**Supplementary material 2 (familiarity-based recognition in probe task)**

One could argue that, to perform the probe task, participants were relying on some kind of familiarity-based recognition rather than search-related processes in the focus of attention. This seems unlikely. It is, for example, difficult to see how familiarity would help recognition in Experiments 3 and 4, in which we used a restricted set of 8 phonologically similar letters as memory material. By the time one is judging the fourth probe, four different memory items have been seen and potentially 3 other letters (as probes) that are very similar to the items that have to be remembered. It seems unlikely that pure familiarity-based recognition would be able to distinguish between a probe letter representing a memory item and a probe letter that is very similar but is not representing a memory item; participants would be suffering from severe proactive interference from all the letters they have been seeing. Familiarity would be worthless in the face of the high degree of proactive interference. Recognition performance was, however, in all experiments, very high, indicating that people used working memory rather than familiarity to respond to the probes.

Furthermore, the use of familiarity-based recognition is not consistent with our data. It has been proposed that set size effects can be used to differentiate between probe recognition based on familiarity arising from activated representations in LTM and probe recognition based on an active search in the central region of WM. In particular, it has been proposed that the presence of set-size effects (increasing RTs as a function of the number of items in WM) can be used as a behavioral marker of the memory list being represented in the central, capacity-limited component of WM, as opposed to LTM (e.g., Burrows & Okada, 1975; Souza, Rerko, & Oberauer, 2014; see Vergauwe & Cowan, 2015, for a similar argumentation). We can examine this in our task in which the number or items in WM increases across probe positions (one item in WM when judging the first probe, two items in WM when judging the second probe, and so on). Using probe position as a proxy for the number of items in WM, we observed that RT increased as a linear function of set size, see Figure below. These functions suggest that participants engaged in an active search through the central component of WM in all four experiments, rather than making a familiarity-based decision.