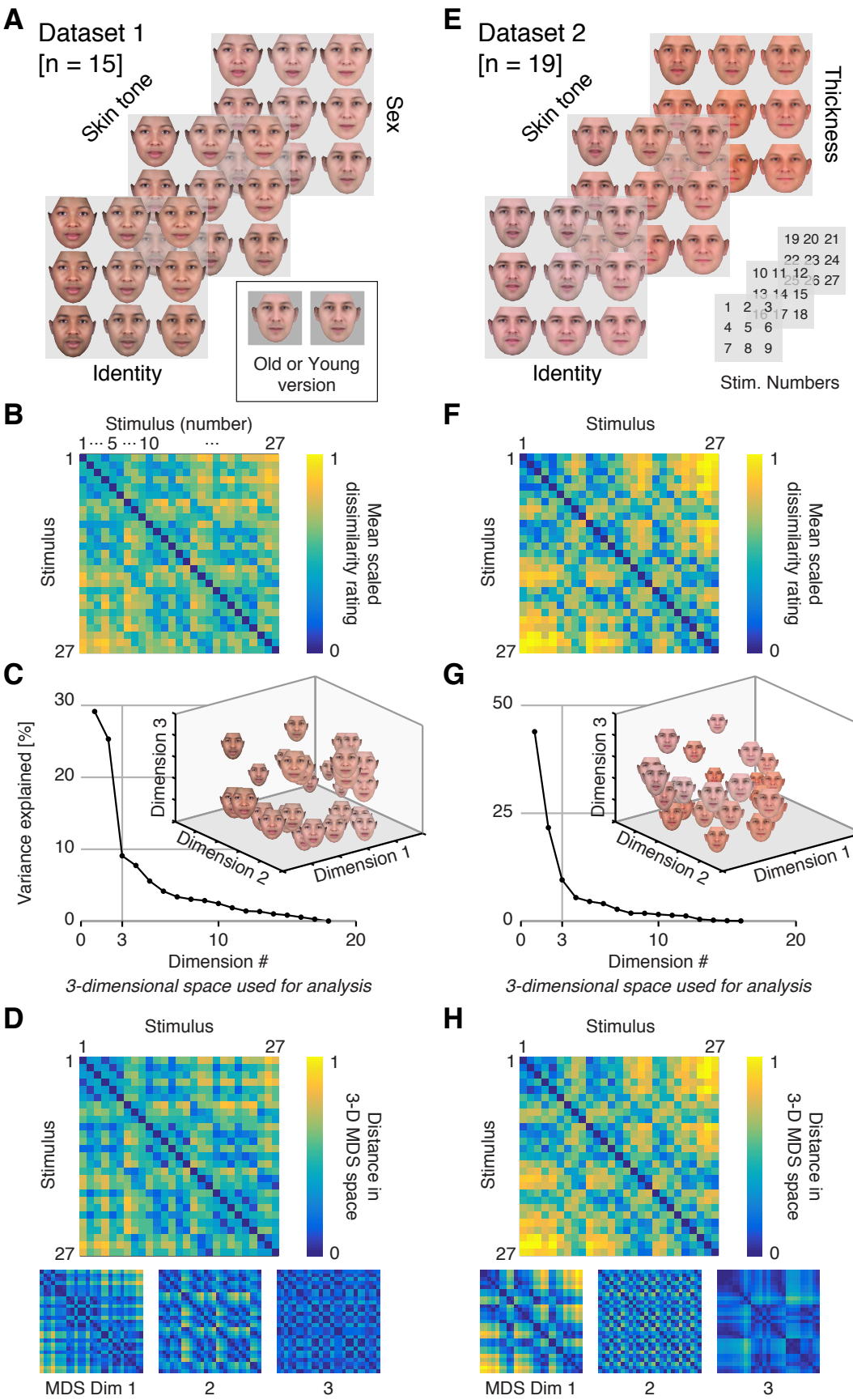
A
Dataset 1
[n = 15]
Skin tone
Sex
Identity
Old or Young
version
E
Dataset 2
[n = 19]
Skin tone
Thickness
Identity
Stim. Numbers
B
Stimulus (number)
Stimulus
Mean scaled
dissimilarity rating
F
Stimulus
Stimulus
Mean scaled
dissimilarity rating
C
Variance explained [%]
Dimension 3
Dimension 2
Dimension 1
Dimension #
3-dimensional space used for analysis
G
Dimension 3
Dimension 2
Dimension 1
Dimension #
3-dimensional space used for analysis
D
Stimulus
Stimulus
Distance in
3-D MDS space
MDS Dim 1
2
3
H
Stimulus
Stimulus
Distance in
3-D MDS space
MDS Dim 1
2
3

Figure S1. Related to Figure 1. Stimuli and perceptual similarity analysis. **(A)** The 27 face stimuli used in the first fMRI experiment (n=15) were generated using GenHead v1.2 (Genemation). Three primary stimulus axes were defined, corresponding to variation in internal facial features, skin tone, and gender. Three points were defined along each axis. A fourth dimension of stimulus variation was the age of the face, which could be set to slightly older or younger (shown inset). The manipulation of the apparent age of the faces during scanning was effectively orthogonal to sequential similarity of the faces. Across the sequence of faces presented during scanning, the manipulation of age was correlated at r=0.06 or less with any of the other dimensions of variation in the faces. **(B)** Average perceptual similarity ratings across subject for all pairings of the 27 faces. Following fMRI scanning, each subject rated the similarity of all pair-wise comparisons of the faces (numbering as in panel **E**, inset), with the pairings presented in a random order. Face pairs were presented together on a laptop LCD screen and remained until the subject supplied a rating. Ratings were given on a scale of 1 to 10 (1 being identical, 10 being completely different). Ratings were standardized within subject by subtracting the mean rating and dividing by the standard deviation of ratings. The standardized ratings were then averaged across subject, scaled between 0 and 1 for display, and presented here as a diagonally symmetric matrix. Similarity ratings provided by different subjects were strongly concordant (the average correlation of the set of ratings from one subject to the average of the remainder of the group was 0.75). Perceptual similarity ratings were also obtained in different subjects for the old and young face sets. The resulting across-subject perceptual dissimilarity matrices were very similar for the old and young face sets (r=0.95). **(C)** The across-subject similarity ratings were submitted to a multidimensional scaling (MDS) analysis in MATLAB (Mathworks, Natick, MA). The proportion of the variance explained by each additional

dimension of the MDS solution is shown. The vertical bar indicates the cutoff of 3 dimensions included, which together explained 70% of the variance in the behavioral data. **(D)** Reconstituted distance matrix for the pair-wise dissimilarity of the faces within the 3-dimensional MDS solution, and for each of the first three dimensions separately. **(E)** The corresponding measures for a second stimulus set. The same procedure was used to generate stimuli, although in this case the 3 primary stimulus axes were identity, skin tone, and face thickness. Inset legend indicates the stimulus numbers for dissimilarity matrices (panels **B**, **D**, **E**, **G**). **(F)** The across-subject (n=19) perceptual dissimilarity matrix for the second stimulus set. Again, there was a high average correlation of the ratings from any one subject to the average of all other subjects (r=0.77). **(G)** MDS analysis of across-subject perceptual dissimilarity ratings for the second stimulus set. The first three dimensions explained 75% of the variance. **(H)** Above: reconstituted distance matrix for the pair-wise dissimilarity of the faces within the 3-dimensional MDS solution, and for each of the first three dimensions separately.

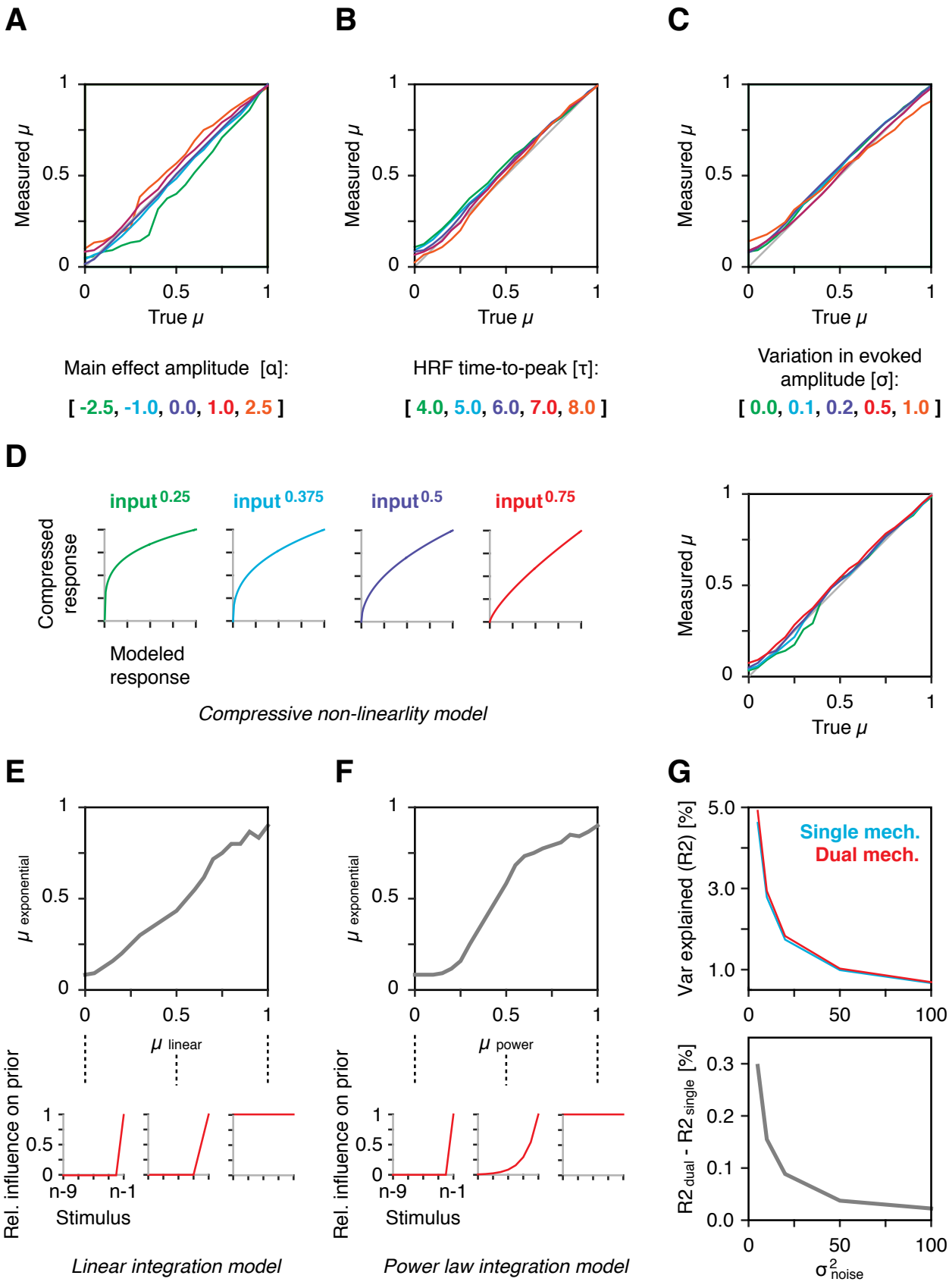
A
B
C
Measured μ
True μ
Main effect amplitude [α]:
[-2.5, -1.0, 0.0, 1.0, 2.5]
HRF time-to-peak [τ]:
[4.0, 5.0, 6.0, 7.0, 8.0]
Variation in evoked amplitude [σ]:
[0.0, 0.1, 0.2, 0.5, 1.0]
D
input 0.25
input 0.375
input 0.5
input 0.75
Compressed response
Modeled response
Compressive non-linearlity model
E
F
G
μ exponential
μ linear
μ power
Rel. influence on prior
Stimulus
n-9
n-1
Linear integration model
Power law integration model
Var explained (R2) [%]
Single mech.
Dual mech.
R2 dual - R2 single [%]
σ2 noise

Figure S2. Related to Figure 2. Model robustness evaluated by simulation. We created a forward model of BOLD fMRI signals and evaluated the performance of the measure of temporal integration. We began with the sequence of presented faces studied in Dataset 1. Given this sequence, we created a model of evoked neural activity. This model had parameters that controlled: the relative amplitude of the "main effect" (the mean response to all face stimuli as compared to the blank trials; α); the standard deviation of noisy variation in the amplitude of neural response evoked by each face (σ); an additional, additive effect of a stimulus following a blank trial (fixed at $\alpha/10$ for all simulations); the amplitude of the modulatory effect of integrated stimulus history (γ). Three different functions were available for stimulus history integration: linear, power-law, and exponential. The time-scale of the integration for each function was controlled with a single parameter (μ). Once the neural response vector was created, it was optionally subject to a compressive power-law non-linearity [S1], with the exponent (ε) serving as a parameter that controlled the degree of compression. Finally, the neural vector was convolved with a standard double-gamma HRF, with the time-to-peak serving as a final parameter of the simulation (τ). The simulated time-series created by this model was then subjected to the analysis described in the main manuscript, which in all cases assumed a fixed value of six seconds for the time-to-peak of the HRF, a linear transformation of neural activity to HRF response (i.e., no compressive non-linearity), and an exponential temporal integrator. The analysis estimated the exponential temporal integration parameter of the simulated time-series ($\mu_{exponential}$). **(A)** In the first simulation we tested our ability to recover the true exponential temporal integration value in the presence of variation in the amplitude of main effects (α; other parameters: γ=0.5, τ=6, σ=0, ε=1). The plot gives the measured temporal integration parameter obtained for a range of modeled μ values for various values of main effects. The close position of

the results from each simulation along the diagonal indicates that we can reliably recover the underlying temporal integration constant in the setting of variations in the relative amplitude and even direction of the main and modulatory effects. **(B)** Variation in the time-to-peak of the HRF used to generate the simulated data (τ; other parameters: γ=0.5, α=1.0, σ=0, ε=1). Deviation of up to 2 seconds of the assumed from the actual time to peak of the HRF had minimal effect upon the accuracy of the measured temporal integration value. **(C)** Variation in the degree of noisy variance in the amplitude of evoked response between the face stimuli (σ; other parameters: γ=0.5, α=1.0, τ=6, ε=1). **(D)** Variation in the degree of compressive non-linearity imposed on the transform of neural activity to BOLD fMRI response (ε; other parameters: γ=0.5, α=1.0, τ=6, σ=0). **(E)** We created simulated data using a linear integration of stimulus history, $f(x) = (1-\mu_{linear}) \cdot x$, where the model parameter μ_{linear} was varied systematically. The analysis continued to assume an exponential temporal integration function (other parameters: γ=0.5, α=1.0, τ=6, σ=0, ε=1). **(F)** Data simulated using a power-law temporal integration function, $f(x) = x^{(1000(1-\mu_{power})/\mu_{power})}$. The analysis continued to assume an exponential temporal integration function (other parameters: γ=0.5, α=1.0, τ=6, σ=0, ε=1). **(G)** We generated simulated data combining two time-scales of the stimulus integration (μ=0 and μ=1) and used the cross-validation analysis procedure described in the main text to compare the variance explained (R2) by the both dual and single mechanism models (top panel). To determine whether the cross-validation procedure over-penalize model complexity under high noise conditions, we added uncorrelated Gaussian noise to the simulated neural responses with different amounts of variance (σ^2_{noise}; other parameters: $\gamma_{\mu=0}$=0.5, $\gamma_{\mu=1}$=0.5, α=1.0, τ=6, σ=0, ε=1). Increasing the level of noise worsens the estimation of the variance explained by a one or two mechanism model, but does not bias the expectation (bottom panel).

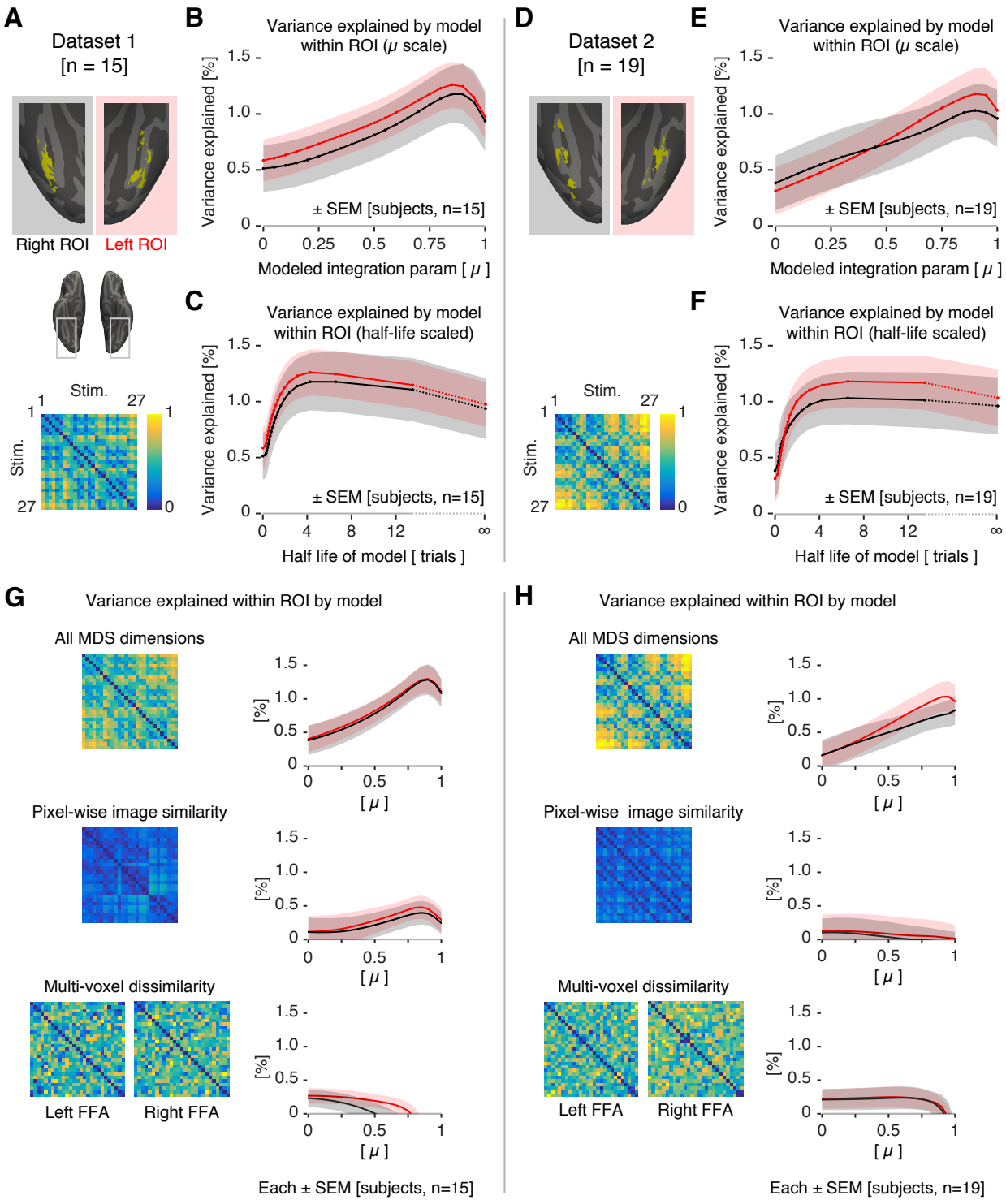
A
Dataset 1
[n = 15]
Right ROI
Left ROI
Stim.
Stim.
1
27
1
27
0
B
Variance explained by model
within ROI (μ scale)
Variance explained [%]
1.5
1.0
0.5
0
± SEM [subjects, n=15]
0
0.25
0.5
0.75
1
Modeled integration param [μ]
C
Variance explained by model
within ROI (half-life scaled)
Variance explained [%]
± SEM [subjects, n=15]
0
4
8
12
∞
Half life of model [trials]
D
Dataset 2
[n = 19]
Stim.
E
Variance explained by model
within ROI (μ scale)
± SEM [subjects, n=19]
Modeled integration param [μ]
F
Variance explained by model
within ROI (half-life scaled)
± SEM [subjects, n=19]
Half life of model [trials]
G
Variance explained within ROI by model
All MDS dimensions
Pixel-wise image similarity
Multi-voxel dissimilarity
Left FFA
Right FFA
[%]
[μ]
Each ± SEM [subjects, n=15]
H
Variance explained within ROI by model
All MDS dimensions
Pixel-wise image similarity
Multi-voxel dissimilarity
Left FFA
Right FFA
Each ± SEM [subjects, n=19]

Figure S3. Related to Figure 2. Fusiform face area (FFA), region of interest analyses analysis. **(A)** The FFA was defined for each hemisphere in Dataset 1 by a group (n=15) contrast of the main effect of stimuli versus blank screen, cropped using a FreeSurfer mask for the fusiform gyrus, and narrowed to the top 400 vertices (approximately 400 mm^2 of surface area). Shown below is the stimulus similarity matrix used in the analyses presented in panels **B** and **C**. **(B)** Average across-subject fits to Dataset 1 for the range of models within the FFA (right hemisphere in black, left hemisphere in red). The plot for the right hemisphere is the same as that in Figure 2B of the main text. **(C)** The data from panel **B**, now re-plotted with the x-axis reflecting the half-life (in stimuli) of the modeled exponential integration function. **(D)** The FFA defined in each hemisphere for Dataset 2 by group (n=19) contrast of the main effect of stimuli versus blank screen, cropped using a FreeSurfer mask for the fusiform gyrus, and narrowed to the top 400 vertices (approximately 400 mm^2 of surface area). **(E)** Average across-subject fits to Dataset 2 for the range of models within the FFA (right hemisphere in black, left hemisphere in red). **(F)** The data from panel **e**, now re-plotted with the x-axis reflecting the half-life (in stimuli) of the modeled exponential integration function. **(G)** Average across-subject fits for Dataset 1 as in **B**, presented for three different stimulus similarity spaces underlying the temporal integration models. Each stimulus space is shown to the left of the resulting data. *Top*: This analysis assumed a perceptual space based on all dimensions of the MDS reconstruction (which is the same as the across-subject average perceived similarity). The proportion of variance explained by the single mechanism model was greater than the variance explained by the dual model in the left hemisphere (variance explained by a single mechanism model: 1.28% ± 0.21% SEM; variance explained by the dual model: 1.02% ± 0.19% SEM; difference paired t-test: $t(14)=2.94$, $p=0.0107$), as well as in the right hemisphere (variance explained by a single mechanism model:

1.27% ± 0.22% SEM; variance explained by the dual model: 0.98% ± 0.20% SEM; difference paired t-test: $t(14)=3.63$, $p=0.0027$). When tests were conducted as the number of subjects with a better fitting model, in both hemispheres, the single mechanism model explains more variance than the combined model in 13 out of the 15 subjects (one-sided binomial test: $p=0.0037$). *Middle*: The assumed stimulus similarity space was derived from pixel-wise similarity of the raw stimulus images. This similarity model performs poorly in accounting for modulatory effects in the FFA. *Bottom*: The assumed stimulus similarity space was derived from the correlation of multi-voxel patterns of evoked response within the FFA (a "representational similarity analysis" [S2]). The average response to each of the 27 stimuli was obtained for each voxel within the FFA in each hemisphere. A stimulus similarity matrix was constructed as 1 minus the Pearson correlation coefficients for the patterns evoked by each face pair. While the resulting similarity matrices corresponded between hemispheres ($r=0.85$), they were weakly related to the measured perceptual similarity of the stimuli ($r=0.00$ and 0.05 for left and right hemispheres, respectively). This similarity matrix explained little positive variance in the data, consistent with prior observations that the voxel-wise structure derived from higher visual areas poorly fits within-category image variation [S3] **(H)** Analyses with alternate stimulus similarity spaces for Dataset 2, following the same conventions described for panel **G**. *Top*: The proportion of variance explained by the single mechanism model was not significantly greater than the variance explained by the dual model in the left hemisphere (variance explained by a single mechanism model: 1.03% ± 0.23% SEM; variance explained by the dual model: 0.92% ± 0.26% SEM; difference paired t-test: $t(18)=1.35$, $p=0.1942$), but was in the right hemisphere (variance explained by a single mechanism model: 0.83% ± 0.22% SEM; variance explained by the dual model: 0.79% ± 0.23% SEM; difference paired t-test: $t(18)=2.55$, $p=0.0202$). When tests were

conducted as the number of subjects with a better fitting model, in the left hemisphere, the single mechanism model explains more variance than the combined model in 11 out of the 19 subjects (one-sided binomial test: p=0.3238); in the right hemisphere, the single mechanism model explains more variance than the combined model in 13 out of the 19 subjects (one-sided binomial test: p=0.0835). *Middle, Bottom*: As these models provide little explanatory power for positive modulations in the FFA, we do not perform model comparisons using them.

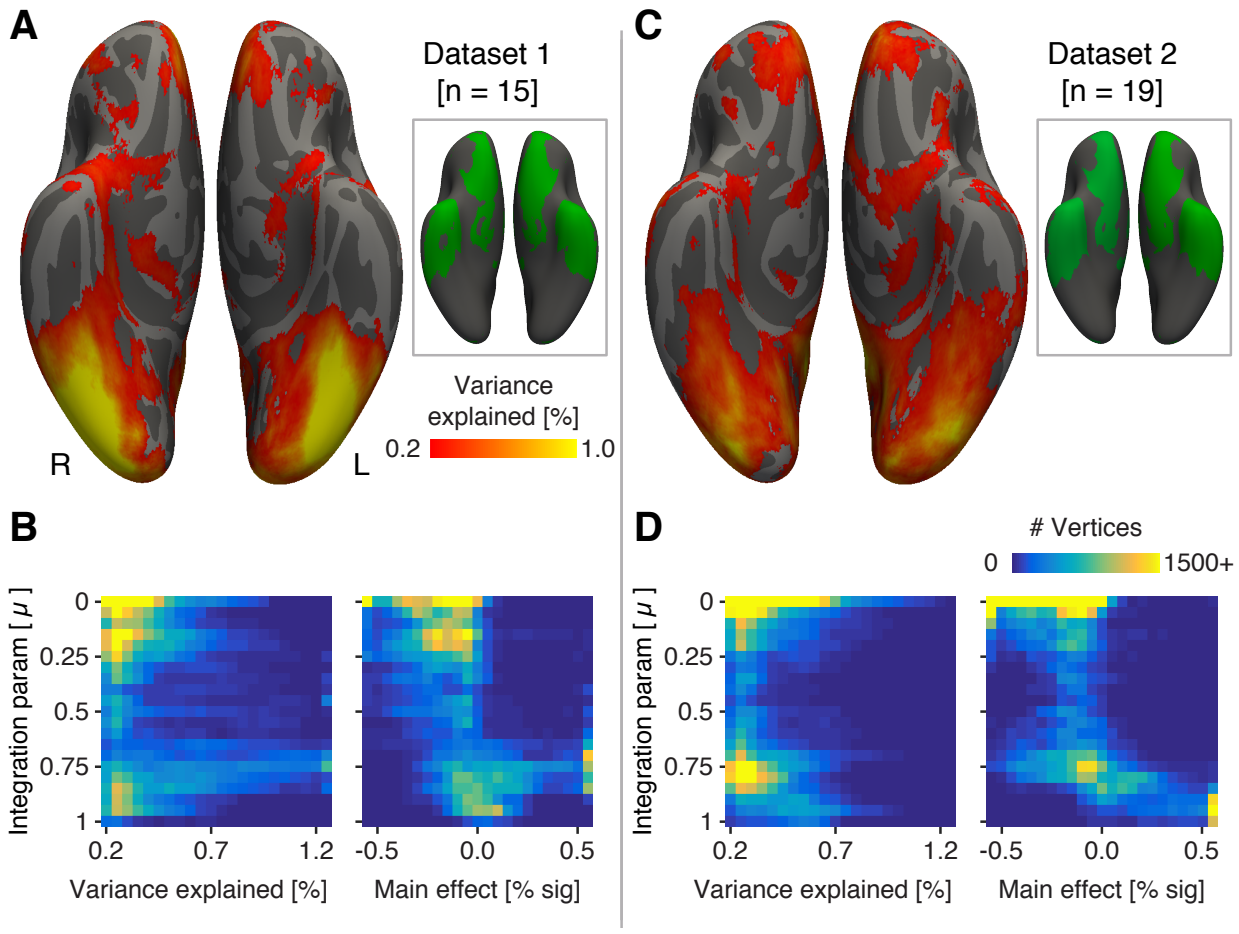

Figure S4. Related to Figure 4. Stimulus history effects on the cortical surface. **(A)** The percentage of variance explained in Dataset 1 by the best fitting temporal integration model at each vertex, thresholded at 0.2%. Shown inset in green are regions with absolute, across subject echoplanar signal values below 500, corresponding to areas with high magnetic susceptibility and thus poorly studied with BOLD fMRI. **(B)** LEFT: Number of vertices binned by their measured temporal integration parameter (μ) and their average, across-subject percentage variance explained. RIGHT: Number of vertices binned by their measured temporal integration parameter (μ) and their average, across-subject main effect of visual stimulation. In both panels, only vertices in which the variance explained by stimulus history was 0.2% or greater are shown. There is little systematic relationship across the cortex between the timescale of temporal integration and the variance explained. **(C, D)** The corresponding measures for Dataset 2.

Supplemental References

S1. Kay, K. N., Winawer, J., Rokem, A., Mezer, A., & Wandell, B. A. (2013). A two-stage cascade model of BOLD responses in human visual cortex. PLoS Comput Biol, 9(5), e1003079.

S2. Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2, 4.

S3. Drucker, D. M., & Aguirre, G. K. (2009). Different spatial scales of shape similarity representation in lateral and ventral LOC. Cerebral Cortex, 10, 2269–2280.