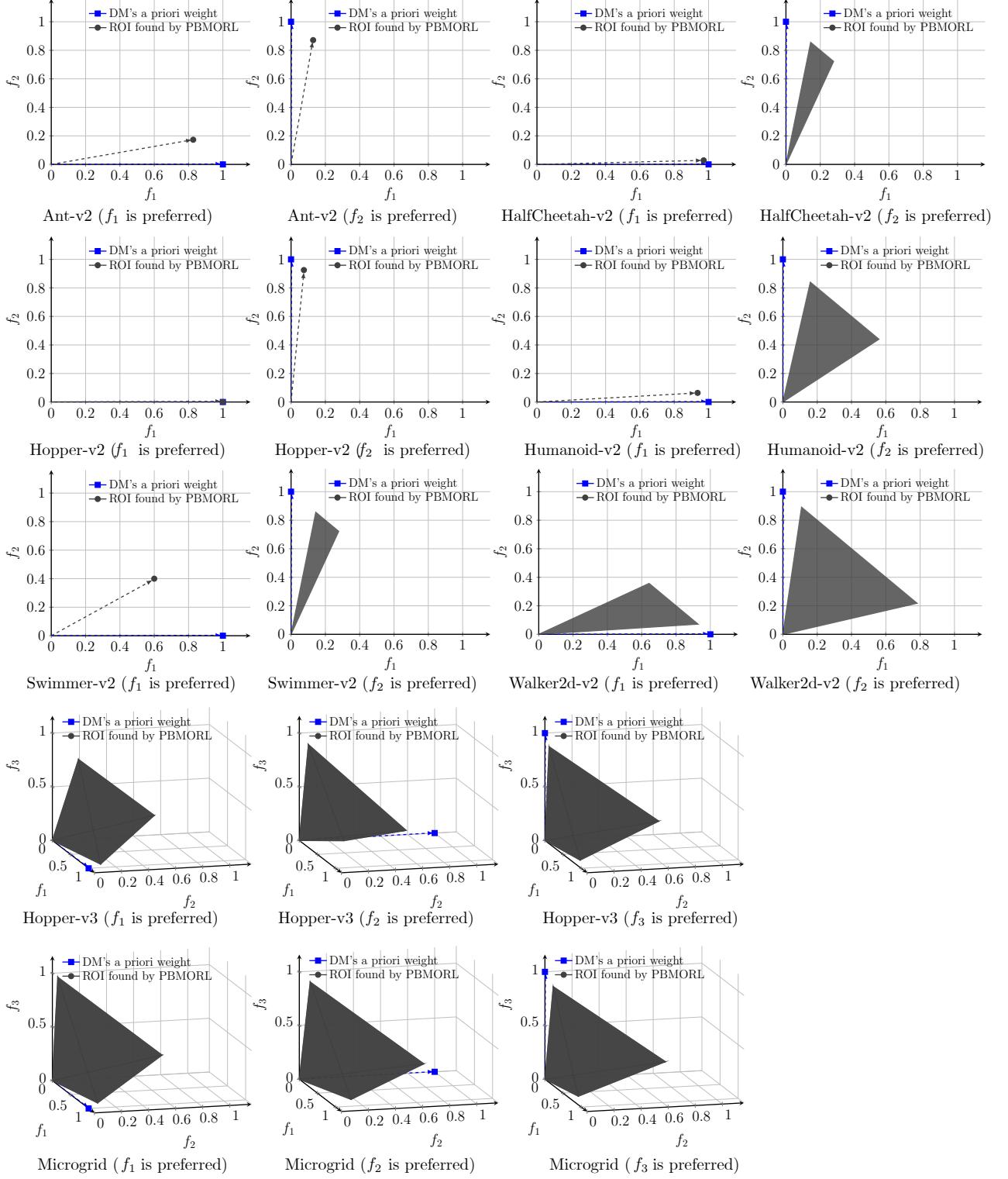


Figure S2: Comparison of the weight vector specified by the DM *a priori* versus the ROI identified by PBMORL (shaded in the gray region).

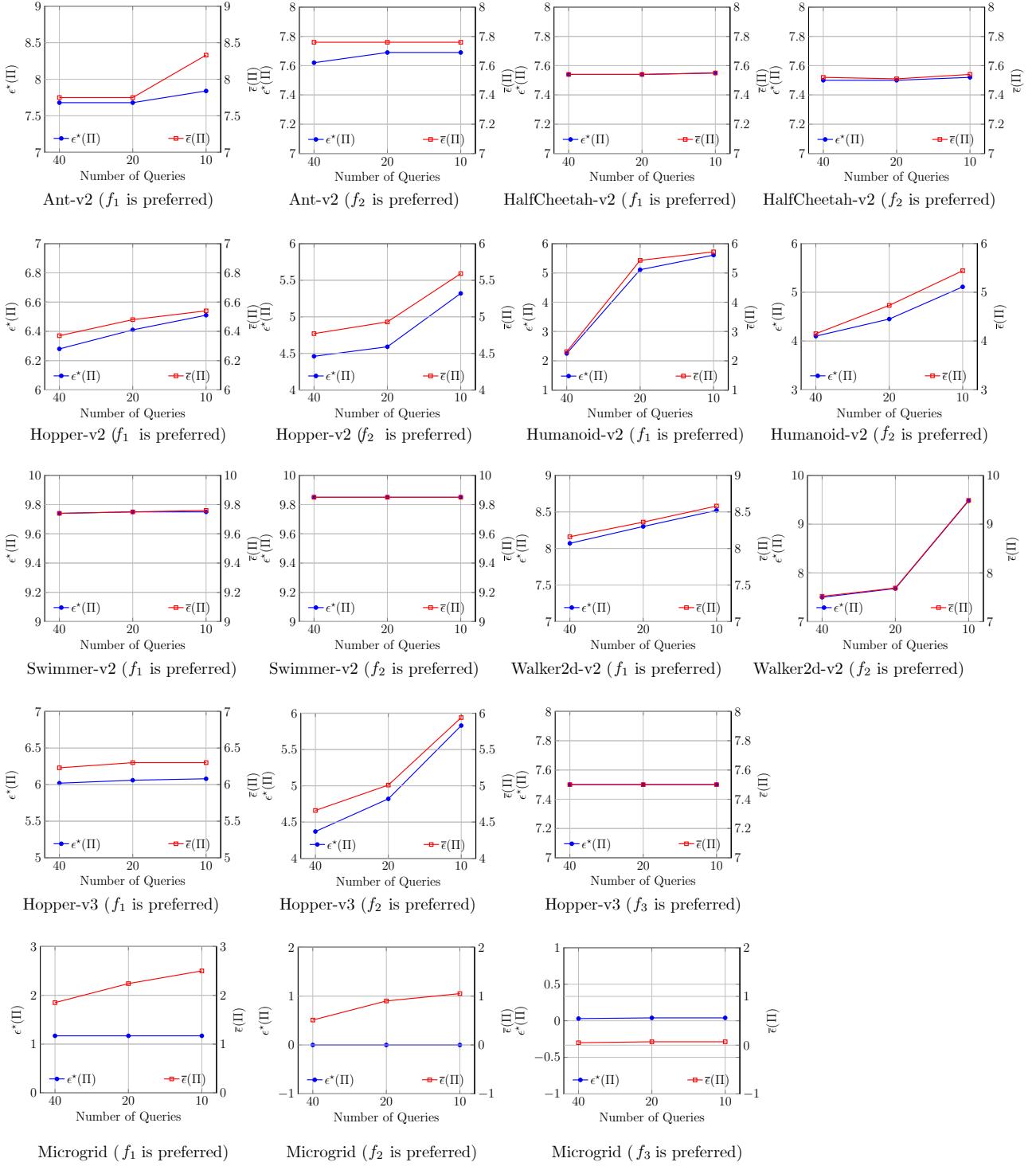


Figure S3: Comparison results of  $\epsilon^*(\Pi)$  and  $\epsilon(\Pi)$  obtained by PBMORL with 10, 20, and 40 interactions, respectively.

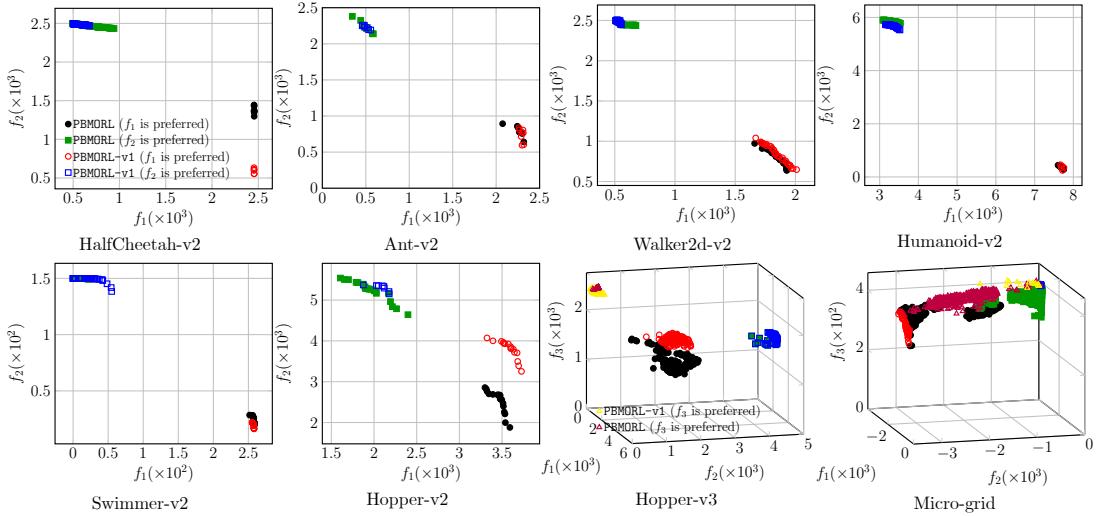


Figure S4: Plots of the non-dominated policies obtained by PBMORL versus PBMORL-TCH with different preferences on each objective.