Fig. 3.3: **A first classification forest and the effect of forest size** T**.** (**a**) Training points belonging to two classes. (**b**) Different training trees produce different partitions and thus different leaf predictors. The colour of tree nodes and edges indicates the class probability of training points going through them. (**c**) In testing, increasing the forest size T produces smoother class posteriors. All experiments were run with $D = 2$ and axis-aligned weak learners. See text for details.

data which is "different" than the training data. The larger the difference, the larger the uncertainty. Thanks to all trees being different from one another, increasing the forest size from $T = 1$ to $T = 200$ produces much smoother posteriors (fig. 3.3c_3). Now we observe higher confidence near the training points and lower confidence away from training regions of space; an indication of good generalization behaviour.

For few trees (*e.g.* $T = 8$) the forest posterior shows clear box-like artifacts. This is due to the use of an axis-aligned weak learner model. Such artifacts yield low quality confidence estimates (especially