Supplemental Image Analysis

To further strengthen the claim that differences in time on task do not explain the observed effects, we estimated an additional first-level model that included a subset of trials on which response time (RT) was matched across How and Why trials. For each participant, we modeled only those How trials that showed a response time less than 1.17 standard deviations (SD) above their average RT to How trials, and only those Why trials that showed a response time more than 1.17 SDs below their average RT to Why trials. (This SD value was chosen because it minimized the RT difference across the two trials types.) Effectively, this procedure removed the fastest Why trials and the slowest How trials from the analysis. Removed trials were modeled separately in the design. As in the primary analysis, we also included the six motion parameters as covariates of no interest. Linear contrasts were then applied to the design to determine regions more active in the How > Why and Why > How contrasts. The resulting contrast images were then gray-matter masked and subjected to a random effects analysis to investigate effects at the group-level. Whole-brain interrogations were conducted using a false discovery rate (FDR) corrected p-value of .05 combined with a cluster size threshold of 5 voxels.

Supplemental Results

In this model, the average RT for How trials (2.922 s) was nearly identical to the average RT for Why trials (2.955 s). Moreover, we found that the regions observed in the primary analysis (Table 1 in the main text) continued to survive FDR correction at $p < .05$ in the whole-brain analyses, except for the ventral premotor area observed in the How > Why contrast, which survived FDR correction at $p < .09$. Given the reduction in the number of trials entering the analysis, as well as on the observation of marginal significance in ventral premotor cortex, we

conclude that all effects reported in the main text are not sufficiently explained by the RT difference.