

Supplemental Information

Part 1. Stability of behavioral and neural measures

Behavioral measures for the spatial WM task were stable within individuals across runs for overall (across conditions) accuracies ($r=0.59$, $p<0.001$) and RTs ($r=0.83$, $p<0.001$). (**Supp. Fig. 1**). We used these overall measures, rather than the H>E measures, because i) they showed substantially greater stability within individuals ($r=0.59$ vs. $r=0.28$ for the accuracies, and $r=0.83$ vs. $r=0.41$ for the RTs), ii) they are more intuitively interpretable, and iii) the H>E measures contain a non-linearity, such that smaller between-condition behavioral differences can be observed in both high performers (when performance is closer to ceiling) and low performers (when performance is closer to floor).

MD H>E neural responses were also stable across runs for each MD ROI individually and collapsing across ROIs ($r=0.73$, $p<0.0001$) (**Supp. Fig. 2**). Thus, both behavioral and neural measures can serve as reliable individual phenotypic markers, and we are justified to probe the relationship between them.

For the specificity-assessment analysis, we used responses in the regions of the left-hemisphere fronto-temporal language network (**Supp. Part 2**). fMRI responses in the language regions were also stable for the Sentences>Nonwords contrast ($r=0.82$, $p<0.0001$), in line with Mahowald & Fedorenko³⁰.

Part 2a. Language Localizer details

All 140 participants performed a language localizer task. The majority ($n=113$) passively read sentences and lists of pronounceable nonwords in a blocked design. The Sentences>Nonwords contrast targets brain regions sensitive to high-level linguistic

processing^{24,25}. Each trial started with 100ms pre-trial fixation, followed by a 12-word-long sentence or a list of 12 nonwords presented on the screen one word/nonword at a time at the rate of 450ms per word/nonword. Then, a line drawing of a hand pressing a button appeared for 400ms, and participants were instructed to press a button whenever they saw the icon, and finally a blank screen was shown for 100ms, for a total trial duration of 6s. The button-press task was included to help participants stay awake and focused. Each block consisted of 3 trials and lasted 18s. Each run consisted of sixteen experimental blocks (eight per condition), and five fixation blocks (14s each), for a total duration of 358s (5min 58s). Each participant performed two runs. Condition order was counterbalanced across runs. The remaining 27 participants performed similar versions of the language localizer with minor differences in the timing and procedure.

Part 2b. Language fROIs definition and response estimation

To define the language fROIs, we used the Group-constrained Subject-Specific (GSS) approach²⁵. In particular, fROIs were constrained to fall within a set of functional parcels that corresponded to the expected gross locations of activations for this contrast and that were generated based on a group-level data representation from a large set of participants. We used six parcels derived from a group-level representation of data for the Sentences>Nonwords contrast in 220 participants. These parcels included three regions in the left frontal cortex: two located in the inferior frontal gyrus, and one located in the middle frontal gyrus; and three regions in the left temporal and parietal cortices spanning the entire extent of the lateral temporal lobe and going posteriorly to the angular gyrus. Within each parcel, we selected the top 10% of most responsive voxels in each individual participant based on the *t*-values for the Sentences>Nonwords contrast. To estimate the fROIs' responses to the Sentences and Nonwords conditions, we used an across-run cross-validation, as for the MD fROIs.

Supplementary figures

Figure 1. Stability of behavioral responses within individuals across runs

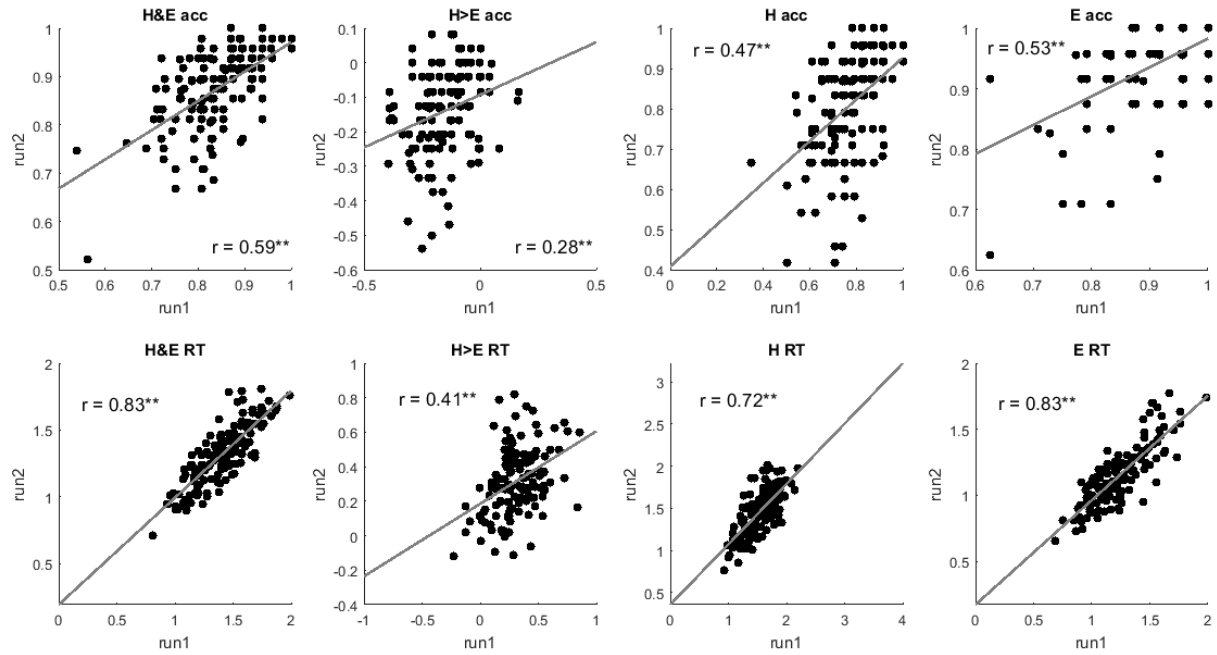
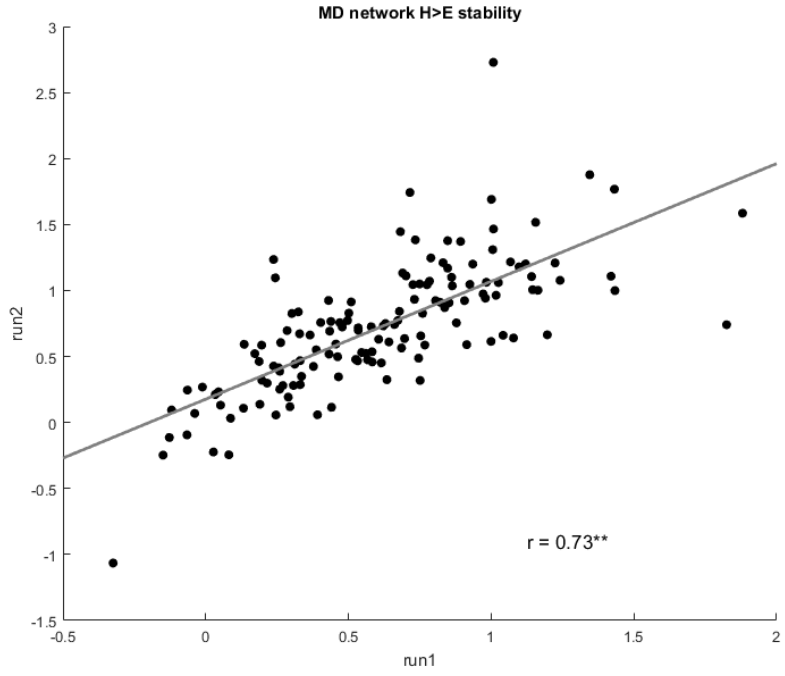


Figure 2. Stability of neural responses within individuals across runs

(a) MD network (across ROIs) Hard>Easy



(b) MD Hard>Easy

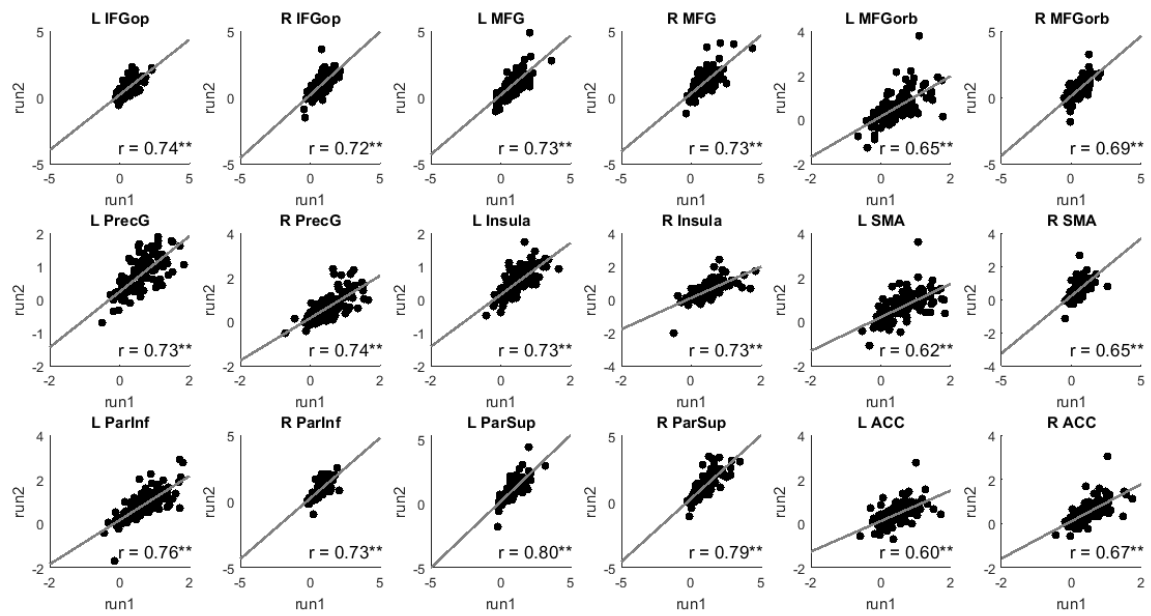
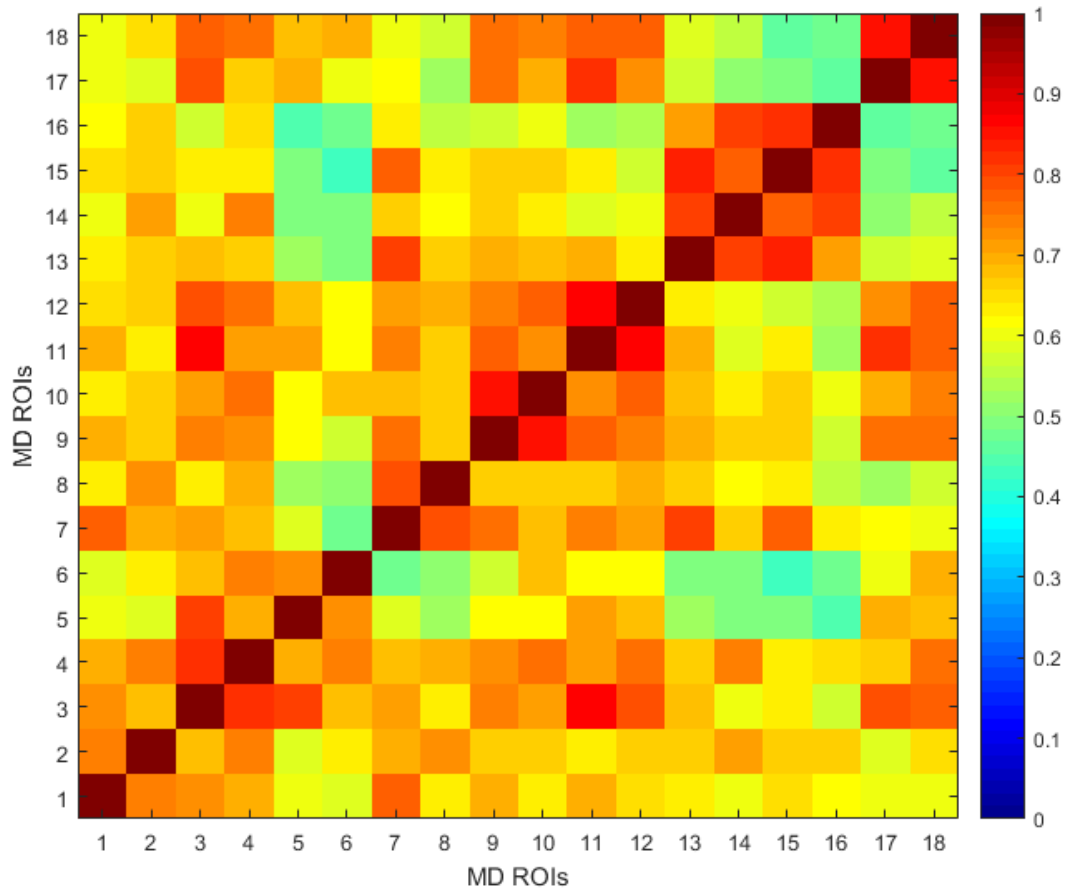


Figure 3. H>E effect size correlation across MD ROIs



Range of correlations (0.44 – 0.87). Average correlation (0.66). All correlations are significant ($p < 0.05$ Bonferroni-corrected for 153 tests).

Figure 4. MD H>E estimates predict performance for independent runs

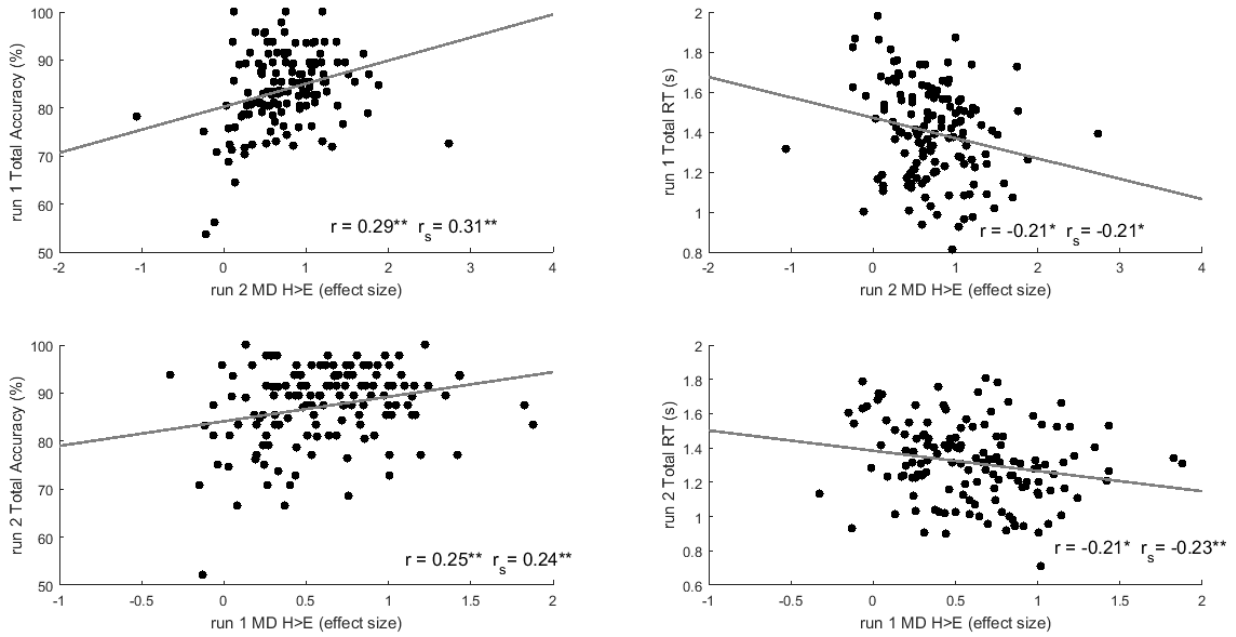


Figure 5. Brain-behavior relationships for the S>N effect in the language network vs. behavior on the spatial WM task and IQ

