Supplementary Materials

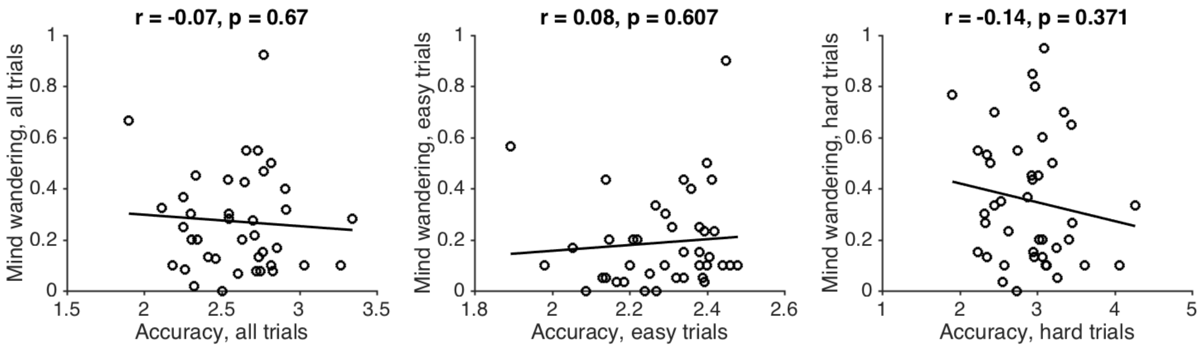

Figure S1. Correlations between mind wandering rate and task accuracy in Experiment 1. **Left:** Accuracy for all trials did not correlate with mind wandering rate for all trials. **Middle**: Mind wandering and accuracy for easy trials (Set Sizes 2 and 3). **Right**: Mind wandering and accuracy for hard trials (Set Sizes 6 and 8).

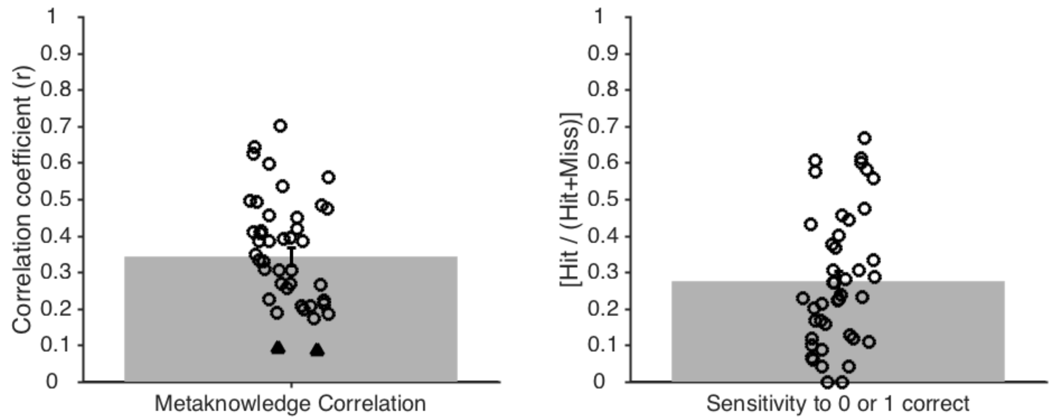

Figure S2. Individual subject values for the metaknowledge correlation analysis and the lapse sensitivity analysis in Experiment 3a. **Left**: Markers represent the individual subject r-values for the correlation between number correct and number confident on each trial. Triangles indicate that the individual's correlation value did not pass the threshold for significance (*p* <.05). The gray bar represents the average correlation value. **Right**: Markers represent individual subjects' lapse sensitivity (proportion of lapses caught). Gray bar represents the average value.

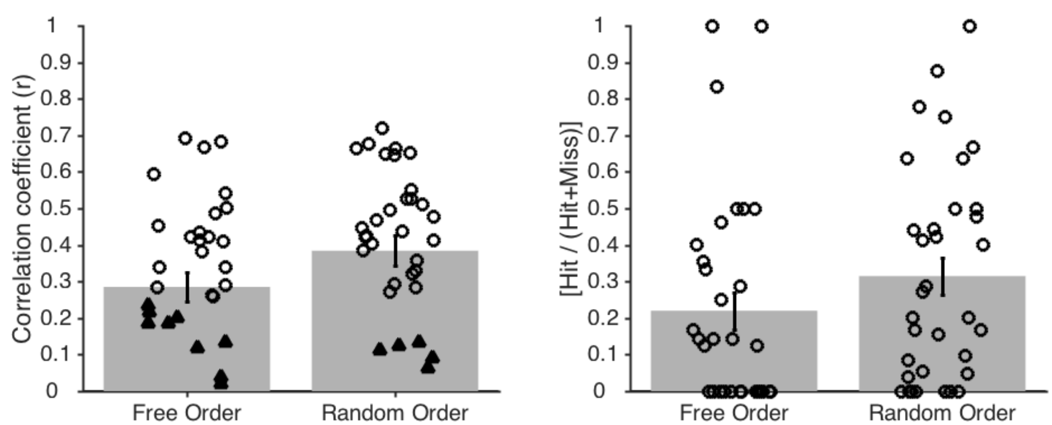

Figure S3. Individual subject values for the metaknowledge correlation analysis and the lapse sensitivity analysis in Experiment 3b, for both the free response-order condition and the random response-order condition. **Left**: Markers represent the individual subject r-values for the correlation between number correct and number confident on each trial. Triangles indicate that the individual's correlation value did not pass the threshold for significance (*p* <.05). The gray bar represents the average value. **Right**: Markers represent individual subjects' proportion of lapses caught. Gray bar represents the average value.

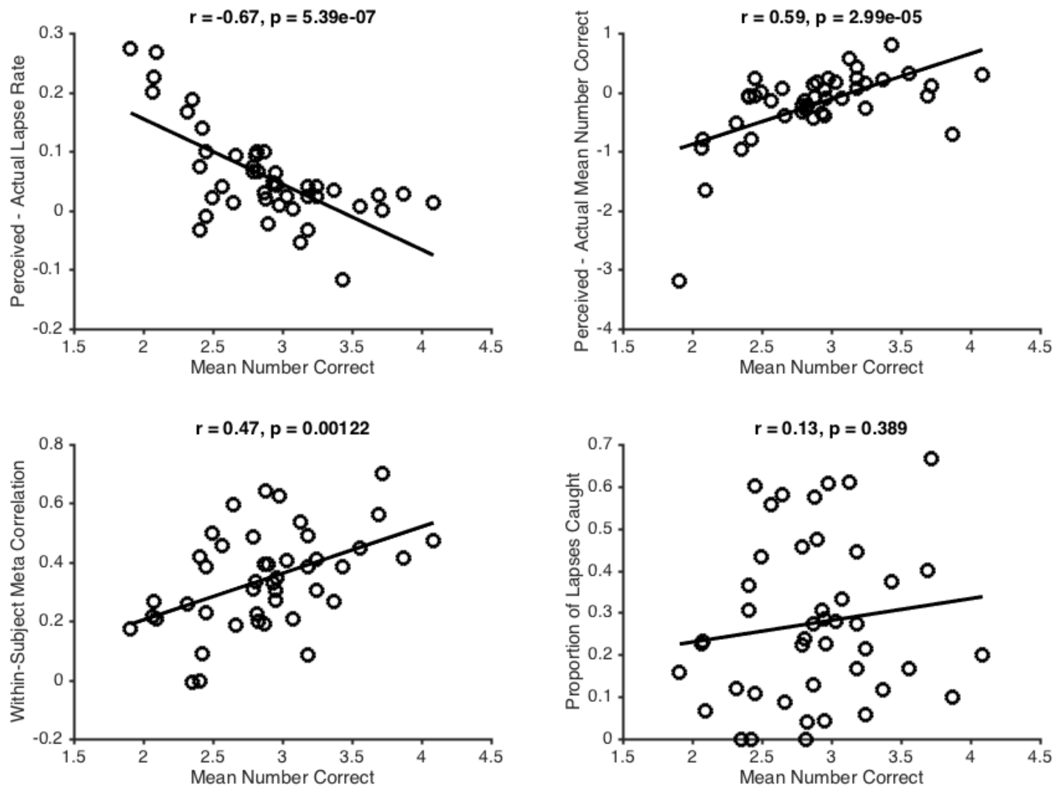

Figure S4. All metaknowledge correlations reported for Experiment 3a. Clockwise, starting at top left: (1) Dunning-Kruger correlation analysis on lapse rate (perceived lapse rate – actual lapse rate). (2) Dunning-Kruger correlation analysis on mean number correct (perceived mean number correct – actual mean number correct. (3) The relationship between average performance and lapse sensitivity (proportion of lapses caught). (4) The relationship between average performance and the metaknowledge correlation score (r value derived from correlating single-trial accuracy and single-trial confidence for each subject).

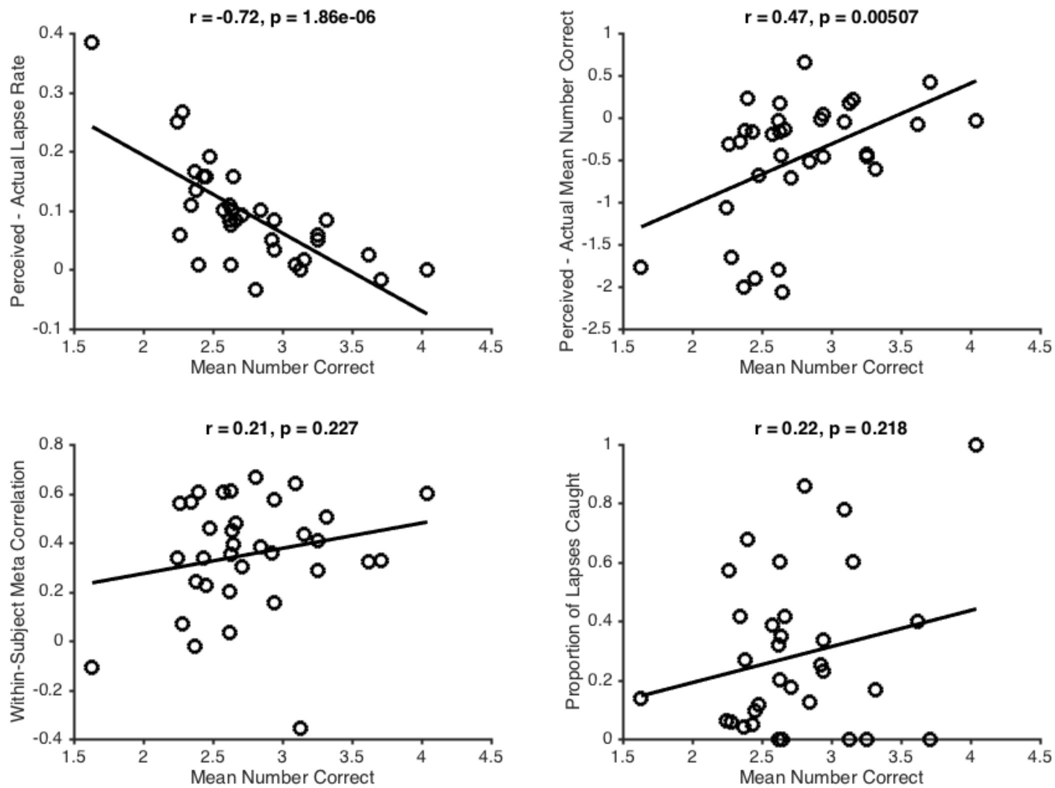

Figure S5. All metaknowledge correlations reported for Experiment 3b. Clockwise, starting at top left: (1) Dunning-Kruger correlation analysis on lapse rate (perceived lapse rate – actual lapse rate). (2) Dunning-Kruger correlation analysis on mean number correct (perceived mean number correct – actual mean number correct. (3) The relationship between average performance and lapse sensitivity (proportion of lapses caught). (4) The relationship between average performance and the metaknowledge correlation score (r value derived from correlating single-trial accuracy and single-trial confidence for each subject).

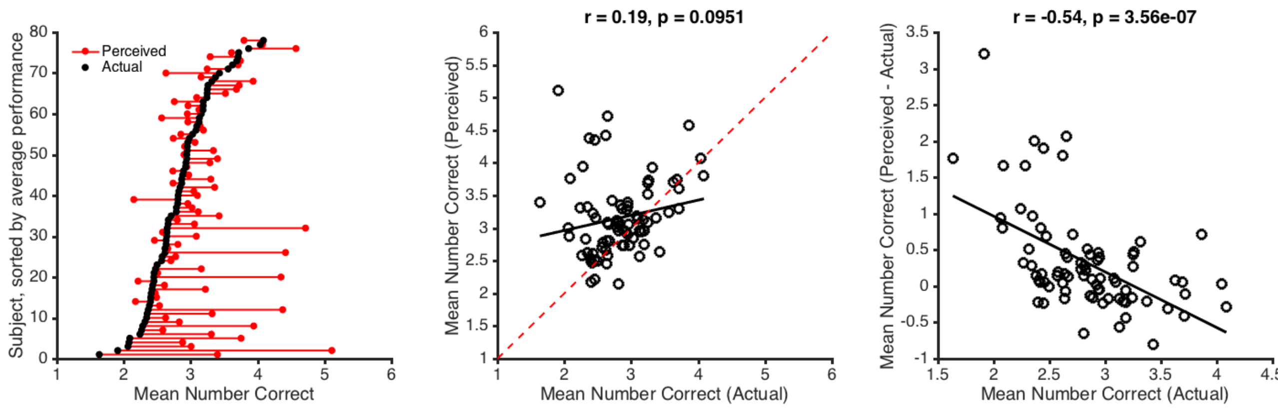

Figure S6. Perceived and actual mean performance, combined across Experiments 3a and 3b. . (Left) Black dots represent each participant's actual mean performance, red dots represent perceived mean performance. Subjects are sorted on the y-axis by actual mean performance. (Middle) Linear correlation between actual mean performance and perceived mean performance. (Right) Correlation between actual mean performance and metacognitive miscalibration (perceived – actual mean performance).

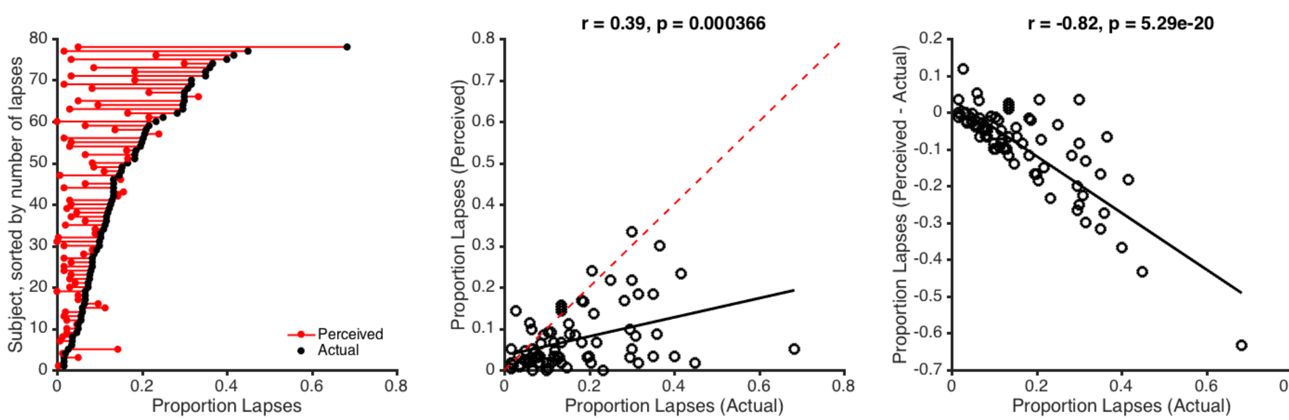

Figure S7. Perceived and actual lapse rates, combined across Experiments 3a and 3b. (Left) Black dots represent each participant's actual lapse rate, red dots represent perceived lapse rate. Subjects are sorted on the y-axis by actual lapse rate. (Middle) Correlation between actual lapse rate and perceived lapse rate. (Right) Correlation between actual lapse rate and metacognitive miscalibration (perceived – actual lapse rate).

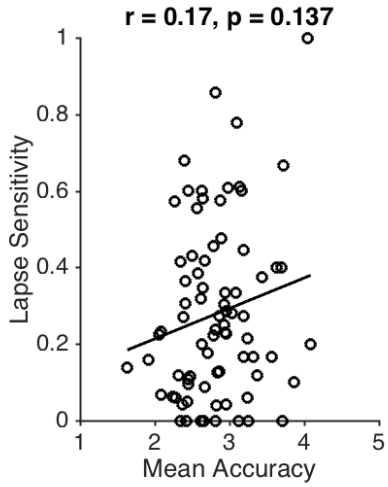

Figure S8. Correlation between lapse sensitivity metric and overall working memory performance.

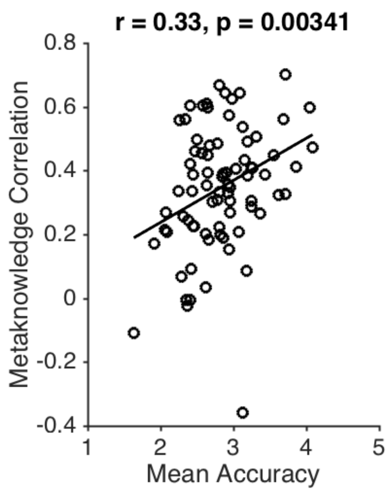

Figure S9. Correlation between trial-by-trial accuracy-confidence correlation and overall working memory performance.

Individual differences analyses (Fig S6 – S9)

To further explicate the link between perceived and actual performance, we have provided the results of some additional analyses in Figures S6 through S9. For these analyses, we combined data for all subjects across Experiments 3a and 3b.

First, we visualized the discrepancy between actual and perceived average performance (Fig. S6). As can be seen in Fig. S6, those with very low working memory performance reported some of the highest overall confidence ratings. In fact, the relationship between overall accuracy and confidence was better described by a U-shaped function (R^2 = .19, RMSE = .54) than by a linear function (R^2 = .04, RMSE = .59). Because of this non-monotonic function, the difference between perceived and actual performance was strongly correlated with a subjects' overall performance level. Second, we considered the link between perceived and actual lapse rate (Fig. S7). Perceived lapse rate was correlated with actual lapse rate, but the correlation fell below the parity line, indicating overconfidence; the distance from the parity line was larger for participants with more lapses, indicating that low-performing participants more greatly underestimated their lapse rate.

Finally, we examined the lapse sensitivity metric (Fig. S8) and the confidence-accuracy correlation metric (Fig. S9). We found that our lapse sensitivity metric (lapses caught / total lapses) did not predict average task performance (r = .17, p = .14). Perhaps the lapse sensitivity metric is not a very good one. First, relatively few trials contribute to this metric, which increases measurement error and may contribute to the measure's poor internal reliability (split-half correlation, r = .45). Second, this metric is not able to control for "baseline" rates of overconfidence. For example, a participant could always be overconfident by exactly 1 item (i.e. say that they got 2 when they got 1, that they got 3 when they got 2). Obviously, a consistent inflation of confidence this would lead to extremely poor lapse sensitivity but excellent insight into trial-by-trial fluctuations in performance. In this sense, then, our trial-by-trial correlation metric should be our strongest piece of evidence that there is any relationship between metacognitive accuracy and overall performance on the task (Figure S9). A small amount of overconfidence (e.g. higher intercept) will not greatly affect the strength of the trial-by-trial correlation between confidence and accuracy. Even though this metric normalizes across varying amounts of "global overconfidence", we still found that, this trial-by-trial correlation metric predicted average performance (r = .33, p = .003).