Both the amygdala and striatum are known to be critical for associative learning. For the striatum, celebrated work in humans and other animals suggests its involvement in learning from prediction errors for reinforcement[1,2]. Such errors occur when there is more or less reward (or punishment) than expected. Supporting this idea, the prediction error, as quantified in theories of conditioning such as the Rescorla-Wagner and temporal difference models, has helped to explain neural signaling in this system across species, including blood oxygenation level–dependent (BOLD) signals in the human striatum[2,3].

However, BOLD activity in the amygdala is not consistently correlated with error signals, even in aversive conditioning tasks[3]. This raises the question of how we might computationally characterize learning signals in the amygdala. Such a specific characterization could shed further light on ideas about the structure's distinct contributions to associative learning. Current theories of amygdala function in humans have highlighted its role in vigilance[4] and the detection of relevant stimuli[5]. Theories of associative learning in animals, such as the Pearce-Hall model[6], describe a more specific and potentially related function for the amygdala[7,8]: the attentional gating of learning. These theories envision that, to learn cue-reinforcer associations, animals track a quantity, known as associability, that reflects the extent to which each cue has previously been accompanied by surprise (positive or negative prediction errors). A cue's associability gates the amount of future learning about the cue on the basis of whether it has been a reliable or poor predictor of reinforcement in the past. In other words, associability controls learning rates dynamically, accelerating learning to cues whose predictions are poor and decelerating it when predictions become reliable.

In nonhuman animals, lesion studies and, more recently, unit recordings have indicated that an important neural substrate for associability is the amygdala[7–9]. To date, there is little direct evidence that the human amygdala might have an analogous role. We hypothesized that the human amygdala codes for associability, which is distinct and complementary to the striatum's coding of prediction error during associative learning. Specifically, we used a computational model to examine an aversive reversal-learning task and asked whether an associability signal similar to that seen in unit recordings in nonhuman animals might be present in the pattern of BOLD signaling in the human amygdala during aversive learning[8].

We asked 17 participants to complete a Pavlovian reversal-

(**Supplementary Tables 4** and **5**). These results leave open the question of whether associability coding in human amygdala is specific to aversive tasks or to other features of our experiment, such as the use of mildly aversive (angry) faces as conditioned stimuli. However, our findings complement previous research that used reward learning tasks in nonhuman animals and found similar roles for the amygdala and the striatum in the computation of associability and prediction error, respectively[8]