**Abstract**

False alarms occur when participants report having perceived a stimulus that was, in fact, never presented. False alarms share this feature with hallucinations - vivid and transient experiences of objects, such as images or sounds, that occur in the absence of a cause in the external world - and have therefore developed into one of the most influential paradigms to study perceptual symptoms in psychosis. Yet to serve as a proxy for hallucinations, false alarms need to satisfy at least two more criteria: First, false alarm should not only represent the erroneous report of a signal, but reflect perceptual experiences that are characterized by a specific content. Second, false alarms should occur at a timescale compatible with the temporal dynamics of hallucinations, which often unfold on the order of seconds to minutes, in particular in early stages of psychotic illness. In this work, we combine Hidden Markov Models with a classification image approach to show that false alarms are more likely to occur in an internal mode of perception, a minute-long state of the brain, during which the content of perception is strongly biased toward previous experiences.