Over the last decades, face learning was predominantly explained through the lens of the averaging hypothesis (Burton et al., 2005; Jenkins & Burton, 2011) originating from computer modelling (Kramer et al., 2018). This rationale suggests that transient variations are discounted whereas stable features are reinforced to construct a robust face representation over time. While this theory explains the advantage for the recognition of variable faces in lab-based face learning studies (Baker et al., 2017; Corpus & Orient, 2022; Ritchie & Burton, 2017) , it does not account for the demonstrated role of peripheral features in more ecological settings. Therefore, we have proposed a cost-efficient theory of face learning that postulates that face representations evolve following a coarse-to-fine prioritisation of facial information based on its relative stability over time. Following a similar parsimony principle, fixed learning contexts should foster coarse facial representations while changing contexts should encourage the identification of diagnostic facial features and a refinement of representations. To test these hypotheses, we designed a face learning experiment to check if variability of peripheral features and context lead to better recognition performance. First, participants learn eight women’s faces with fixed or changing hairstyles, presented against fixed and changing scenes while their eye movements are recorded. Participants then perform a speeded old/new task and provide confidence ratings. These allow the calculation of area under ROC curves for each condition. Based on the cost-efficient theory, we predict that inner facial features of faces learned with varying hairstyles and/or contexts will be better recognized than faces with stable peripheral features and/or bound to a single context. Indeed, varying learning situations should yield the refinement of facial representations, unlike in stable learning conditions. We expect these effects to be associated with different ratios of fixations on faces and background scenes in the different learning conditions.