

Human electrocorticography reveals a common neurocognitive pathway for mentalizing about the self and others

Kevin Tan

UCLA <https://orcid.org/0000-0003-4299-996X>

Amy Daitch

Stanford

Pedro Pinheiro-Chagas

Stanford

Kieran Fox

Stanford University <https://orcid.org/0000-0002-6923-4580>

Josef Parvizi

Stanford University <https://orcid.org/0000-0001-8520-7948>

Matthew Lieberman (✉ mdlieber99@gmail.com)

UCLA

Article

Keywords: Theory of Mind, Default Mode Network, Neuronal Population Activity, Functional Specialization

Posted Date: March 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-257986/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on April 8th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-29510-2>.

1 **Human electrocorticography reveals a common neurocognitive pathway**
2 **for mentalizing about the self and others**

3
4
5 Kevin M. Tan¹, Amy L. Daitch², Pedro Pinheiro-Chagas², Kieran C.R. Fox^{2,3}, Josef
6 Parvizi^{2,3}, Matthew D. Lieberman¹

7
8
9 ¹ Social Cognitive Neuroscience Laboratory, Department of Psychology, University of California,
10 Los Angeles, CA, USA

11
12 ² Laboratory of Behavioral and Cognitive Neuroscience, Department of Neurology and
13 Neurological Sciences, Stanford University, CA, USA

14
15 ³ School of Medicine, Stanford University, CA, USA
16
17

18 Counts:

- 19 • Title = 14
20 • Abstract = 150
21 • Main text (intro, results, discussion) = 4523
22 • Methods = 2000
23 • References = 97
24

1 **Abstract**

2 Hundreds of neuroimaging studies show that mentalizing (i.e., theory of mind) recruits
3 default mode network (DMN) regions with remarkable consistency. Nevertheless, the social-
4 cognitive functions of individual DMN regions remain unclear, perhaps due to the limited
5 spatiotemporal resolution of neuroimaging. We used electrocorticography (ECoG) to record
6 neuronal population activity while 16 human subjects judged the psychological traits of
7 themselves and others. Self- and other-mentalizing recruited near-identical neuronal
8 populations in a common spatiotemporal sequence: activations were earliest in visual cortex,
9 followed by temporoparietal DMN regions, and finally medial prefrontal cortex. Critically, regions
10 with later activations showed greater functional specificity for mentalizing, greater self/other
11 differentiation, and stronger associations with behavioral response times. Moreover, other-
12 mentalizing evoked slower and lengthier activations than self-mentalizing across successive
13 DMN regions, suggesting temporally extended demands on higher-level processes. Our results
14 reveal a common neurocognitive pathway for self- and other-mentalizing that follows a
15 hierarchy of functional specialization across DMN regions.

16

1 **Introduction**

2 Humans are social by nature: our central nervous systems have evolved many
3 mechanisms to support our rich and complex social worlds¹. Although high levels of sociality
4 are seen throughout the animal kingdom^{2,3}, humans are exceptional in their capacity for
5 mentalizing – the ability to consider the mental states of others and oneself^{4,5}. The field of
6 social neuroscience seeks to understand how mentalizing and other social functions are
7 implemented at the level of brain and biology⁶. In humans, social neuroscience primarily relies
8 on functional magnetic resonance imaging (fMRI), a neuroimaging modality with higher
9 anatomical resolution but low temporal resolution⁷. Hundreds of fMRI studies have shown that
10 mentalizing recruits default mode network (DMN) regions – including medial prefrontal cortex
11 (mPFC), temporoparietal junction (TPJ), and posteromedial cortex (PMC) – with remarkable
12 consistency across countless mentalizing tasks instantiated in various sensory modalities^{5,8-12}.
13 Nevertheless, the specific social cognitive functions of individual DMN regions remain unclear.
14 When seen through fMRI, DMN regions appear to respond concurrently, yet electrophysiological
15 studies demonstrate that critical neurocognitive dynamics occur at millisecond timescales
16 throughout the DMN¹³. Thus, the limited temporal resolution of fMRI may preclude more precise
17 neurocognitive accounts of mentalizing and its component processes.

18 Several studies have investigated the fast spatiotemporal dynamics of mentalizing using
19 source-space electroencephalography (EEG) and magnetoencephalography (MEG),
20 neuroimaging modalities with high temporal resolution but coarser spatial resolution¹⁴. These
21 studies reveal a general spatiotemporal sequence of brain recruitment, starting in visual cortex,
22 followed by mirror neuron system regions (MNS; e.g., intraparietal sulcus, premotor cortex), and
23 lastly DMN regions¹⁵⁻²⁰. These findings exemplify the emerging consensus that visual
24 representations are used by MNS to identify observable actions (e.g., grasping for food), which
25 are then used by DMN to infer unobservable mental states (e.g., hunger)^{8,21-24}. Taken together,

1 MEG/EEG studies of mentalizing suggest that visual cortex, MNS, and DMN act as a hierarchical
2 neurocognitive pathway that transforms low-level visual inputs into high-level mentalistic
3 inferences. However, despite broad agreement at the network level, these studies report
4 inconsistent recruitment across individual DMN regions. These inconsistencies may reflect
5 limitations in MEG/EEG source localization, particularly in deeper regions such as mPFC and
6 PMC²⁵, which were not sampled in many of these studies. As such, the spatiotemporal
7 dynamics of mentalizing processing across individual DMN regions remains unclear.

8 Leveraging the benefits of human intracranial electrophysiology²⁶, we sought a more
9 spatiotemporally precise and mechanistic account of mentalizing by exploring neuronal
10 population activity across individual DMN regions and beyond. We show that self- and other-
11 mentalizing recruit a common neurocognitive pathway characterized by complex and
12 hierarchical processing dynamics. Our findings demonstrate that high temporal resolution
13 methods can provide critical insights on the neural mechanisms of human social cognition.

15 **Results**

16 *Data and design*

17 We recruited sixteen human subjects who had electrocorticography (ECoG) electrodes
18 neurosurgically implanted onto the cortical surface for epilepsy monitoring and treatment
19 (Supplementary Table 2). Our behavioral task (Fig. 1a) consisted of true/false text prompts for
20 three experimental conditions of interest: self-mentalizing (e.g. “I am honest”), other-mentalizing
21 (e.g. “my neighbor is honest”), and a control cognitive task involving simple arithmetic (e.g.
22 “9+86=95”). ECoG was recorded from all 2125 electrode sites in our subject cohort (Fig. 1b).
23 Sites showing epileptic activity were excluded from primary analyses. We analyzed the high-
24 frequency broadband signal (HFB; 70-180 Hz), which reflects the aggregate spiking of neuronal
25 populations immediately adjacent to an electrode site^{26,27}.

1 We began by parcellating the brain into seven regions-of-interests (ROIs; Fig. 1c) using
2 each subject's native-space cortical surface reconstruction (see Methods). We included six
3 DMN ROIs that are strongly implicated in mentalizing¹⁰: temporoparietal junction (TPJ), anterior
4 temporal lobe (ATL), posteromedial cortex (PMC), anteromedial prefrontal cortex (amPFC),
5 dorsomedial prefrontal cortex (dmPFC), and ventromedial prefrontal cortex (vmPFC). Visual
6 cortex was included as a control ROI. Out of 2125 electrode sites, 555 were included in our
7 ROIs.

8 We examined HFB responses at each site in two ways: single-trial and trial-averaged
9 analyses. Single-trial analysis examined HFB responses during individual trials relative to the
10 pre-stimulus baseline preceding each trial (Fig. 1de). Single-trial analysis captured four key
11 metrics of the HFB response: onset, peak, and offset latencies, as well as peak power ($p_{\text{FDR}} < .05$;
12 corrected for number of timepoints, trials, and sites; see Fig. 1e). Trial-averaged analysis used
13 linear mixed-effects models (LMEMs) to generate mean timecourses of HFB responses to each
14 task condition relative to pre-stimulus baseline (Fig. 1f). Trial-averaged analysis identified sites
15 with significant activations or deactivations for each task condition ($p_{\text{FDR}} < .05$; corrected for
16 number of timepoints and sites). See Figure 2 for trial-averaged analysis of exemplar ROI sites.

17

18 *A gradient of functional specificity for mentalizing from visual cortex to mPFC*

19 To explore the fast spatiotemporal dynamics of mentalizing processing, we first
20 examined its functional-anatomical foundations. To this end, we began by identifying sites that
21 showed any significant response to self- or other-mentalizing, regardless of functional
22 specificity (Fig. 3ab). Using trial-averaged results, sites were considered 'mentalizing-active'
23 (light turquoise) or 'mentalizing-deactive' (orange) if they produced higher or lower HFB power,
24 respectively, relative to pre-stimulus baseline ($p_{\text{FDR}} < .05$). Sites were considered 'mentalizing-
25 nonresponsive' if they produced non-significant HFB responses to mentalizing. We found

1 mentalizing-active sites in nearly all parts of cortex (Fig. 3a). In contrast, mentalizing-deactive
2 sites were rarer and concentrated in somatomotor and executive regions. Overall, the majority
3 of sites throughout the brain were mentalizing-nonresponsive (57% whole-brain; Fig. 3b).

4 Next, we examined the functional specificity of mentalizing-active sites (Fig. 3cd). Using
5 trial-averaged results, we identified which mentalizing-active sites also produced significant
6 ($p_{\text{FDR}} < .05$) HFB responses to the cognitive task (arithmetic). Sites were considered 'mentalizing-
7 specific' (light + dark turquoise) if they were active for mentalizing but were nonresponsive or
8 deactive for arithmetic. Sites were considered 'non-specific' (pink) if they coactivated for
9 mentalizing and arithmetic. At the whole-brain level, most mentalizing-active sites were non-
10 specific (60%), while the remaining mentalizing-specific sites (40%) were unevenly distributed
11 across cortex (Fig. 3cd). Within our ROIs, the lowest mentalizing-specificity was found in visual
12 cortex (3%). Intermediate mentalizing-specificity was found in TPJ, ATL, and PMC (range: 57-
13 58%). Very high mentalizing-specificity was found in amPFC, dmPFC, and vmPFC (range: 95-
14 100%). Taken together, these results show a gradient of mentalizing-specificity from visual
15 cortex to mPFC.

16

17 *A spatiotemporal sequence of mentalizing activations from visual cortex to mPFC*

18 Next, we explored the timing of mentalizing-evoked activations across our ROIs. To this
19 end, we analyzed single-trial HFB metrics of onset, peak, and offset latencies from mentalizing-
20 active ROI sites (Fig. 3ef). Pairwise ROI contrasts (Fig. 3f) were performed for each latency
21 metric using LMEMs ($p_{\text{FDR}} < .05$; corrected for number of unique ROI pairs) that controlled for
22 heterogeneity across mentalizing type (self/other), behavioral response times (RT_{Task}), and ROI
23 coverage (see Methods).

24 Pairwise ROI contrasts ($p_{\text{FDR}} < .05$; Fig. 3f) revealed that visual cortex produced the
25 earliest activation onsets ($M=94$ ms, $SE=5$) of any ROI. Next, mid-latency onsets were produced

1 by TPJ ($M=316$ ms, $SE=9$), ATL ($M=355$ ms, $SE=12$), and PMC ($M=367$ ms, $SE=8$), with non-
2 significant differences between them. Later onsets were produced by amPFC ($M=486$ ms,
3 $SE=17$) and dmPFC ($M=493$ ms, $SE=14$), with non-significant differences between them. Finally,
4 the latest onsets were produced by vmPFC ($M=594$ ms, $SE=25$). However, three vmPFC sites on
5 the posterior orbital surface produced very early activations (<200 ms; Figs. 5a, 2k, S7) that met
6 outlier exclusion criteria for ROI analyses (see Methods). Peak latencies showed the same
7 pattern of cross-ROI differences as onsets, though differences across mPFC ROIs were non-
8 significant (Fig. 3f). Offset latencies exhibited the least differentiation across ROIs. The earliest
9 offsets were produced by visual cortex, TPJ, and ATL, with non-significant differences between
10 them. The latest offsets were produced by PMC, amPFC, dmPFC, and vmPFC, with non-
11 significant differences between them. Crucially, despite these robust cross-ROI latency
12 differences, all ROIs activated concurrently from 594-1908 ms, with even longer activation
13 overlap between specific ROI pairs (Fig. 3e).

14 In sum, mentalizing evoked largely concurrent activations across all ROIs, as might be
15 expected from neuroimaging literature. Nonetheless, fine-scale cross-ROI differences in onset,
16 peak, and offset latencies depict an overarching spatiotemporal sequence of activation from
17 visual cortex to mPFC. Spatial gradients in activation latency were also seen across individual
18 sites throughout cortex (Fig. 5a).

19

20 *Activation onset latency predicts mentalizing-specificity*

21 Thus far, we have revealed spatial gradients in the timing (Fig. 3ef) and functional
22 specificity (Fig. 3cd) of neuronal population responses to mentalizing. To examine the
23 correspondence between these gradients, we correlated mean onset latencies (black squares in
24 Fig. 3f) and mentalizing-specificity (percentage of mentalizing-specific sites; Fig. 3d) across all
25 seven ROIs. Despite this small sample, we found a near-perfect positive correlation between

1 onset latency and mentalizing-specificity ($r(5)=0.98, p=9.32e-5$), indicating that ROIs with later
2 activations had greater mentalizing-specificity.

3 These results portray a hierarchical neurocognitive pathway²⁸⁻³¹ that reflects both the
4 timing and specificity of mentalizing processing across ROIs. This pathway begins in visual
5 cortex, which featured the earliest activations and the least mentalizing-specificity.
6 Temporoparietal DMN regions (TPJ, ATL, and PMC) appear to be intermediate stages of the
7 pathway, as these regions featured mid-latency activations and intermediate mentalizing-
8 specificity. The highest stages of the pathway appear to be mPFC regions (amPFC, dmPFC, and
9 vmPFC), which featured the latest activations and overwhelming mentalizing-specificity.

10

11 *Self- and other-mentalizing share a common neuroanatomical basis*

12 To fractionate mentalizing's neurocognitive pathway in further detail, we explored the
13 anatomical interrelations between self- and other-mentalizing. Using trial-averaged results, we
14 first identified sites that produced significant ($p_{FDR}<.05$) HFB activations for each mentalizing
15 type (Fig. 4ab). Sites that activated for only one mentalizing type were considered 'self-only' or
16 'other-only' (irrespective of responses to the cognitive arithmetic task). Sites that activated for
17 both mentalizing types were further analyzed for self/other selectivity using single-trial metrics
18 of peak power (i.e., activation magnitude). Sites with significant ($p_{FDR}<.05$; corrected for number
19 of sites) peak power differences were considered 'self-greater' (self>other) or 'other-greater'
20 (other>self), while sites with nonsignificant differences were considered 'non-selective'
21 (self=other). Note that percentages reported in this section and Figure 4b only consider
22 mentalizing-active sites.

23 We found that mentalizing-active sites were overwhelmingly coactive for both
24 mentalizing types, regardless of self/other selectivity (non-selective + self-greater + other-
25 greater = 91% whole-brain overlap; Fig. 4b). Moreover, non-selective sites formed the largest

1 single category in all ROIs (range: 35-82%) and the whole-brain (67%). Among selective sites, we
2 compared amounts of self-selective (self-only + self-greater) versus other-selective (other-only
3 + other-greater) sites (McNemar χ^2 , Yates-corrected; $df=1$). This revealed that other-selective
4 sites significantly ($p<.05$) outnumbered self-selective sites in visual cortex (4% self/33% other;
5 $\chi^2=13.88$), PMC (4% self/42% other; $\chi^2=13.14$), and the whole-brain (11% self/22% other;
6 $\chi^2=23.14$). Non-significant self/other differences were found in TPJ (26% self/29% other; $\chi^2=0$),
7 ATL (4% self/21% other; $\chi^2=1.50$), amPFC (35% self/30% other; $\chi^2=0$), dmPFC (7% self/11%
8 other; $\chi^2=0$), and vmPFC (18% self/45% other; $\chi^2=0.57$). In sum, self- and other-mentalizing
9 recruited near-identical neuronal populations in a broadly non-selective manner, though
10 selective sites were predominantly other-selective in visual cortex and PMC.

11

12 *Other-mentalizing evokes slower and longer activations in later-stage ROIs*

13 Given the largely-overlapping neuroanatomy of self- and other-mentalizing, we wondered
14 whether self/other differences might be better characterized by fast spatiotemporal functional
15 dynamics. We therefore compared single-trial metrics of the timing (onset, peak, and offset
16 latencies) and magnitude (peak power) of HFB activations evoked by self- and other-mentalizing
17 in ROI sites (Fig. 4cd & Table 1). Each HFB metric was separately analyzed for Condition (other-
18 self) differences using LMEMs ($p<.05$).

19 We found that self- and other-mentalizing evoked a common spatiotemporal sequence
20 of activations across ROIs (Fig. 4c). Within this sequence, onset latencies showed non-
21 significant Condition differences. However, significantly later peaks were evoked by other-
22 versus self-mentalizing in PMC, amPFC, dmPFC, and vmPFC. Similarly, significantly later offsets
23 were evoked by other- versus self-mentalizing in all DMN ROIs except TPJ. Analysis of
24 activation duration (number of active timepoints) confirms that significantly longer activations
25 were evoked by other- versus self-mentalizing in PMC, amPFC, dmPFC, and vmPFC (see

1 Supplementary Report & Fig. S1). In contrast, self/other magnitude differences were only
2 significant in visual cortex and PMC, where greater peak power was evoked by other- versus
3 self-mentalizing (Fig. 4d). Critically, LMEMs indicated that these results were dissociable from
4 self/other differences in RT_{Task} , among other confounds (see Methods).

5 Lastly, we examined whether these self/other functional differences became stronger in
6 ROIs with later activations. We correlated Condition effect sizes for the onset, peak, and offset
7 HFB metrics (b -coefficients in Table 1) with mean onset, peak, and offset latencies (black
8 squares in Fig. 3f), respectively, across all seven ROIs. We found significant positive
9 correlations in all latency metrics (onset latency: $r(5)=.79, p=.036$; peak latency: $r(5)=.91, p=.004$,
10 offset latency: $r(5)=.88, p=.009$), indicating that self/other latency differences strengthened
11 across successive ROIs. Moreover, self/other differences in activation duration also
12 strengthened across successive ROIs (see Supplementary Report). In contrast, self/other
13 magnitude differences (peak power) were not correlated with the sequence of ROI recruitment
14 (mean peak latencies): $r(5)= -.32, p=.481$.

15 In sum, we found that self- and other-mentalizing activated near-identical sites in a
16 common spatiotemporal sequence (Fig. 4a-c). Within this sequence, other-mentalizing evoked
17 slower and longer activations than self-mentalizing in succeeding DMN ROIs (Figs. 4c & S1).
18 Critically, these self/other timing differences strengthened across successive ROIs. Taken
19 together, later-stage ROIs showed greater self/other differentiation, which was primarily
20 characterized by the timing, rather than magnitude, of activations (Table 1).

21

22 *Behavioral response times are best predicted by TPJ and dmPFC activity*

23 To explore the relationship between neuronal population activity and mentalizing task
24 performance, we compared behavioral response times (RT_{Task}) with single-trial HFB metrics
25 (e.g. onsets, peaks, offsets) from mentalizing-active ROI sites using LMEMs (Table 1). Onset

1 latencies in visual cortex, TPJ, PMC, amPFC, and dmPFC significantly ($p < .05$) positively
2 predicted RT_{Task} . Unsurprisingly, peak and offset latencies in all ROIs significantly positively
3 predicted RT_{Task} . In contrast, peak power (i.e., activation magnitude) in only TPJ and dmPFC
4 significantly predicted RT_{Task} (Fig. 4e). Lastly, we examined the correlations between RT_{Task}
5 effect size (b -coefficients in Table 1) and mean activation latencies (Fig. 3f) across ROIs using
6 the same correlation method as described above. This revealed that ROIs with later activations
7 had stronger RT_{Task} associations (peak latency: $r(5) = .96, p = .0008$; offset latency: $r(5) = .78,$
8 $p = .038$; peak power: $r(5) = .77, p = .041$), though the correlation for onset latency was non-
9 significant ($r(5) = .49, p = .269$). In sum, activations in later-stage ROIs had the strongest RT_{Task}
10 associations. Nevertheless, TPJ and dmPFC best predicted RT_{Task} , as they were the only ROIs
11 with significant RT_{Task} effects in all HFB metrics.

12

13 *Summary of task-evoked neuronal population activity*

14 To summarize the spatiotemporal dynamics of neuronal population responses to each
15 task condition, we identified which sites produced significant HFB responses (relative to pre-
16 stimulus baseline; $p_{FDR} < .05$) within specific time-windows (Fig. 6). From 0-250 ms, activations
17 were largely localized to visual cortex and showed little differentiation across task conditions.
18 From 250-500ms, activations spread beyond visual cortex to encompass temporoparietal and
19 lateral frontal cortex. During this time window, mentalizing and the cognitive task (arithmetic)
20 began to diverge. In temporoparietal DMN regions, self- and other-mentalizing primarily evoked
21 activations, while arithmetic evoked a mix of activations and deactivations. Self-mentalizing
22 also evoked some mPFC activations. From 500-750 ms, all task conditions evoked mPFC
23 responses. Self-mentalizing evoked stronger mPFC activations than other-mentalizing. In
24 contrast, arithmetic evoked mPFC deactivations, which continued for all successive time
25 windows. From 750-1000 ms, self- and other-mentalizing evoked similar responses. From 1000-

1 1500 ms, other-mentalizing evoked more sustained activations than self-mentalizing,
2 particularly in amPFC and dmPFC. From 1500-2000 ms, mentalizing-evoked activations
3 weakened, except in vmPFC, which sustained strong activations for other-mentalizing.
4

5 *Effects of task condition on behavioral response times*

6 Linear mixed-effects model (LMEM) analysis of behavioral response times (RT_{Task})
7 revealed that self-mentalizing elicited the fastest responses ($M=2559$ ms, $SE=56$), followed by
8 other-mentalizing ($M=2935$ ms, $SE=60$), and lastly the cognitive task ($M=3936$ ms, $SE=60$).
9 Compared to self-mentalizing, significantly later RT_{Task} was evoked by other-mentalizing ($b=381$
10 ms, $SE=107$, $p=3.71e-4$). Compared to other-mentalizing, significantly later RT_{Task} was evoked by
11 the cognitive task ($b=1041$ ms, $SE=201$, $p=2.48e-7$).
12

13 **Discussion**

14 Using electrocorticography (ECoG), we probed the neurocognitive substrates of
15 mentalizing with unprecedented spatiotemporal resolution. We found that mentalizing about the
16 self and others recruited near-identical neuronal populations (Fig. 4ab) in a common
17 spatiotemporal sequence (Figs. 4c & 6). Within our ROIs, activations began in visual cortex, then
18 spread to temporoparietal DMN regions (TPJ, ATL, and PMC), and lastly to mPFC regions
19 (amPFC, dmPFC, and vmPFC; Fig. 3ef). Critically, regions with later activations exhibited greater
20 functional specialization for mentalizing as measured by three metrics: functional specificity for
21 mentalizing versus arithmetic (Fig. 3d), self/other differentiation in activation latencies (Fig. 4c),
22 and prediction of behavioral response times (RT_{Task} ; Fig. 4e & Table 1). Taken together, these
23 results portray a common neurocognitive pathway²⁸⁻³¹ for self- and other-mentalizing, beginning
24 in visual cortex (low specialization), ascending through temporoparietal DMN areas
25 (intermediate specialization), then reaching its apex in mPFC regions (high specialization).

1 Our results are consistent with gradient-based models of brain function, which posit that
2 concrete sensorimotor processing in unimodal regions (e.g., visual cortex) gradually yields to
3 increasingly abstract and inferential processing in ‘high-level’ transmodal regions like mPFC^{32,33}.
4 We found that the strength of self/other differences in activation latencies increased along a
5 gradient from visual cortex to vmPFC. Specifically, other-mentalizing evoked slower (Fig. 4c)
6 and lengthier (Fig. S1) activations than self-mentalizing in successive DMN ROIs (Table 1).
7 Thus, perhaps because we know ourselves better than others, other-mentalizing may require
8 longer computation times at more abstract and inferential levels of processing. What might
9 these hierarchical neurocognitive dynamics imply about fMRI findings?

10 Hundreds of fMRI studies on mentalizing consistently suggest that: (1) TPJ and dmPFC
11 are most crucial among a network of mentalizing regions^{6,8,10,12,34–37}, and (2) dmPFC is selective
12 for thinking about others over oneself^{38–41}. However, when examined with ECoG, we found that
13 both pieces of received wisdom are not what they seem. Below, we discuss our results in
14 relation to (1) then (2). Afterwards, we discuss our findings at the systems level.

15 Unsurprisingly, we found that DMN regions such as TPJ and dmPFC contained higher
16 proportions of ‘mentalizing-specific’ sites relative to the whole-brain average (i.e., sites that
17 activated for mentalizing but not for the arithmetic cognitive task; Fig. 3cd). The spatial
18 distribution of such sites roughly resembles the ‘mentalizing network’ reported in countless
19 fMRI studies^{5,6,8–12}. However, our DMN ROIs were by no means functionally homogenous.
20 Relative to other ROIs, TPJ and dmPFC activity best predicted RT_{Task} (Fig. 4e & Table 1),
21 supporting the notion that both regions are most crucial for mentalizing
22 performance^{6,10,34,36,37,42,43}.

23 We also found numerous functional distinctions between TPJ and dmPFC, which is
24 surprising given their remarkably similar functional profiles in fMRI literature^{5,6,8–12,44}.
25 Specifically, we found that TPJ produced earlier activations (Fig. 3ef) that were notably coactive

1 for mentalizing and arithmetic (cognitive task; Fig. 3cd). Indeed, the onsets and offsets of TPJ
2 activations were the earliest of any DMN ROI (Fig. 3f). In contrast, dmPFC produced significantly
3 later activations (Fig. 3ef) that were overwhelmingly mentalizing-specific (Fig. 3cd), indicating
4 that dmPFC sits at a higher level of mentalizing's neurocognitive pathway than TPJ. Moreover,
5 aggregate ROI analyses revealed no significant self/other differences in TPJ (Table 1), while
6 dmPFC showed robust self/other timing differences (Fig. 4c), suggesting that dmPFC is more
7 sensitive to differences in mentalistic content. Furthermore, even when controlling for self/other
8 differences, dmPFC had stronger RT_{Task} associations than TPJ in all HFB metrics (Fig. 4e &
9 Table 1). Strikingly, unlike TPJ, the offsets of dmPFC activations closely preceded RT_{Task} (within
10 ~200 ms; Fig. S3), suggesting that dmPFC is more deeply involved in the final stages of
11 mentalistic decision-making. Taken together, while TPJ and dmPFC are both clearly crucial for
12 mentalizing performance, dmPFC appears more specialized for mentalizing itself.

13 Given the marked functional differentiation between TPJ and dmPFC, what specific
14 neurocognitive roles might they play in mentalizing? In social neuroscience, TPJ is often
15 considered to be a functionally-specific locus for explicit belief reasoning^{45,36,46,37}. Yet here, TPJ
16 appears less functionally specialized relative to dmPFC (Figs. 3c-f & 4c-e). To explain this
17 discrepancy, we suggest that TPJ provides crucial antecedents for explicit belief reasoning in
18 dmPFC. Given TPJ's central role in automatic evaluations of thematic semantics⁴⁷⁻⁵⁵, we
19 propose that TPJ automatically represents integrative psycho-semantic models of exemplar
20 contexts for a given inference. In simpler terms, TPJ may help us 'see' the psycho-semantic
21 gestalt of a given situation⁵⁶. Accordingly, tasks that 'show' concrete mentalistic content (e.g.,
22 social animations, reading the mind in the eyes) reliably recruit TPJ but not dmPFC, while tasks
23 that require mentalistic logical inferences (e.g., false belief, trait judgments) reliably recruit
24 dmPFC in addition to TPJ¹⁰⁻¹². Thus, when mentalistic content feels 'seeable' from perceptual
25 processing, TPJ could generate mentalistic inferences without explicit belief reasoning. Indeed,

1 work on implicit and spontaneous mentalizing consistently find that TPJ (but not dmPFC)
2 encodes an actor's beliefs without any explicit reasoning^{16,57-65}. Taken together, TPJ may
3 implicitly set the psycho-semantic stage for explicit belief reasoning that occurs later in dmPFC
4 when necessary (e.g., the present trait judgment task; Fig. 1a).

5 The dmPFC may be well-suited for explicit belief reasoning⁶⁶⁻⁶⁹. We found substantial
6 concurrent activation between dmPFC and all other ROIs (Fig. 3ef), suggesting that dmPFC
7 could work iteratively with lower-level regions to refine what is 'seen', thus providing dmPFC
8 with increasingly-useful inputs from which to draw better inferences^{70,71}. Moreover, studies on
9 strategic reasoning show that dmPFC can arbitrate between multiple mental models⁷² and
10 prospective choices⁷³ by simultaneously evaluating multiple possibilities⁷⁴ through 'fuzzy'
11 propositional reasoning^{71,75}. As such, dmPFC may arbitrate between multiple TPJ-generated
12 exemplar contexts to help extract the most relevant and enduring semantic features for a given
13 psychological inference. Taken together, dmPFC may integrate and refine representations
14 throughout mentalizing's neurocognitive pathway to strategically reason about minds.

15 As for mentalizing about the self or others, fMRI studies routinely suggest that amPFC is
16 'self-selective' while dmPFC is 'other-selective'³⁸⁻⁴¹. In both regions, what underlying neuronal
17 population dynamics could result in stronger hemodynamic responses for one mentalizing type
18 over another? The standard assumption would be that the magnitude (i.e., intensity) of neuronal
19 activations differs across mentalizing type. This might be seen in aggregate ROI activity, or
20 perhaps across individual ROI sites. We tested both possibilities by examining HFB activation
21 magnitudes (e.g., peak power) across amPFC and dmPFC. Unexpectedly, aggregate ROI
22 analysis of peak power revealed nonsignificant self/other differences in both ROIs (Fig. 4d &
23 Table 1). Similarly, across individual sites, we found that the vast majority of dmPFC sites
24 showed nonsignificant self/other differences in peak power (82% non-selective), while amPFC
25 contained near-equal amounts of self-selective sites (35%) and other-selective sites (30%; Fig.

1 4ab). Critically, self-selective sites and other-selective sites did not significantly outnumber each
2 other within both ROIs. However, amPFC did contain a notable proportion of sites that only
3 produced significant activations for self-mentalizing (25%; Fig. 4b), perhaps explaining amPFC's
4 'self-selective' hemodynamic responses. In sum, we found that dmPFC produced equivalent
5 activation magnitudes for both mentalizing types, which appears inconsistent with reports of
6 'other-selective' dmPFC responses in fMRI studies.

7 We instead found robust self/other differences in the *timing* of dmPFC activations.
8 Specifically, other-mentalizing evoked slower and lengthier HFB activations compared to self-
9 mentalizing (Fig. 4c & S1). In other words, dmPFC activity remained significantly above baseline
10 for longer during other-mentalizing (see Figs. 2e, 5g & 6). This suggests a different account of
11 why dmPFC produces stronger hemodynamic responses for mentalizing about others over
12 oneself. The typical story is that dmPFC is highly specialized for thinking about other people's
13 minds³⁸⁻⁴¹. However, we found that both mentalizing types recruited the same neuronal
14 populations (100% overlap) at equivalent intensities (82% non-selective) in dmPFC (Fig. 4abd).
15 Alternatively, dmPFC could be sensitive to the inherently greater difficulty of other-mentalizing,
16 which may necessitate additional processing cycles before completion; it should take longer to
17 reason about other people's minds. Self-mentalizing may be simplified by the rich compendium
18 of accessible knowledge we have about ourselves, thus resulting in brief but equally intense
19 processing. Given that standard fMRI analysis does not distinguish the intensity of response
20 from the duration of response, it appears the latter has been mistaken for the former.

21 At the systems level, we revealed complex and hierarchical processing dynamics across
22 mentalizing's cortical pathway. Mentalizing about the self and others evoked similar
23 spatiotemporal sequences of activation onsets, peaks, and offsets, revealing an overarching
24 processing sequence from visual cortex to vmPFC (Fig. 4c). The sequence of onsets may depict
25 an initial 'feedforward sweep' of coarser processing^{28,76} along this pathway. Indeed, onset

1 latencies were insensitive to self/other differences (Table 1). Onsets were followed by
2 considerable concurrent activations across all ROIs, which were sustained until the sequence of
3 activation offsets (Figs. 3ef, 4c, 2, 5b-e, 6 & S2). Given that concurrent activation is considered
4 an index of recurrent processing^{28,77,78}, distinct regions of mentalizing's cortical pathway may
5 largely work together within an overarching spatiotemporal sequence. For example, PMC's
6 lengthy activations (Fig. 3e & S1) bridged the gap between the earlier onsets of temporoparietal
7 DMN ROIs and the later offsets of mPFC ROIs (see white squares in Fig. 3f), supporting the idea
8 that PMC is the posterior DMN hub that helps coordinate processing between temporoparietal
9 and prefrontal DMN regions^{44,79}. Intriguingly, in aggregate analyses, self/other differences did
10 not reach significance until concurrent activation was achieved across all ROIs (e.g., peak and
11 offset metrics; Figs. 4cd & Table 1), perhaps signifying the importance of recurrent processes in
12 self/other differentiation. In sum, mentalizing may be supported by a brief initial 'feedforward
13 sweep' of coarser processing along the cortical pathway, followed by substantial recurrent
14 processing that may integrate and refine representations across cortical pathway regions.
15 These dynamics could obscure cross-regional functional distinctions in fMRI studies.

16 This study is not without confounds and limitations. Some of these limitations are
17 inherent to ECoG: the use of epileptic patients, inconsistent brain coverage across subjects, and
18 sampling bias for cortical gyri²⁶. Although these limitations were mitigated to the best of our
19 ability (see Methods), they cannot be completely ameliorated. Thus, our ECoG findings could be
20 corroborated by examining healthy subjects with recent advances in source-space EEG/MEG,
21 such as ultra-high density EEG⁸⁰, optically-pumped MEG⁸¹, and laminar source localization⁸².
22 Another important confound was the sparse right-hemisphere coverage of our cohort, which
23 may limit the interpretability of our ATL and TPJ results^{53,83}. However, the few right-hemisphere
24 sites in ATL and TPJ appear functionally similar to their left-hemisphere homologues (Figs. S4-
25 S7, Supplementary Tables 3-4). Another limitation was the use of cross-regional timing

1 differences to infer feedforward and recurrent processes. Future work could use effective
2 connectivity analyses to better reveal the directionality and causality of information flow during
3 mentalizing. An additional limitation was the short pre-stimulus baseline in our task (200 ms),
4 which sometimes contained residual activity from previous trials, thus likely resulting in
5 artifactual 'deactivations' in somatomotor sites (Fig. 6). Another confound was the greater word
6 count of other-mentalizing prompts (e.g., "My neighbor is...") versus self-mentalizing prompts
7 (e.g., "I am..."; Fig. 1a), which could conceivably result in longer or stronger activations for other-
8 mentalizing, although we consider this possibility unlikely outside of visual cortex (Fig. 4a-d).
9 Relatedly, another limitation may be the ease of our mentalizing task, which involves less
10 abstract mentalization than, for instance, a typical false belief task⁸⁴.

11 Distributed hierarchical processing is a central organizing principle of neurocognitive
12 systems^{78,28,33,31}. Characterizing such hierarchies has enabled incisive neuromechanistic
13 accounts of many psychological functions^{30,85}. Here we provide a comprehensive
14 electrophysiological exploration of the human social brain, revealing that mentalizing is
15 characterized by complex and hierarchical neurocognitive dynamics at millisecond, millimeter,
16 and cross-regional scales. While many questions remain, our findings contribute to a solid
17 foundation upon which more conclusive neurocognitive accounts of mentalizing can be built.

18

19 **Methods**

20 All computational procedures and analyses herein were implemented in MATLAB unless
21 otherwise specified⁸⁶.

22

23 *Subjects*

24 We employed a cohort of sixteen patients that underwent neurosurgical treatment for
25 drug-resistant epilepsy (Supplementary Table 2). Each patient provided written informed

1 consent to participate in the study, which was approved by the Stanford Institutional Review
2 Board. As part of their presurgical evaluation, patients were implanted with ECoG at Stanford
3 University Medical Center. The anatomical placement of electrode sites was determined
4 according to each patient's clinical needs. Patients were included in this study's subject cohort
5 if they had electrode coverage in key DMN regions: mPFC, PMC, TPJ, and ATL. Each patient was
6 monitored in the hospital for six to ten days prior to surgery, during which the study was
7 conducted.

8

9 *Experimental Task*

10 ECoG data was recorded while subjects performed an event-related experimental task
11 with six conditions (trial types; Fig.1a). Five of these conditions required true/false responses to
12 written prompts, while one condition consisted of cued rest. Two conditions featured
13 mentalizing prompts, either about oneself (e.g., "I am honest") or others (e.g., "My neighbor is
14 honest"). Subjects were instructed to select a single neighbor (current or past) as the target for
15 other-mentalizing. Cognitive task trials consisted of basic arithmetic (e.g., "9 + 86 = 95"). Two
16 conditions featured memory-related prompts: episodic (e.g., "I ate candy yesterday") and self-
17 semantic (e.g., "I eat a lot of candy"). The cued rest condition required no response and
18 displayed a fixation crosshair for 5-10 seconds. The memory and rest conditions were not
19 relevant to the current analyses have been reported elsewhere⁸⁷⁻⁸⁹. Stimuli were presented in a
20 random order and were self-paced, advancing to the next trial after the subjects' response, or up
21 to 15 seconds if no response. The inter-trial interval (ITI) occurred -200-0 ms before each trial.
22 The experiment was broken into two separate runs (mean run duration = 12.50 ± 1.64 minutes).
23 Subjects were allowed a short break in between the experimental runs. On average, each run
24 featured 25 trials of each sentence condition, 40 cognitive trials, and 36 rest trials. Each non-
25 rest trial contained unique prompts; prompts were not repeated within subject. Responses were

1 made via a handheld keypad using either the '1' (true) or '2' (false) key. Subjects were instructed
2 to perform the task as accurately and as quickly as possible. All stimuli were presented in white
3 font on a black background using Psychophysics Toolbox 3
4 (<http://psychtoolbox.org/HomePage>). Behavioral response times (RT_{task}) across task conditions
5 were analyzed through a linear mixed-effects model (LMEM) with Condition nested within
6 Subject to account for subject-specific variance.

7

8 *Electrocorticography data acquisition*

9 ECoG recordings were obtained via 2125 subdural electrodes (Fig. 1b). Electrodes
10 (platinum plates with diameter of 1.2-2.3 mm) were implanted subdurally onto the cortical
11 surface in grids or strips with center-to-center interelectrode spacing of 4-10 mm (Adtech
12 Medical Instruments). Electrodes were connected to a multichannel recording system (Nihon
13 Kohden; Tucker Davis Technologies) with sampling rate of 1,000 Hz or above. Anatomical data
14 was acquired using a GE 3-Tesla SIGNA Magnetic Resonance Imaging (MRI) scanner at
15 Stanford University. A T1-weighted anterior-posterior commissure-aligned pulse sequence was
16 used. T1 data was resampled to 1 mm isotropic voxels, then segmented to distinguish gray and
17 white matter using FreeSurfer⁹⁰. To facilitate electrode localization, postimplant computerized
18 tomography (CT) scans were coregistered to the preoperative MRI anatomical brain volume⁹¹.
19 For each patient, electrodes sites were localized in BiImage Suite⁹² and displayed on the
20 patients' own reconstructed 3D cortical surface using the iELVis toolbox⁹³. Electrode positions
21 were corrected for postimplantation brain shift, allowing for the accurate anatomical
22 localization of electrodes sites⁹⁴.

23

24 *Defining regions of interest (ROIs) and brain networks*

1 Each subject's native-space cortical surface reconstruction (e.g., Fig. S2) was used to
2 classify electrode sites into *a priori* ROIs that are strongly implicated in mentalizing, with visual
3 cortex included as a control ROI (Fig. 1c). MNI-based parcellation was avoided due to known
4 transformation inconsistencies in ECoG⁹³. ROIs were defined through FreeSurfer cortical
5 parcellation combined with visual inspection of anatomical landmarks. The ROI for 'visual
6 cortex' consisted of occipital cortex, lingual gyrus, posterior fusiform gyrus, and posterior
7 inferotemporal cortex. The 'ATL' ROI consisted of a bilateral anterior subregion of temporal
8 cortex with precentral sulcus as the posterior bound, comprising the temporal poles and
9 adjacent sections of entorhinal cortex and superior, middle, and inferior temporal sulci/gyri. The
10 'TPJ' ROI was a bilateral posterior subregion of inferior parietal lobule with lateral sulcus as the
11 anterior bound, comprising angular gyrus and adjacent sections of supramarginal gyrus and
12 superior temporal sulcus/gyrus. The 'PMC' ROI consisted of precuneus, posterior cingulate, and
13 retrosplenial cortex. The 'amPFC' ROI was an mPFC subregion bounded between the ventral and
14 dorsal reaches of corpus callosum. The 'dmPFC' ROI was a mPFC subregion ventrally bounded
15 by the amPFC ROI and posteriorly bounded by the callosal rostrum. The 'vmPFC' ROI was an
16 mPFC subregion dorsally bounded by the amPFC ROI and posteriorly bounded by the callosal
17 rostrum, including the medial orbitofrontal surface.

18

19 *ECoG preprocessing*

20 Preprocessing was performed on a single-subject/single-electrode basis using custom
21 routines (https://github.com/LBCN-Stanford/Preprocessing_pipeline). First, data were notch
22 filtered for power-line noise (57-63 Hz) and harmonics (117-123 Hz, 177-183 Hz). Electrodes
23 were discarded from further analyses if they were marked as pathological or 'noisy' by
24 postclinical evaluation. The data was then rereferenced by subtracting the mean signal of the
25 remaining electrodes from each electrode's signal. The rereferenced data underwent time-

1 frequency decomposition into 4-200 Hz spectra in 1-10 Hz bands using 5-cycle Morlet wavelet
2 transforms. The power of the signal in each frequency band was z-transformed across time;
3 this partially corrects for the 1/frequency decay of neurophysiological signals and improves
4 interpretability. Data was then epoched into trials that were time-locked to stimulus onsets,
5 ranging from 200 ms pre-stimulus to 5000 ms post-stimulus (e.g. -200-5000 ms epochs). For
6 each trial and frequency increment, baseline correction was performed by subtracting the mean
7 power across the pre-stimulus baseline period (-200–0 ms) from all timepoints within a trial. To
8 reconstruct the high-frequency broadband (HFB) signal, the primary signal of interest, z-
9 transformed power of frequency bands within 70-180 Hz were averaged to produce a single
10 HFB timecourse per electrode. Trials were rejected from further analyses if they featured
11 epileptic high-frequency oscillations. Lastly, HFB signal from each electrode was low-pass
12 filtered with a gaussian window (width=50 ms) for further analysis.

13

14 *Within-site analyses*

15 Within-site analyses (Fig. 1d-f) were performed to provide the bases for the primary
16 multi-site analyses. Data after a behavioral response (RT_{task}) were discarded. All multiple
17 comparisons corrections herein involved maintaining the False Discovery Rate (FDR) under
18 0.05, with p -values adjusted accordingly (p_{FDR}), using the Benjamini-Yekutieli procedure for data
19 with any dependence structure⁹⁵. Correlations between functional specialization metrics and
20 mean HFB latency metrics were performed using Pearson correlations.

21

22 *Trial-averaged analysis*

23 To identify sites with statistically-significant HFB responses ($p_{\text{FDR}} < .05$; corrected for
24 number of timepoints and sites within subject), linear mixed-effects models (LMEMs) were used
25 to analyze HFB power (Fig. 1d) during each task condition. The intercept (null distribution)

1 consisted of timepoints within the pre-stimulus baseline (-200–0 ms). Each peri-stimulus
2 timepoint (0–RT_{task} ms) was represented as a separate dummy variable. The intercept was
3 nested within trial to account for trial-specific variance. Restricted maximum likelihood
4 estimation using full-Cholesky parametrization was used to account for autocorrelation and
5 unequal variance between model terms⁹⁶. This LMEM specification estimates the mean HFB
6 response (β) for each timepoint and task condition (Fig. 1f). To dampen spikes and other noise,
7 timepoints were not considered significant unless $p_{FDR} < .05$ was maintained for 50 ms
8 consecutively. For each task condition, sites were considered ‘active’ or ‘deactive’ if evoked HFB
9 power was significantly higher or lower than pre-stimulus baseline; if sites produced both, the
10 polarity of the greatest deflection was used. Sites with nonsignificant differences from baseline
11 were considered ‘nonresponsive’.

12

13 Single-trial analysis

14 Single-trial analysis was performed to provide four key metrics of the HFB response:
15 onset, peak, and offset latencies, along with peak power (Fig. 1e). For each trial, timepoints
16 between stimulus onset and RT_{task} were run through a sliding window test (width: ± 10 ms) to
17 reveal timepoints with significant HFB responses. Observations (z-scored HFB power; see Fig.
18 1d) in each sliding window were tested against observations from the pre-stimulus baseline via
19 two-sample Welch’s *t*-tests to account for unequal variances and sample sizes. This analysis
20 identified timepoints within individual trials that featured significant stimulus-evoked responses
21 (brown areas; $p_{FDR} < .05$, corrected for number of timepoints, trials, and sites) relative to the pre-
22 stimulus baseline (ITI; -200-0ms) preceding each trial. Onset latency (green squares) is the
23 earliest timepoint with a significant response. Peak latency and peak power (white squares) are
24 the timepoint and magnitude, respectively, of the greatest significant response. Offset latency is
25 the latest timepoint with a significant response (red squares). To dampen spikes and other

1 noise, timepoints were not considered significant unless $p_{FDR} < .05$ was maintained for 50 ms
2 consecutively. Outlier observations were discarded if greater than three median absolute
3 deviations (MAD) from site or six MAD from ROI (within task condition). Sites were excluded
4 from ROI analyses if over 50% of observations exceeded outlier thresholds. Of all ROI sites, only
5 three sites in vmPFC were excluded.

6

7 *Multi-site analyses*

8 Multi-site analyses used results from within-site analyses as response measures.

9

10 *Functional specificity and selectivity*

11 Functional specificity was categorized using results from trial-averaged analysis within
12 sites. Sites were considered 'mentalizing-specific' if they produced significant ($p_{FDR} < .05$;
13 corrected for number of sites) activations for mentalizing (mentalizing-active) but not the
14 cognitive task (cognitive-nonresponsive or cognitive-deactive). Sites with significant
15 coactivations for mentalizing and the cognitive task were considered 'non-specific'. All
16 mentalizing-active sites were further analyzed for 'selective' activations to self- or other-
17 mentalizing (not considering cognitive task). Sites that activated for only one mentalizing type
18 were considered 'self-only' or 'other-only' (not considering cognitive task). Sites that activated
19 for both mentalizing types were analyzed for self/other selectivity by comparing single-trial
20 metrics of HFB peak power via Welch's t-tests ($p_{FDR} < .05$; corrected for number of sites). This
21 resulted in three additional categories: 'self-greater' (self > other), 'other-greater' (other > self),
22 and 'non-selective' (other = self).

23

24 *ROI analyses of onsets, peaks, and offsets*

1 To reveal the spatiotemporal dynamics of neuronal activation evoked by self- and other-
2 mentalizing, four key metrics of HFB activity were used as dependent variables: onset, offset,
3 peak latency, and peak power (derived from single-trial analyses). Each metric was analyzed
4 separately, first using full-factorial LMEMs, which determined appropriate follow-up tests that
5 used reduced LMEMs. This approach minimizes Type I, Type II, and Type III errors⁹⁷. The full-
6 factorial LMEMs included the main effects of ROI, Condition, and RT_{Task} and all possible
7 interactions. All effects were nested within Subject, and Condition was also nested within Site.
8 Full-factorial LMEMs underwent omnibus tests that used Satterthwaite approximation for
9 degrees of freedom to account for unequal variances and sample sizes⁹⁷. Follow-up tests were
10 performed for categorical factors that produced significant ($p < .05$) omnibus results. Within-ROI
11 follow-up tests were performed using reduced LMEMs with Site and Subject as nesting terms.
12 Pairwise ROI contrasts were performed using reduced LMEMs with Site and Subject as nesting
13 terms. Pairwise ROI tests only included subjects with sites in both ROIs and were FDR-corrected
14 for number of ROI pairs. These LMEM specifications were designed to distinguish experimental
15 effects from nuisance variance (e.g., heterogenous ROI coverage and RT_{task} across subjects and
16 task conditions).

17

18 Grand-average ROI timecourses

19 To reveal aggregate timecourses of HFB responses within ROIs (Fig. 5b-h), we examined
20 mean HFB timecourses (trial-averaged β -coefficients) from ROI sites that had significant
21 ($p_{\text{FDR}} < .05$; corrected for number of site and timepoints) responses for a given condition. The
22 intercept (null distribution) consisted of timepoints within the pre-stimulus baseline (-200–0
23 ms). Peri-stimulus timepoints (0–3000 ms) were represented as separate dummy variables.
24 The intercept was nested within Site and Subject to account for site- and subject-specific
25 variance.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Whole-brain HFB responses within time-windows

To provide a broad overview of the neuronal spatiotemporal dynamics evoked by each task condition, we performed whole-brain analysis of significant HFB responses within specific time-windows (Fig. 6). To this end, we used results from within-site trial-averaged analysis (see Fig. 1f & 2) to calculate mean statistics within the specified time windows ($p_{FDR} < .05$; corrected for number of sites and time-windows).

Acknowledgements

We are thankful to the Laboratory of Behavioral and Cognitive Neuroscience for generously providing access to this data and for providing technical support. We also thank members of the Social Cognitive Neuroscience Lab for their continued support. We are also thankful to Carolyn Parkinson for her continued support. This work was supported by National Institute of Neurological Disorders and Stroke Grant R01NS078396, National Institute of Mental Health Grant 1R01MH109954-01, NSF Grant BCS1358907 (all to J.P.), and National Science Foundation Graduate Research Fellowship DGE-1650604 (all to K.M.T.).

Author contributions

J.P. and A.L.D. contributed to experimental design and data acquisition. J.P., A.L.D, P.P., and K.M.T. developed analysis tools. K.M.T. performed data analyses. All authors contributed to writing of the manuscript.

Ethics declarations

All authors declare no conflicts of interest.

1 **References**

- 2 1. Lieberman, M. D. *Social: Why Our Brains Are Wired to Connect*. (Crown, 2013).
- 3 2. Leonhardt, S. D., Menzel, F., Nehring, V. & Schmitt, T. Ecology and Evolution of
4 Communication in Social Insects. *Cell* **164**, 1277–1287 (2016).
- 5 3. Fox, K. C. R., Muthukrishna, M. & Shultz, S. The social and cultural roots of whale and
6 dolphin brains. *Nature Ecology & Evolution* **1**, 1699–1705 (2017).
- 7 4. Schaafsma, S. M., Pfaff, D. W., Spunt, R. P. & Adolphs, R. Deconstructing and
8 reconstructing theory of mind. *Trends in Cognitive Sciences* **19**, 65–72 (2015).
- 9 5. Frith, C. D. & Frith, U. The Neural Basis of Mentalizing. *Neuron* **50**, 531–534 (2006).
- 10 6. Lieberman, M. D. Social Cognitive Neuroscience. in *Handbook of Social Psychology*
11 (American Cancer Society, 2010). doi:10.1002/9780470561119.socpsy001005.
- 12 7. Friston, K. J. Functional and Effective Connectivity: A Review. *Brain Connectivity* **1**, 13–36
13 (2011).
- 14 8. Van Overwalle, F. & Baetens, K. Understanding others' actions and goals by mirror and
15 mentalizing systems: A meta-analysis. *NeuroImage* **48**, 564–584 (2009).
- 16 9. Mars, R. B. *et al.* On the relationship between the “default mode network” and the “social
17 brain”. *Front. Hum. Neurosci* **6**, 189 (2012).
- 18 10. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: A
19 meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*
20 **42**, 9–34 (2014).
- 21 11. Amft, M. *et al.* Definition and characterization of an extended social-affective default
22 network. *Brain Struct Funct* **220**, 1031–1049 (2015).
- 23 12. Molenberghs, P., Johnson, H., Henry, J. D. & Mattingley, J. B. Understanding the minds of
24 others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews* **65**, 276–291
25 (2016).

- 1 13. Fox, K. C. R., Foster, B. L., Kucyi, A., Daitch, A. L. & Parvizi, J. Intracranial Electrophysiology
2 of the Human Default Network. *Trends in Cognitive Sciences* **22**, 307–324 (2018).
- 3 14. Schoffelen, J.-M. & Gross, J. Source connectivity analysis with MEG and EEG. *Human Brain*
4 *Mapping* **30**, 1857–1865 (2009).
- 5 15. Vistoli, D., Brunet-Gouet, E., Baup-Bobin, E., Hardy-Bayle, M.-C. & Passerieux, C. Anatomical
6 and temporal architecture of theory of mind: A MEG insight into the early stages.
7 *NeuroImage* **54**, 1406–1414 (2011).
- 8 16. Van Duynslaeger, M., Van Overwalle, F. & Verstraeten, E. Electrophysiological time course
9 and brain areas of spontaneous and intentional trait inferences. *Soc Cogn Affect Neurosci*
10 **2**, 174–188 (2007).
- 11 17. Mossad, S. I. *et al.* Thinking about the thoughts of others; temporal and spatial neural
12 activation during false belief reasoning. *NeuroImage* **134**, 320–327 (2016).
- 13 18. Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C. & Kessler, K. Rhythm makes
14 the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in
15 embodied perspective taking. *Cortex* **75**, 68–81 (2016).
- 16 19. Yuk, V. *et al.* Do you know what I'm thinking? Temporal and spatial brain activity during a
17 theory-of-mind task in children with autism. *Developmental Cognitive Neuroscience* **34**,
18 139–147 (2018).
- 19 20. Zhang, L., Gan, J. Q., Zheng, W. & Wang, H. Spatiotemporal Phase Synchronization in
20 Adaptive Reconfiguration from Action Observation Network to Mentalizing Network for
21 Understanding Other's Action Intention. *Brain Topography* **31**, 447–467 (2018).
- 22 21. Spunt, R. P. & Lieberman, M. D. Dissociating Modality-Specific and Supramodal Neural
23 Systems for Action Understanding. *Journal of Neuroscience* **32**, 3575–3583 (2012).

- 1 22. Barrett, L. F. & Satpute, A. B. Large-scale brain networks in affective and social
2 neuroscience: towards an integrative functional architecture of the brain. *Current Opinion in*
3 *Neurobiology* **23**, 361–372 (2013).
- 4 23. Catmur, C. Understanding intentions from actions: Direct perception, inference, and the
5 roles of mirror and mentalizing systems. *Consciousness and Cognition* **36**, 426–433
6 (2015).
- 7 24. Tan, K. M., Burklund, L. J., Craske, M. G. & Lieberman, M. D. Posttraumatic stress disorder
8 and the social brain: Affect-related disruption of the default and mirror networks.
9 *Depression and Anxiety* **36**, 1058–1071 (2019).
- 10 25. Akalin Acar, Z. & Makeig, S. Effects of Forward Model Errors on EEG Source Localization.
11 *Brain Topogr* **26**, 378–396 (2013).
- 12 26. Parvizi, J. & Kastner, S. Promises and limitations of human intracranial
13 electroencephalography. *Nature Neuroscience* **21**, 474 (2018).
- 14 27. Dubey, A. & Ray, S. Cortical Electrocorticogram (ECoG) Is a Local Signal. *J. Neurosci.* **39**,
15 4299–4311 (2019).
- 16 28. Felleman, D. J. & Van Essen, D. C. Distributed Hierarchical Processing in the Primate
17 Cerebral Cortex. *Cerebral Cortex* **1**, 1–47 (1991).
- 18 29. Mormann, F. *et al.* Latency and Selectivity of Single Neurons Indicate Hierarchical
19 Processing in the Human Medial Temporal Lobe. *J. Neurosci.* **28**, 8865–8872 (2008).
- 20 30. Diaconescu, A. O. *et al.* A computational hierarchy in human cortex. *arXiv:1709.02323 [q-*
21 *bio]* (2017).
- 22 31. Hilgetag, C. C. & Goulas, A. 'Hierarchy' in the organization of brain networks. *Philosophical*
23 *Transactions of the Royal Society B: Biological Sciences* **375**, 20190319 (2020).
- 24 32. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of
25 macroscale cortical organization. *PNAS* **113**, 12574–12579 (2016).

- 1 33. Huntenburg, J. M., Bazin, P.-L. & Margulies, D. S. Large-Scale Gradients in Human Cortical
2 Organization. *Trends in Cognitive Sciences* **22**, 21–31 (2018).
- 3 34. Apperly, I. A., Samson, D., Chiavarino, C. & Humphreys, G. W. Frontal and Temporo-Parietal
4 Lobe Contributions to Theory of Mind: Neuropsychological Evidence from a False-Belief
5 Task with Reduced Language and Executive Demands. *Journal of Cognitive Neuroscience*
6 **16**, 1773–1784 (2004).
- 7 35. Samson, D., Apperly, I. A., Chiavarino, C. & Humphreys, G. W. Left temporoparietal junction
8 is necessary for representing someone else’s belief. *Nature Neuroscience* **7**, 499–500
9 (2004).
- 10 36. Saxe, R. & Powell, L. J. It’s the Thought That Counts: Specific Brain Regions for One
11 Component of Theory of Mind. *Psychol Sci* **17**, 692–699 (2006).
- 12 37. Saxe, R. The right temporo-parietal junction: a specific brain region for thinking about
13 thoughts. *Handbook of theory of mind* 1–35 (2010).
- 14 38. Denny, B. T., Kober, H., Wager, T. D. & Ochsner, K. N. A Meta-Analysis of Functional
15 Neuroimaging Studies of Self and Other Judgments Reveals a Spatial Gradient for
16 Mentalizing in Medial Prefrontal Cortex. *J Cogn Neurosci* **24**, 1742–1752 (2012).
- 17 39. Bzdok, D. *et al.* Segregation of the human medial prefrontal cortex in social cognition.
18 *Front. Hum. Neurosci.* **7**, (2013).
- 19 40. de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. Large-Scale Meta-
20 Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *The*
21 *Journal of Neuroscience* **36**, 6553–6562 (2016).
- 22 41. Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M. & Tan, K. M. Social, Self, (Situational),
23 and Affective Processes in Medial Prefrontal Cortex (MPFC): Causal, Multivariate, and
24 Reverse Inference Evidence. *Neuroscience & Biobehavioral Reviews* (2019)
25 doi:10.1016/j.neubiorev.2018.12.021.

- 1 42. Cook, J. L. Task-relevance dependent gradients in medial prefrontal and temporoparietal
2 cortices suggest solutions to paradoxes concerning self/other control. *Neuroscience &*
3 *Biobehavioral Reviews* **42**, 298–302 (2014).
- 4 43. Spunt, R. P., Meyer, M. L. & Lieberman, M. D. The Default Mode of Human Brain Function
5 Primes the Intentional Stance. *Journal of Cognitive Neuroscience* **27**, 1116–1124 (2015).
- 6 44. Andrews-Hanna, J. R., Smallwood, J. & Spreng, R. N. The default network and self-
7 generated thought: component processes, dynamic control, and clinical relevance. *Ann N Y*
8 *Acad Sci* **1316**, 29–52 (2014).
- 9 45. Saxe, R. & Kanwisher, N. People thinking about thinking people: The role of the temporo-
10 parietal junction in “theory of mind”. *NeuroImage* **19**, 1835–1842 (2003).
- 11 46. Saxe, R. & Wexler, A. Making sense of another mind: The role of the right temporo-parietal
12 junction. *Neuropsychologia* **43**, 1391–1399 (2005).
- 13 47. Sass, K., Krach, S., Sachs, O. & Kircher, T. Lion – tiger – stripes: Neural correlates of
14 indirect semantic priming across processing modalities. *NeuroImage* **45**, 224–236 (2009).
- 15 48. Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends in Cognitive*
16 *Sciences* **15**, 527–536 (2011).
- 17 49. Geng, J. J. & Vossel, S. Re-evaluating the role of TPJ in attentional control: Contextual
18 updating? *Neurosci Biobehav Rev* **37**, 2608–2620 (2013).
- 19 50. Davey, J. *et al.* Automatic and Controlled Semantic Retrieval: TMS Reveals Distinct
20 Contributions of Posterior Middle Temporal Gyrus and Angular Gyrus. *J. Neurosci.* **35**,
21 15230–15239 (2015).
- 22 51. Lewis, G. A., Poeppel, D. & Murphy, G. L. The neural bases of taxonomic and thematic
23 conceptual relations: An MEG study. *Neuropsychologia* **68**, 176–189 (2015).

- 1 52. Lambon Ralph, M. A., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and
2 computational bases of semantic cognition. *Nature Reviews Neuroscience* **18**, 42–55
3 (2017).
- 4 53. Murphy, C. *et al.* Fractionating the anterior temporal lobe: MVPA reveals differential
5 responses to input and conceptual modality. *NeuroImage* **147**, 19–31 (2017).
- 6 54. Teige, C. *et al.* Dynamic semantic cognition: Characterising coherent and controlled
7 conceptual retrieval through time using magnetoencephalography and chronometric
8 transcranial magnetic stimulation. *Cortex* **103**, 329–349 (2018).
- 9 55. Patel, G. H., Sestieri, C. & Corbetta, M. The evolution of the temporoparietal junction and
10 posterior superior temporal sulcus. *Cortex* **118**, 38–50 (2019).
- 11 56. Lieberman, M. D. Gestalt cortex: Seeing minds, matter, and meaning. (Under Review).
- 12 57. Ma, N., Vandekerckhove, M., Overwalle, F. V., Seurinck, R. & Fias, W. Spontaneous and
13 intentional trait inferences recruit a common mentalizing network to a different degree:
14 Spontaneous inferences activate only its core areas. *Social Neuroscience* **6**, 123–138
15 (2011).
- 16 58. Kestemont, J., Vandekerckhove, M., Ma, N., Van Hoeck, N. & Van Overwalle, F. Situation
17 and person attributions under spontaneous and intentional instructions: an fMRI study.
18 *Soc Cogn Affect Neurosci* **8**, 481–493 (2013).
- 19 59. Kovács, Á. M., Kühn, S., Gergely, G., Csibra, G. & Brass, M. Are All Beliefs Equal? Implicit
20 Belief Attributions Recruiting Core Brain Regions of Theory of Mind. *PLOS ONE* **9**, e106558
21 (2014).
- 22 60. Hyde, D. C., Betancourt, M. A. & Simon, C. E. Human temporal-parietal junction
23 spontaneously tracks others' beliefs: A functional near-infrared spectroscopy study.
24 *Human Brain Mapping* **36**, 4831–4846 (2015).

- 1 61. Schuwerk, T., Vuori, M. & Sodian, B. Implicit and explicit Theory of Mind reasoning in
2 autism spectrum disorders: The impact of experience. *Autism* **19**, 459–468 (2015).
- 3 62. Naughtin, C. K. *et al.* Do implicit and explicit belief processing share neural substrates?
4 *Human Brain Mapping* **38**, 4760–4772 (2017).
- 5 63. Schneider, D., Slaughter, V. P. & Dux, P. E. Current evidence for automatic Theory of Mind
6 processing in adults. *Cognition* **162**, 27–31 (2017).
- 7 64. Bardi, L., Six, P. & Brass, M. Repetitive TMS of the temporo-parietal junction disrupts
8 participant's expectations in a spontaneous Theory of Mind task. *Soc Cogn Affect Neurosci*
9 **12**, 1775–1782 (2017).
- 10 65. Boccadoro, S. *et al.* Defining the neural correlates of spontaneous theory of mind (ToM):
11 An fMRI multi-study investigation. *NeuroImage* **203**, 116193 (2019).
- 12 66. Mitchell, J. P. Mentalizing and Marr: An information processing approach to the study of
13 social cognition. *Brain Research* **1079**, 66–75 (2006).
- 14 67. Krol, M. & Krol, M. Is Recursive Belief Inference the Engine of Mentalizing? *J. Neurosci.* **30**,
15 15711–15712 (2010).
- 16 68. Tamir, D. I., Thornton, M. A., Contreras, J. M. & Mitchell, J. P. Neural evidence that three
17 dimensions organize mental state representation: Rationality, social impact, and valence.
18 *PNAS* **113**, 194–199 (2016).
- 19 69. Tamir, D. I. & Thornton, M. A. Modeling the Predictive Social Mind. *Trends in Cognitive*
20 *Sciences* **22**, 201–212 (2018).
- 21 70. Cunningham, W. A. & Zelazo, P. D. Attitudes and evaluations: a social cognitive
22 neuroscience perspective. *Trends in Cognitive Sciences* **11**, 97–104 (2007).
- 23 71. Mitchell, C. J., Houwer, J. D. & Lovibond, P. F. The propositional nature of human
24 associative learning. *Behavioral and Brain Sciences* **32**, 183–198 (2009).

- 1 72. Lee, S. W., Shimojo, S. & O’Doherty, J. P. Neural Computations Underlying Arbitration
2 between Model-Based and Model-free Learning. *Neuron* **81**, 687–699 (2014).
- 3 73. Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D. & Daw, N. D. Model-based choices
4 involve prospective neural activity. *Nature Neuroscience* **18**, 767–772 (2015).
- 5 74. Barron, H. C., Dolan, R. J. & Behrens, T. E. J. Online evaluation of novel choices by
6 simultaneous representation of multiple memories. *Nature Neuroscience* **16**, 1492–1498
7 (2013).
- 8 75. Mitchell, J. P. Social psychology as a natural kind. *Trends in Cognitive Sciences* **13**, 246–
9 251 (2009).
- 10 76. Alilović, J., Timmermans, B., Reteig, L. C., van Gaal, S. & Slagter, H. A. No Evidence that
11 Predictions and Attention Modulate the First Feedforward Sweep of Cortical Information
12 Processing. *Cereb Cortex* **29**, 2261–2278 (2019).
- 13 77. Lamme, V. A. & Roelfsema, P. R. The distinct modes of vision offered by feedforward and
14 recurrent processing. *Trends in neurosciences* **23**, 571–579 (2000).
- 15 78. Hebb, D. O. *The organization of behavior; a neuropsychological theory*. (Wiley, 1949).
- 16 79. Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease.
17 *Brain* **137**, 12–32 (2014).
- 18 80. Robinson, A. K. *et al.* Very high density EEG elucidates spatiotemporal aspects of early
19 visual processing. *Scientific Reports* **7**, (2017).
- 20 81. Boto, E. *et al.* On the Potential of a New Generation of Magnetometers for MEG: A
21 Beamformer Simulation Study. *PLOS ONE* **11**, e0157655 (2016).
- 22 82. Bonaiuto, J. J. *et al.* Non-invasive laminar inference with MEG: Comparison of methods and
23 source inversion algorithms. *NeuroImage* **167**, 372–383 (2018).

- 1 83. Igelstrom, K. M., Webb, T. W., Kelly, Y. T. & Graziano, M. S. A. Topographical Organization
2 of Attentional, Social, and Memory Processes in the Human Temporoparietal Cortex.
3 *eNeuro* **3**, (2016).
- 4 84. Dodell-Feder, D., Koster-Hale, J., Bedny, M. & Saxe, R. fMRI item analysis in a theory of mind
5 task. *NeuroImage* **55**, 705–712 (2011).
- 6 85. Demirtaş, M. *et al.* Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale
7 Neural Dynamics. *Neuron* **0**, (2019).
- 8 86. Mathworks. *MATLAB 2019b*. (2019).
- 9 87. Foster, B. L., Rangarajan, V., Shirer, W. R. & Parvizi, J. Intrinsic and Task-Dependent
10 Coupling of Neuronal Population Activity in Human Parietal Cortex. *Neuron* **86**, 578–590
11 (2015).
- 12 88. Daitch, A. L. & Parvizi, J. Spatial and temporal heterogeneity of neural responses in human
13 posteromedial cortex. *PNAS* **115**, 4785–4790 (2018).
- 14 89. Raccach, O., Daitch, A. L., Kucyi, A. & Parvizi, J. Direct Cortical Recordings Suggest Temporal
15 Order of Task-Evoked Responses in Human Dorsal Attention and Default Networks. *J.*
16 *Neurosci.* **38**, 10305–10313 (2018).
- 17 90. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
- 18 91. Hermes, D., Miller, K. J., Noordmans, H. J., Vansteensel, M. J. & Ramsey, N. F. Automated
19 electrocorticographic electrode localization on individually rendered brain surfaces.
20 *Journal of Neuroscience Methods* **185**, 293–298 (2010).
- 21 92. Papademetris, X. *et al.* BiImage Suite: An integrated medical image analysis suite: An
22 update. *Insight J* **2006**, 209 (2006).
- 23 93. Groppe, D. M. *et al.* iELVis: An open source MATLAB toolbox for localizing and visualizing
24 human intracranial electrode data. *Journal of neuroscience methods* **281**, 40–48 (2017).

- 1 94. Dykstra, A. R. *et al.* Individualized localization and cortical surface-based registration of
2 intracranial electrodes. *NeuroImage* **59**, 3563–3570 (2012).
- 3 95. Yekutieli, D. & Benjamini, Y. Resampling-based false discovery rate controlling multiple test
4 procedures for correlated test statistics. *Journal of Statistical Planning and Inference* **82**,
5 171–196 (1999).
- 6 96. Pinheiro, J. C. & Bates, D. M. Unconstrained parametrizations for variance-covariance
7 matrices. *Stat Comput* **6**, 289–296 (1996).
- 8 97. Maxwell, S. E., Delaney, H. D. & Kelley, K. *Designing Experiments and Analyzing Data: A*
9 *Model Comparison Perspective*. (Routledge, 2017). doi:10.4324/9781315642956.
- 10

Figure 1: Data and design. **a)** Schematic of the behavioral task. ITI is used as the pre-stimulus baseline. **b)** MNI maps of all electrode sites. Each color represents a different subject. Note: MNI maps in the main figures plot all sites on left hemisphere for display purposes. **c)** MNI maps of ROI sites (colored circles) and non-ROI sites (black dots). Sites were anatomically parcellated into ROIs using each subject's native-space cortical surface (see Methods). **Panel d-f)** depict the single-site analysis pipeline using data from an exemplar site in mid-cingulate cortex. **d)** Heatmap of HFB power during other-mentalizing across timepoints (x-axis) and trials (y-axis). Black areas indicate timepoints after RT_{Task} , which were discarded from analyses. **e)** Heatmap of single-trial analysis results using the data in Panel D. Brown areas indicate timepoints with significant activations ($p_{FDR} < .05$; corrected for number of timepoints, trials, and sites) relative to the pre-stimulus baseline preceding each trial (-200-0 ms). Gray areas indicate timepoints with nonsignificant responses. Single-trial analysis provides four key metrics of the HFB response. Onset latency (green squares) is the earliest timepoint with a significant response. Peak latency and peak power (white squares) are the timepoint and magnitude, respectively, of the strongest significant response. Offset latency is the latest timepoint with a significant response (red squares). **f)** Timecourses of evoked HFB power (β) estimated by trial-averaged analysis. Thick solid lines indicate significant responses relative to the pre-stimulus baseline ($p_{FDR} < .05$; corrected for number of timepoints and sites). Thin dashed lines indicate nonsignificant HFB responses. Shaded areas indicate standard error of β . This analysis was used to identify sites with significant HFB responses for each task condition. **Abbreviations:** ROI = region of interest; MNI = Montreal Neurological Institute; s = seconds; ms = millisecond; RT_{Task} = behavioral response time; ITI = inter-trial interval (pre-stimulus baseline); Visual = visual cortex; ATL = anterior temporal lobe; TPJ = temporoparietal junction; PMC = posteromedial cortex; amPFC = anteromedial prefrontal cortex; dmPFC = dorsomedial prefrontal cortex; vmPFC = ventromedial prefrontal cortex; HFB = high-frequency broadband; FDR = false discovery rate; NS = nonsignificant ($p_{FDR} > .05$)

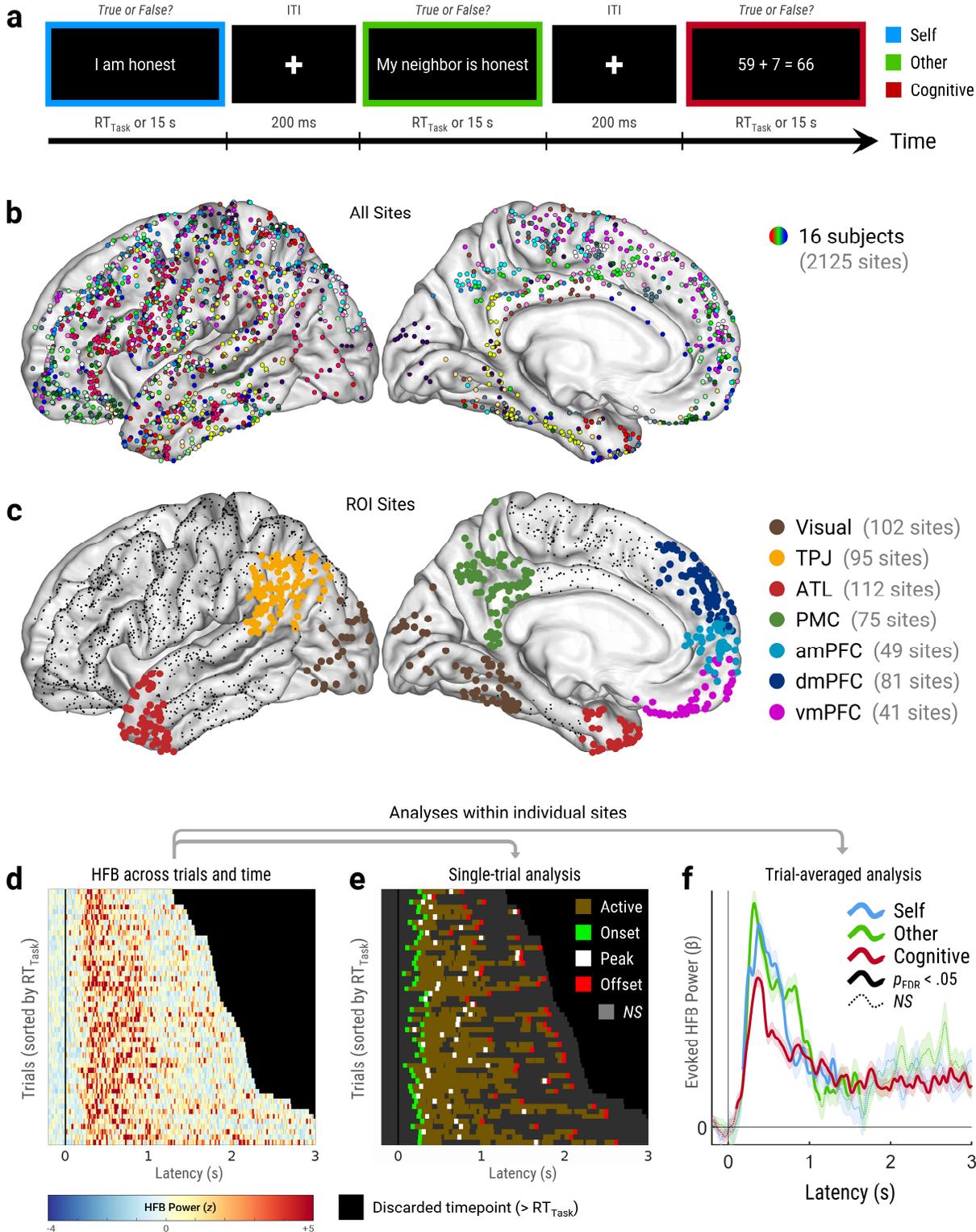


Figure 2: Exemplar ROI sites. **a)** MNI map of exemplar ROI sites. Circle fill color indicates self/other selectivity, which was measured by t -tests of single-trial HFB peak power ($p_{FDR} < .05$; corrected for number of sites). Circle outline color indicates significant HFB response to the cognitive task, if any. All sites are plotted on left hemisphere for display purposes. **Panels b-k** show timecourses of evoked HFB power (β) estimated by trial-averaged analysis of the ROI sites indicated in Panel A. Thick solid lines indicate significant HFB responses relative to the pre-stimulus baseline ($p_{FDR} < .05$; corrected for number of timepoints and sites). Thin dashed lines indicate nonsignificant HFB responses. Shaded areas indicate standard error of β . †excluded from ROI-level analyses due to outlier thresholds.

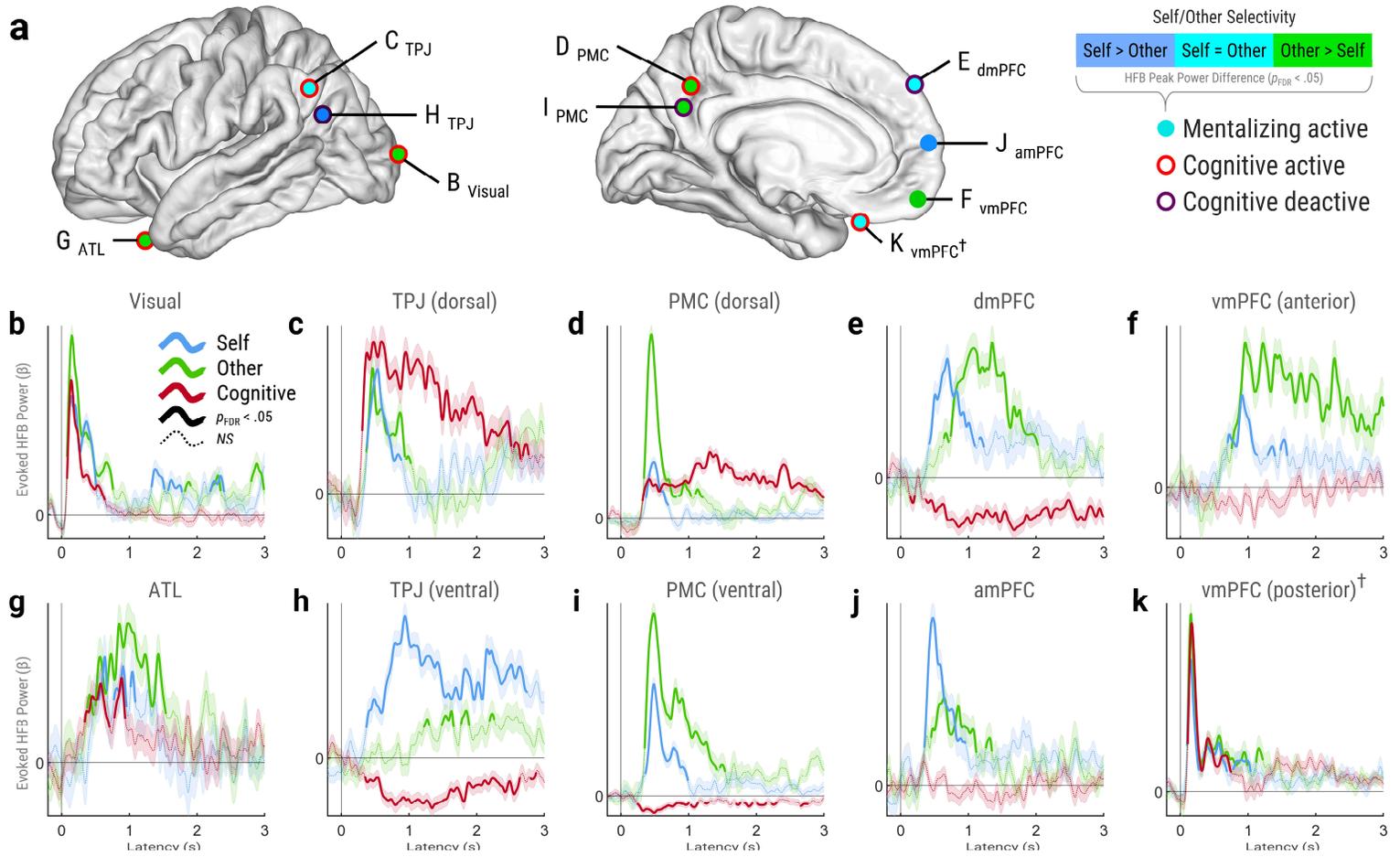


Figure 3: A neurocognitive pathway for mentalizing. **Panels ac** show MNI maps with approximate ROI outlines. All sites plotted on left hemisphere. **Panels bdef** show results from precise native-space ROI parcellation. ‘Whole Brain’ refers to all relevant sites in the entire brain. **a)** MNI map of sites identified as active, inactive, or nonresponsive for mentalizing via trial-averaged analysis ($p_{FDR} < .05$; corrected for number of timepoints and sites). **b)** Percentages of sites exhibiting the response types in Panel A. **c)** Functional specificity of mentalizing-active sites as identified by trial-averaged analysis ($p_{FDR} < .05$; corrected for number of timepoints and sites). Sites were considered ‘mentalizing-specific’ (light and dark turquoise) if they were mentalizing-active but not cognitive-active. Sites were considered ‘non-specific’ (pink) if they coactivated for mentalizing and the cognitive task. **d)** Percentages of mentalizing-active sites featuring the response types in Panel C. **e)** Mean activation latencies evoked by mentalizing across ROIs. The left and right edges of the bars indicate onsets and offsets, respectively, while diamonds indicate peaks (see Fig. 1e). Error bars depict standard error of the mean. **f)** Pairwise ROI contrasts for onset, peak, and offset latencies ($p_{FDR} < .05$; corrected for number of unique ROI pairs). The black diagonal squares show mean latencies for each ROI. The off-diagonal squares show estimated latency differences between ROI pairs, such that ROI(x) - ROI(y). Blue squares indicate significantly earlier latencies in ROI(x) versus ROI(y). Orange squares indicate significantly later latencies in ROI(x) versus ROI(y). White squares indicate nonsignificant differences. Contrast results were simply inverted across the diagonal. Each contrast was restricted to subjects with sites in both ROIs. **Abbreviations:** Mz = mentalizing (collapsed across self and other); Cog = cognitive task (arithmetic).

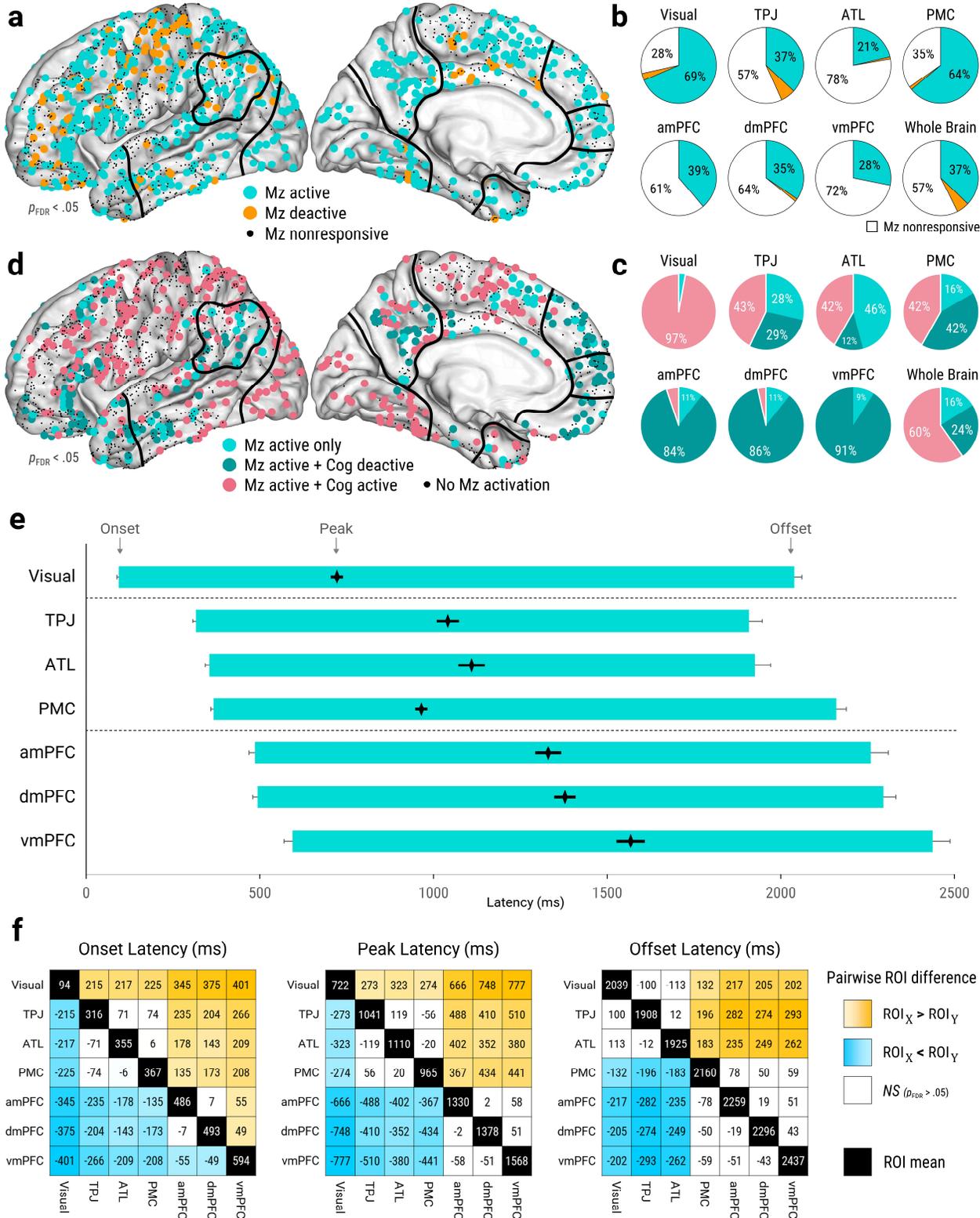


Figure 4: Self/other differences. **a)** Functional anatomy of self- and other-mentalizing. Circles indicate sites with significant HFB activations for both self- and other-mentalizing, colored by the t-score of self/other differences in HFB peak power ($p_{FDR} < .05$; corrected for number of sites). Squares indicate sites with significant activations for only one mentalizing type. Dots indicate sites with nonsignificant mentalizing activations. All sites plotted on left hemisphere with approximate ROI outlines. **b)** Percentages of mentalizing-active sites featuring the response types in Panel A. **c)** Mean activation latencies across mentalizing type and ROIs. The left and right edges of the bars indicate onsets and offsets, respectively, while diamonds indicate peaks. Error bars depict standard error. Asterisks indicate significant self/other differences in peak (black) and offset (red) latencies ($p < .05$, controlled for RT_{Task}). **d)** Mean HFB peak power across mentalizing type and ROI. Asterisks indicate significant self/other differences ($p < .05$, controlled for RT_{Task}). **e)** Scatterplots of HFB peak power and RT_{Task} in all mentalizing trials. Slopes (b) are shown as black diagonal lines, indicating the change in HFB peak power (z-scored) for every one-second increase in RT_{Task} . Asterisks indicate significant slopes ($p < .05$, controlled for self/other differences).

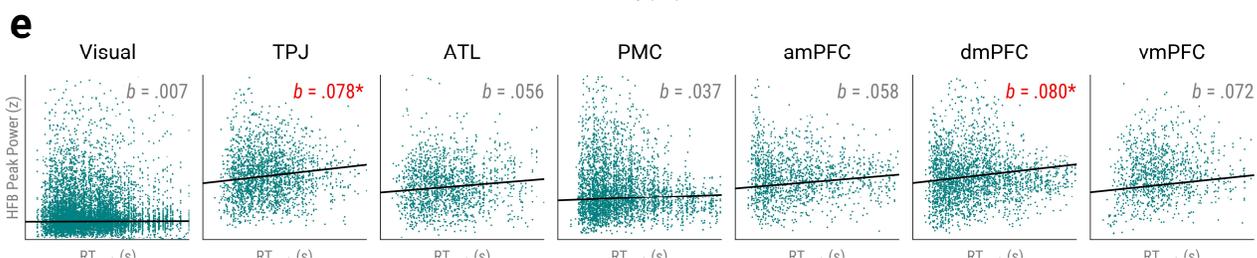
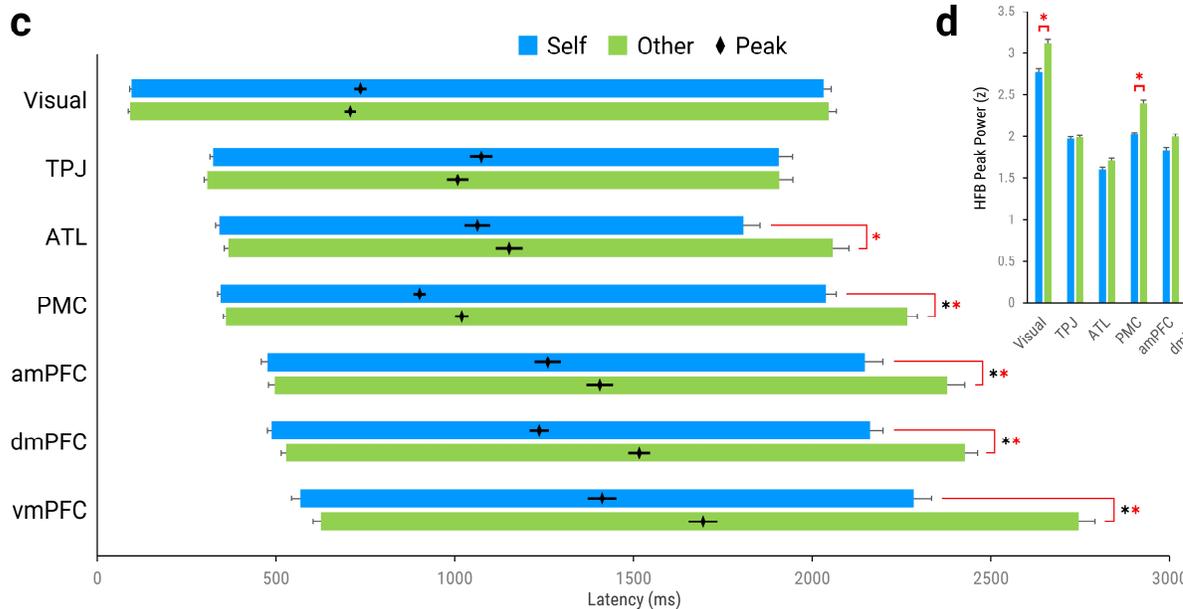
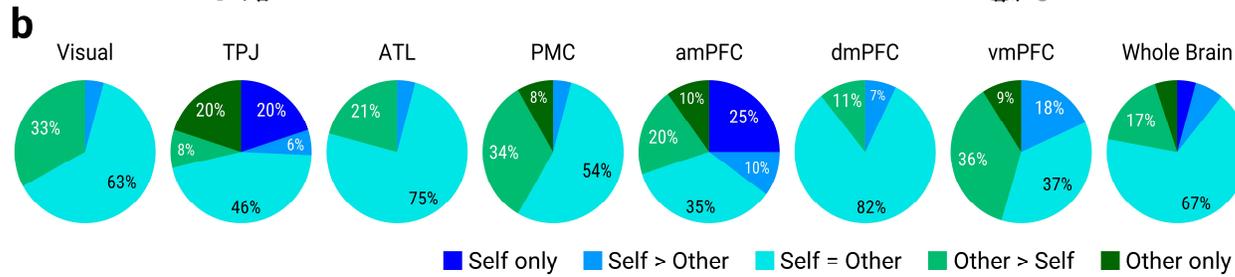
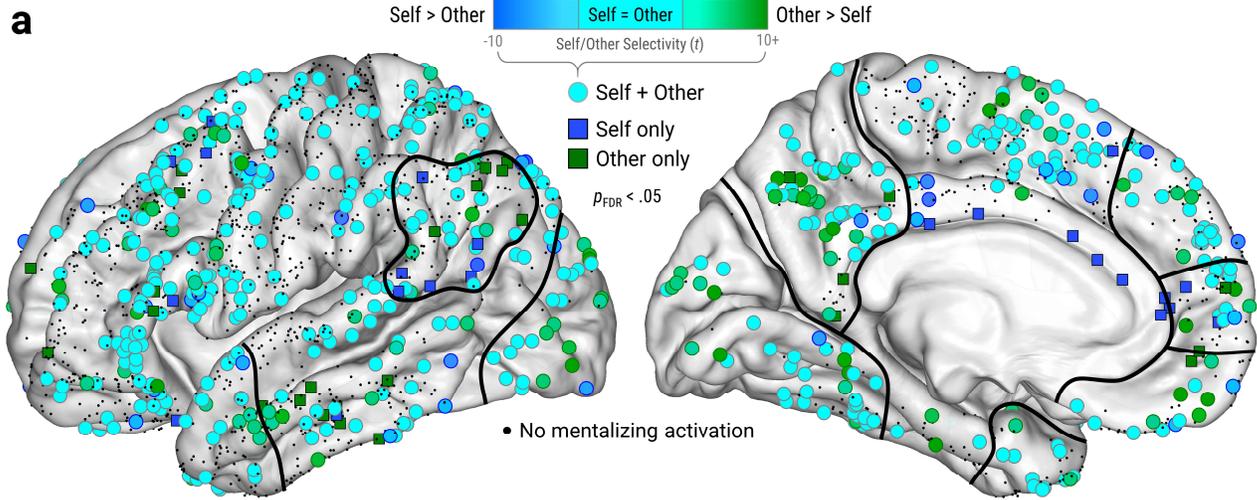


Figure 5: Single-site onset latencies and grand-average ROI timecourses. **a)** Mean mentalizing HFB onset latencies of mentalizing-active sites using single-trial analysis (see Figure 1E). Sites are overlaid on MNI maps with approximate ROI outlines. **b-h)** Grand-average HFB timecourses of ROI sites for each task condition. Thick solid lines indicate significant HFB responses relative to the pre-stimulus baseline ($p_{FDR} < .05$; corrected for number of timepoints and ROIs). Thin dashed lines indicate nonsignificant HFB responses. Shaded areas indicate standard error of β .

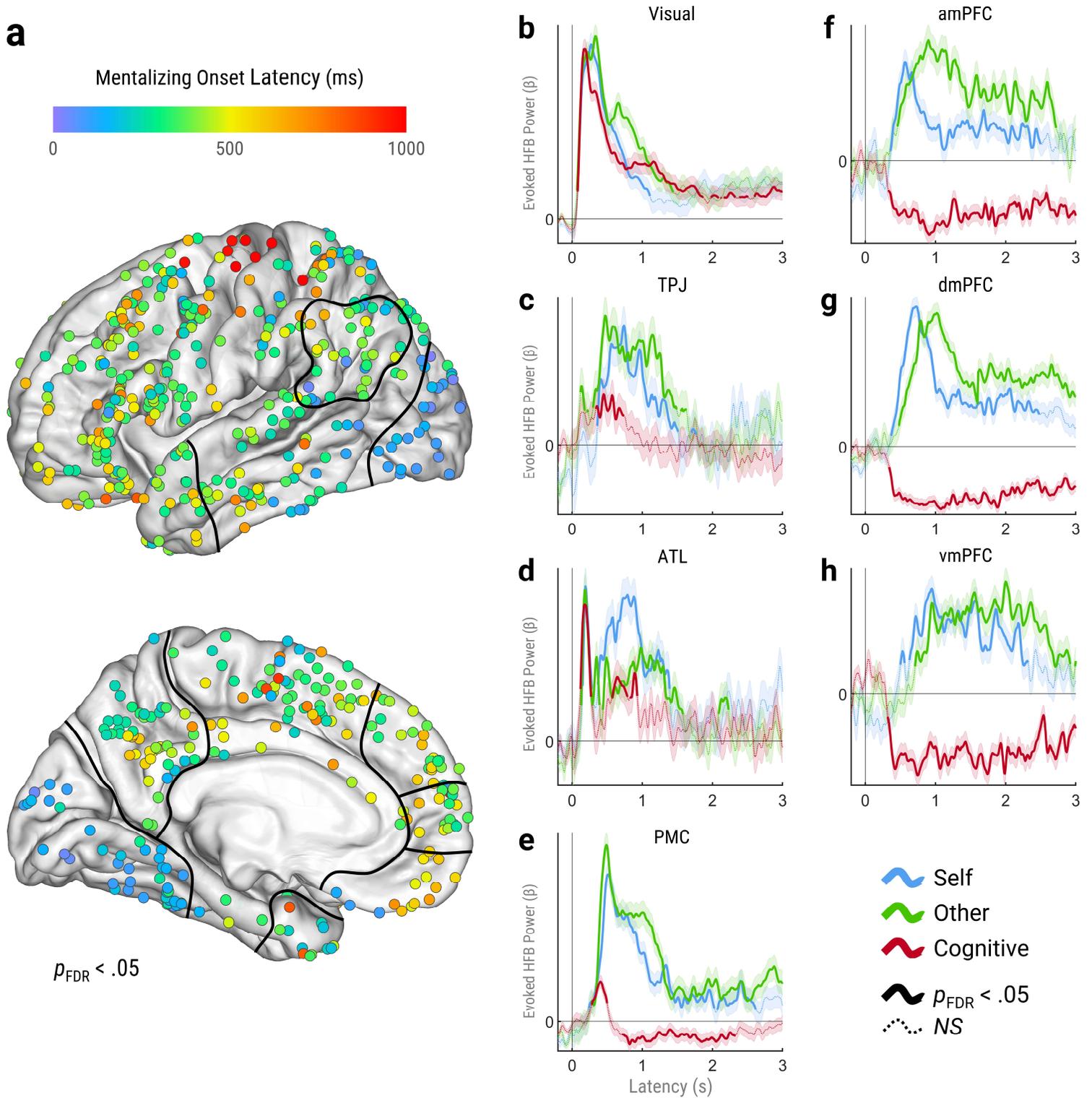


Figure 6: Summary of task-evoked neuronal responses. Sites with significant HFB responses (relative to pre-stimulus baseline) in during specific time windows ($p_{FDR} < .05$; corrected for number of sites and time windows). Sites plotted on left hemisphere with approximate ROI outlines for display purposes.

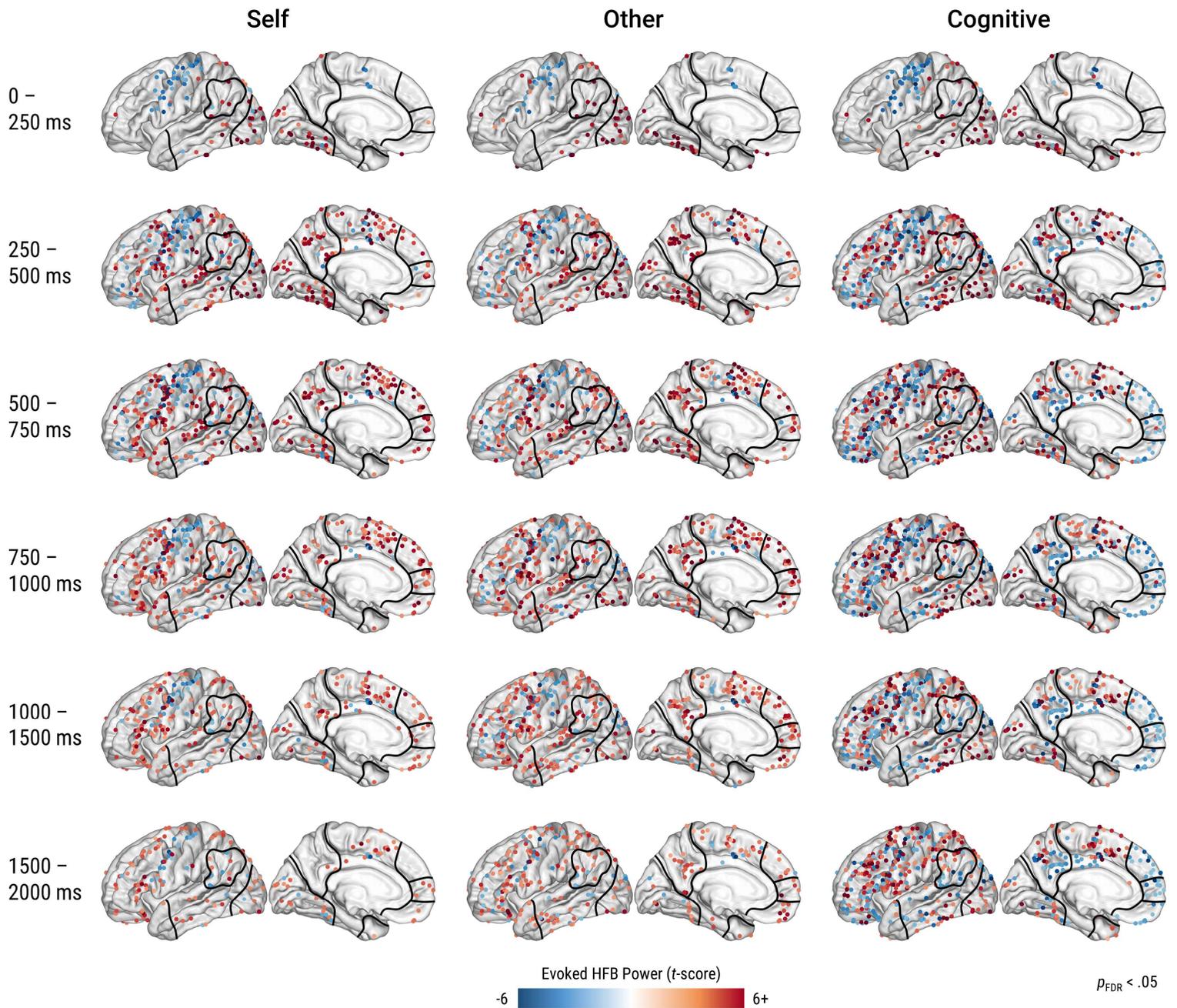


Table 1. ROI-level results for each single-trial HFB metric. The top section shows omnibus tests of the full-factorial linear mixed-effects models (LMEM). The middle and bottom sections show the effects of Condition (other-self) and RT_{Task} within each ROI, with Condition and RT_{Task} controlling for each other. All LMEMs accounted for subject- and site-specific heterogeneity by including Subject and Site as nested grouping factors. Unequal variances, sample sizes, and autocorrelation were accounted for by using full-Cholesky covariance matrices and Satterthwaite approximation for degrees of freedom (*DF*). **Abbreviations:** ROI = region of interest; ms = milliseconds; Cond = condition (other - self); RT_{Task} = behavioral response time; *b* = LMEM coefficient; Visual = visual cortex; TPJ = temporoparietal junction; ATL = anterior temporal lobe; PMC = posteromedial cortex; amPFC = anteromedial prefrontal cortex; dmPFC = dorsomedial prefrontal cortex; vmPFC = ventromedial prefrontal cortex

Omnibus	Onset Latency (ms)			Peak Latency (ms)			Offset Latency (ms)			Peak Power (z)		
	<i>F</i>	<i>DF</i> _{1/2}	<i>p</i>	<i>F</i>	<i>DF</i> _{1/2}	<i>p</i>	<i>F</i>	<i>DF</i> _{1/2}	<i>p</i>	<i>F</i>	<i>DF</i> _{1/2}	<i>p</i>
ROI	24.65	6/30	<.001	3.544	6/34	0.008	3.573	6/32	0.009	0.854	6/68	0.533
Cond (Other-Self)	0.518	1/33	0.477	15.42	1/26	0.001	35.37	1/14	<.001	12.18	1/94	0.001
RT _{Task}	38.26	1/92	<.001	352.6	1/16	<.001	2140	1/16	<.001	17.17	1/13	0.001
ROI x Cond	1.681	6/18	0.183	2.973	6/62	0.013	1.914	6/30	0.115	1.375	6/29	0.258
ROI x RT _{Task}	9.327	6/30	<.001	12.25	6/31	<.001	6.483	6/34	<.001	0.388	6/19	0.878
Cond x RT _{Task}	3.286	1/67	0.074	0.088	1/12	0.772	0.152	1/42	0.698	0.646	1/31	0.428
ROI x Cond x RT _{Task}	2.136	6/40	0.070	1.194	6/47	0.326	1.111	6/50	0.369	1.059	6/23	0.415
Cond in ROI	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
Visual _{Other-Self}	-5	10	0.648	-6	16	0.719	9	25	0.704	0.332	0.036	<.001
TPJ _{Other-Self}	-4	34	0.901	-17	20	0.402	2	14	0.895	0.048	0.030	0.110
ATL _{Other-Self}	14	25	0.583	38	44	0.382	97	43	0.024	0.106	0.077	0.169
PMC _{Other-Self}	28	19	0.144	43	17	0.010	112	37	0.002	0.249	0.024	<.001
amPFC _{Other-Self}	21	26	0.410	131	31	<.001	156	40	<.001	0.131	0.099	0.184
dmPFC _{Other-Self}	45	33	0.172	168	28	<.001	158	45	<.001	0.004	0.027	0.888
vmPFC _{Other-Self}	31	34	0.352	190	47	<.001	231	43	<.001	0.276	0.162	0.089
RT _{Task} in ROI	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
Visual	0.004	0.002	0.016	0.195	0.026	<.001	0.629	0.033	<.001	0.007	0.030	0.811
TPJ	0.081	0.019	<.001	0.297	0.048	<.001	0.655	0.078	<.001	0.078	0.024	0.001
ATL	-0.009	0.011	0.407	0.284	0.052	<.001	0.572	0.069	<.001	0.056	0.032	0.075
PMC	0.035	0.014	0.013	0.261	0.033	<.001	0.659	0.042	<.001	0.037	0.027	0.175
amPFC	0.057	0.021	0.007	0.300	0.050	<.001	0.670	0.064	<.001	0.058	0.040	0.146
dmPFC	0.090	0.016	<.001	0.358	0.033	<.001	0.776	0.044	<.001	0.080	0.019	<.001
vmPFC	0.045	0.024	0.065	0.404	0.087	<.001	0.724	0.079	<.001	0.072	0.045	0.108

Figures

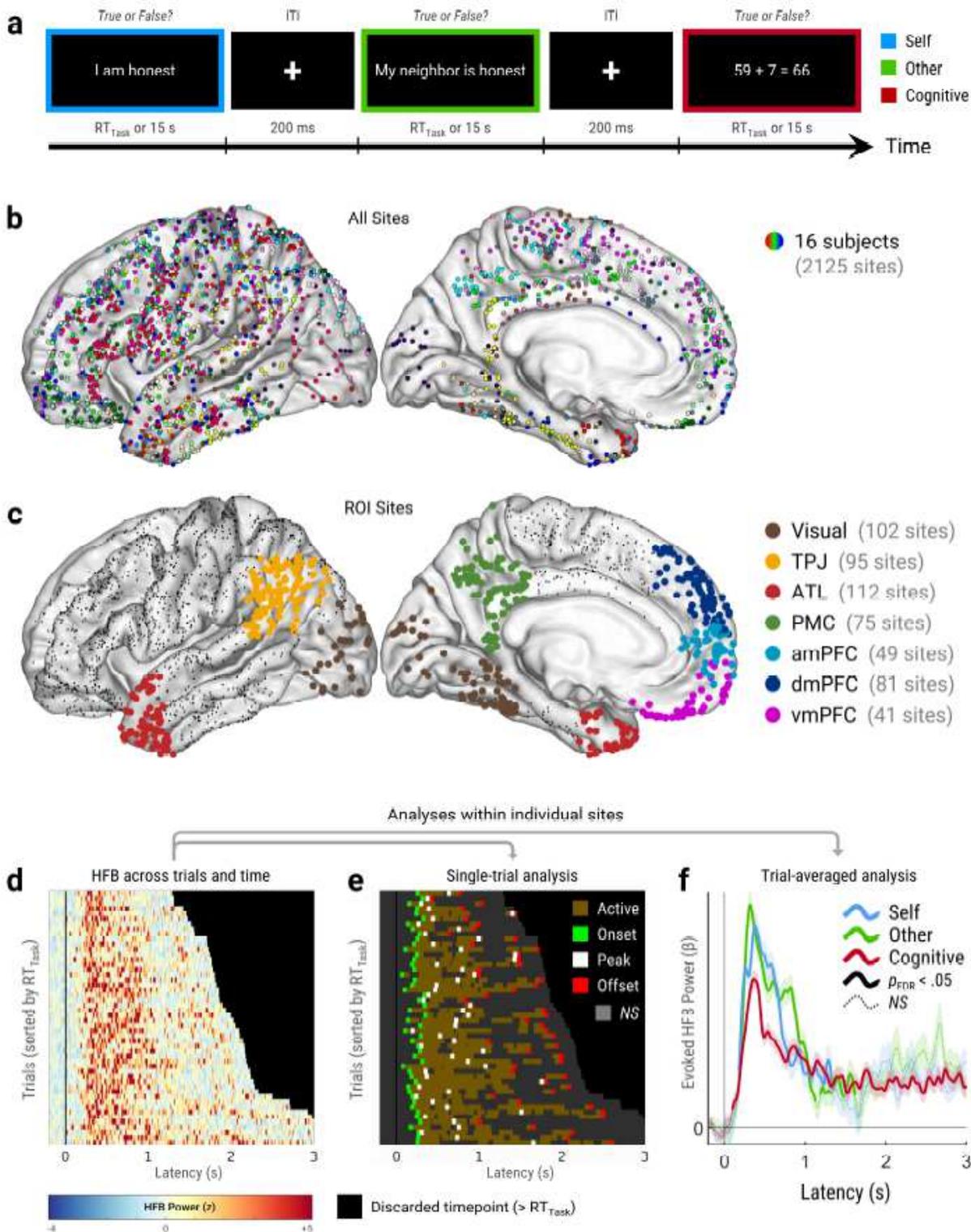


Figure 1

Data and design. a) Schematic of the behavioral task. ITI is used as the pre-stimulus baseline. b) MNI maps of all electrode sites. Each color represents a different subject. Note: MNI maps in the main figures plot all sites on left hemisphere for display purposes. c) MNI maps of ROI sites (colored circles) and non-

ROI sites (black dots). Sites were anatomically parcellated into ROIs using each subject's native-space cortical surface (see Methods). Panels d-f depict the single-site analysis pipeline using data from an exemplar site in mid-cingulate cortex. d) Heatmap of HFB power during other-mentalizing across timepoints (x-axis) and trials (y-axis). Black areas indicate timepoints after RTTask, which were discarded from analyses. e) Heatmap of single-trial analysis results using the data in Panel D. Brown areas indicate timepoints with significant activations ($p_{FDR} < .05$; corrected for number of timepoints, trials, and sites) relative to the pre-stimulus baseline preceding each trial (-200-0 ms). Gray areas indicate timepoints with nonsignificant responses. Single-trial analysis provides four key metrics of the HFB response. Onset latency (green squares) is the earliest timepoint with a significant response. Peak latency and peak power (white squares) are the timepoint and magnitude, respectively, of the strongest significant response. Offset latency is the latest timepoint with a significant response (red squares). f) Timecourses of evoked HFB power (β) estimated by trial-averaged analysis. Thick solid lines indicate significant responses relative to the pre-stimulus baseline ($p_{FDR} < .05$; corrected for number of timepoints and sites). Thin dashed lines indicate nonsignificant HFB responses. Shaded areas indicate standard error of β . This analysis was used to identify sites with significant HFB responses for each task condition. Abbreviations: ROI = region of interest; MNI = Montreal Neurological Institute; s = seconds; ms = millisecond; RTTask = behavioral response time; ITI = inter-trial interval (pre-stimulus baseline); Visual = visual cortex; ATL = anterior temporal lobe; TPJ = temporoparietal junction; PMC = posteromedial cortex; amPFC = anteromedial prefrontal cortex; dmPFC = dorsomedial prefrontal cortex; vmPFC = ventromedial prefrontal cortex; HFB = high-frequency broadband; FDR = false discovery rate; NS = nonsignificant ($p_{FDR} > .05$)

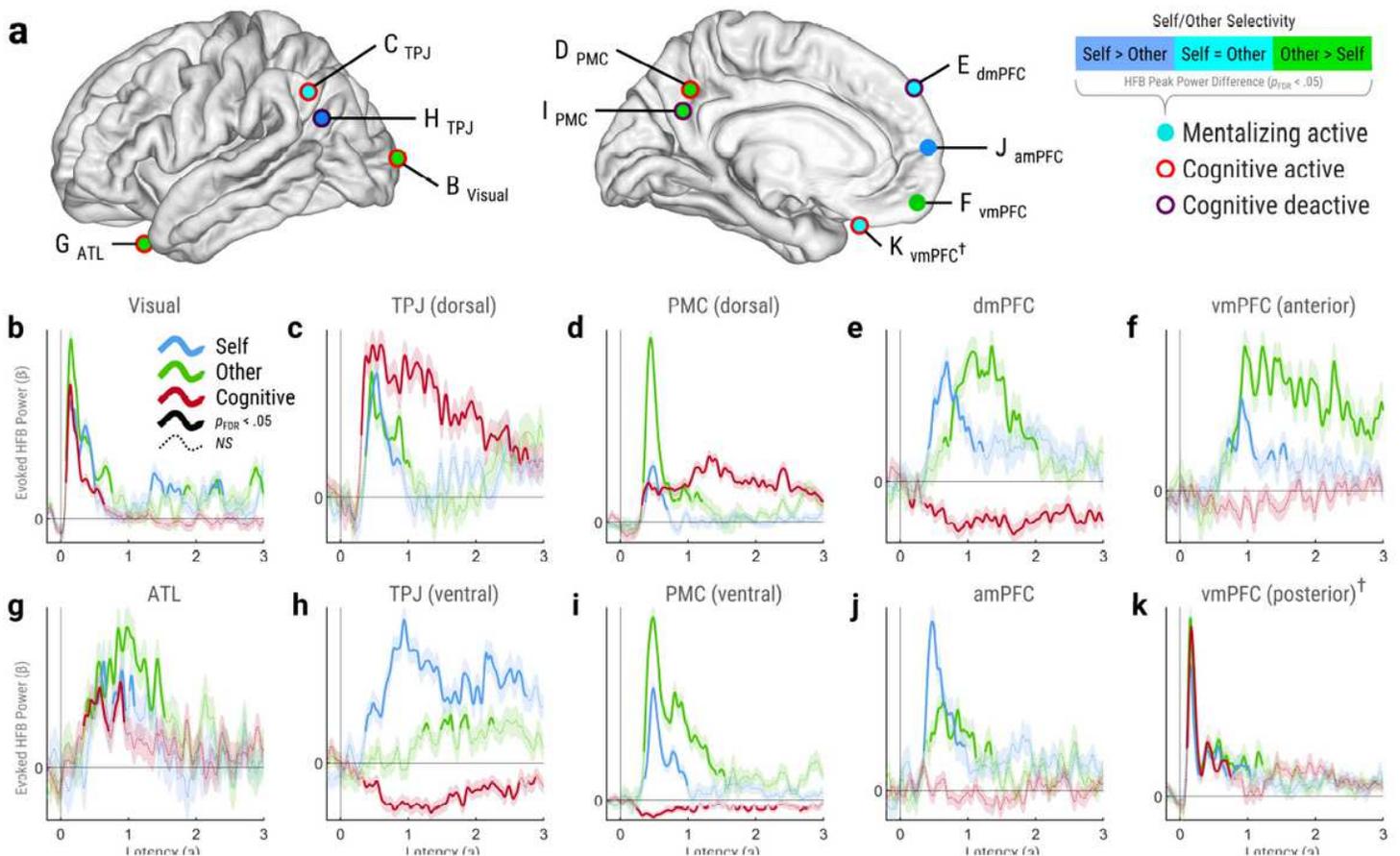


Figure 2

Exemplar ROI sites. a) MNI map of exemplar ROI sites. Circle fill color indicates self/other selectivity, which was measured by t-tests of single-trial HFB peak power ($pFDR < .05$; corrected for number of sites). Circle outline color indicates significant HFB response to the cognitive task, if any. All sites are plotted on left hemisphere for display purposes. Panels b-k show timecourses of evoked HFB power (β) estimated by trial-averaged analysis of the ROI sites indicated in Panel A. Thick solid lines indicate significant HFB responses relative to the pre-stimulus baseline ($pFDR < .05$; corrected for number of timepoints and sites). Thin dashed lines indicate nonsignificant HFB responses. Shaded areas indicate standard error of β . †excluded from ROI-level analyses due to outlier thresholds.

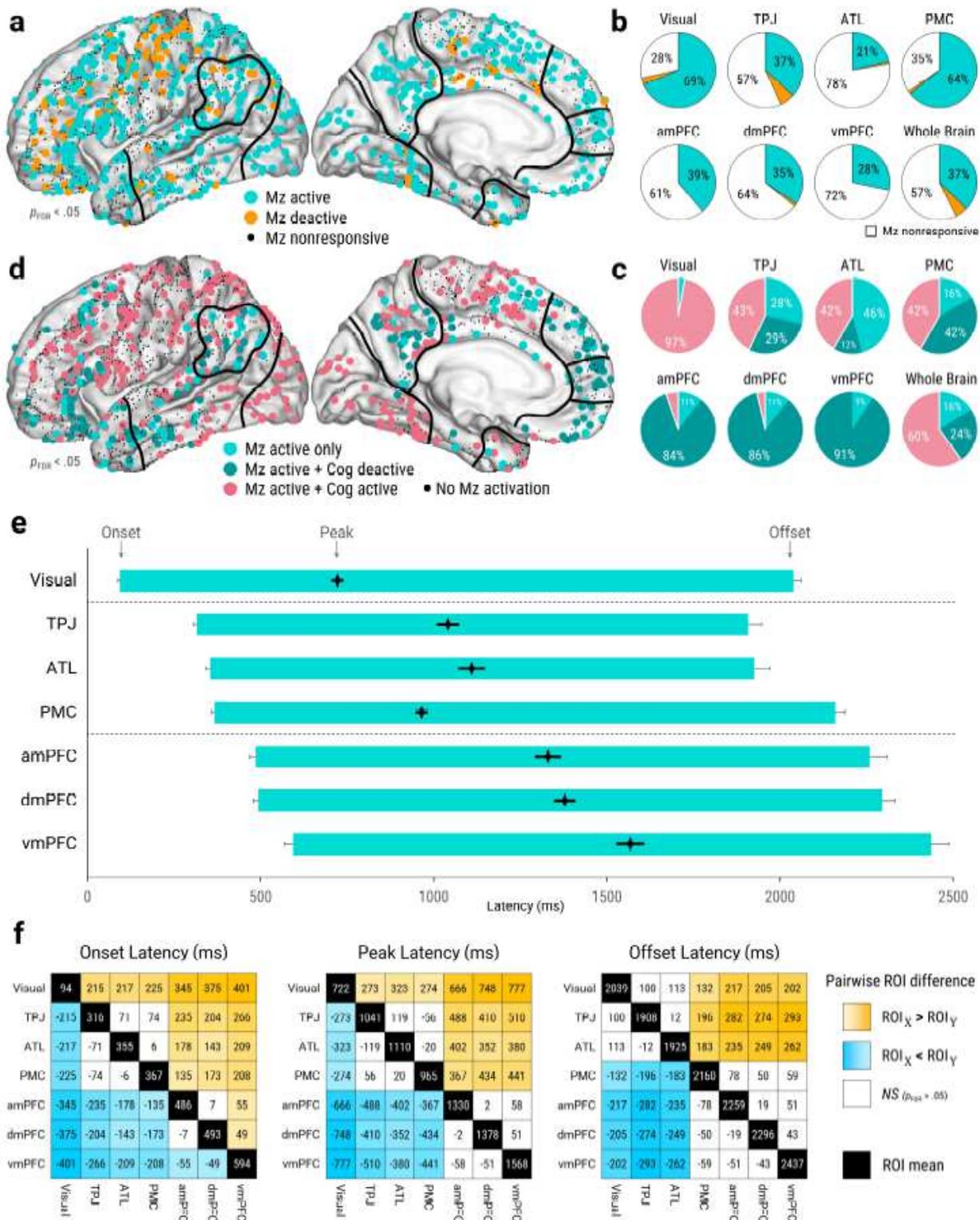


Figure 3

A neurocognitive pathway for mentalizing. Panels ac show MNI maps with approximate ROI outlines. All sites plotted on left hemisphere. Panels bdef show results from precise native-space parcellation. ‘Whole Brain’ refers to all relevant sites in the entire brain. a) MNI map of sites identified as active, inactive, or nonresponsive for mentalizing via trial-averaged analysis ($p_{FDR} < .05$; corrected for number of timepoints and sites). b) Percentages of sites exhibiting the response types in Panel A. c) Functional

specificity of mentalizing-active sites as identified by trial-averaged analysis ($pFDR < .05$; corrected for number of timepoints and sites). Sites were considered 'mentalizing-specific' (light and dark turquoise) if they were mentalizing-active but not cognitive-active. Sites were considered 'non-specific' (pink) if they coactivated for mentalizing and the cognitive task. d) Percentages of mentalizing-active sites featuring the response types in Panel C. e) Mean activation latencies evoked by mentalizing across ROIs. The left and right edges of the bars indicate onsets and offsets, respectively, while diamonds indicate peaks (see Fig. 1e). Error bars depict standard error of the mean. f) Pairwise ROI contrasts for onset, peak, and offset latencies ($pFDR < .05$; corrected for number of unique ROI pairs). The black diagonal squares show mean latencies for each ROI. The off-diagonal squares show estimated latency differences between ROI pairs, such that $ROI(x) - ROI(y)$. Blue squares indicate significantly earlier latencies in $ROI(x)$ versus $ROI(y)$. Orange squares indicate significantly later latencies in $ROI(x)$ versus $ROI(y)$. White squares indicate nonsignificant differences. Contrast results were simply inverted across the diagonal. Each contrast was restricted to subjects with sites in both ROIs. Abbreviations: Mz = mentalizing (collapsed across self and other); Cog = cognitive task (arithmetic).

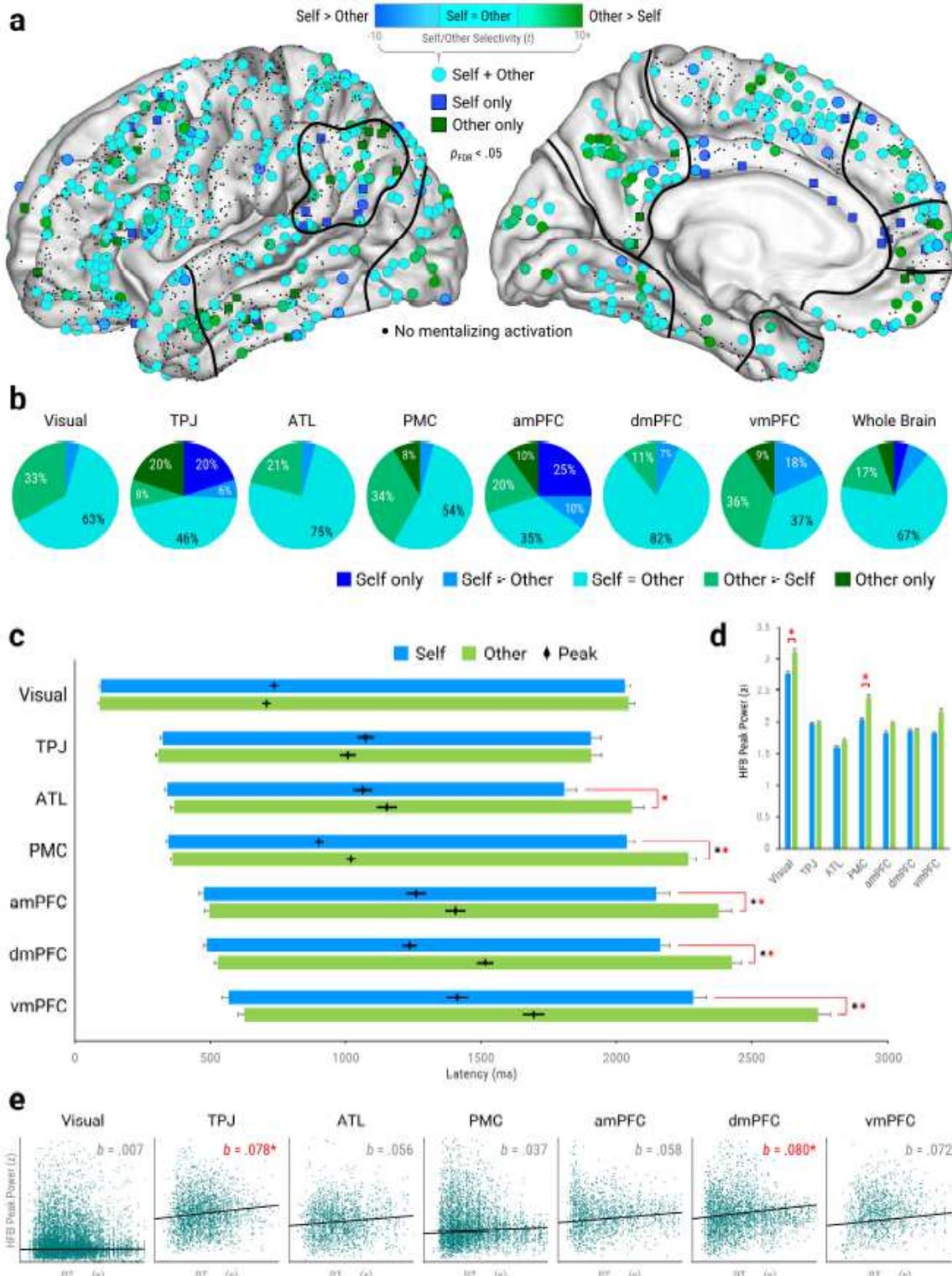


Figure 4

Self/other differences. a) Functional anatomy of self- and other-mentalizing. Circles indicate sites with significant HFB activations for both self- and other-mentalizing, colored by the t-score of self/other differences in HFB peak power ($p_{FDR} < .05$; corrected for number of sites). Squares indicate sites with significant activations for only one mentalizing type. Dots indicate sites with nonsignificant mentalizing activations. All sites plotted on left hemisphere with approximate ROI outlines. b) Percentages of

mentalizing-active sites featuring the response types in Panel A. c) Mean activation latencies across mentalizing type and ROIs. The left and right edges of the bars indicate onsets and offsets, respectively, while diamonds indicate peaks. Error bars depict standard error. Asterisks indicate significant self/other differences in peak (black) and offset (red) latencies ($p < .05$, controlled for RTTask). d) Mean HFB peak power across mentalizing type and ROI. Asterisks indicate significant self/other differences ($p < .05$, controlled for RTTask). e) Scatterplots of HFB peak power and RTTask in all mentalizing trials. Slopes (b) are shown as black diagonal lines, indicating the change in HFB peak power (z-scored) for every one-second increase in RTTask. Asterisks indicate significant slopes ($p < .05$, controlled for self/other differences).

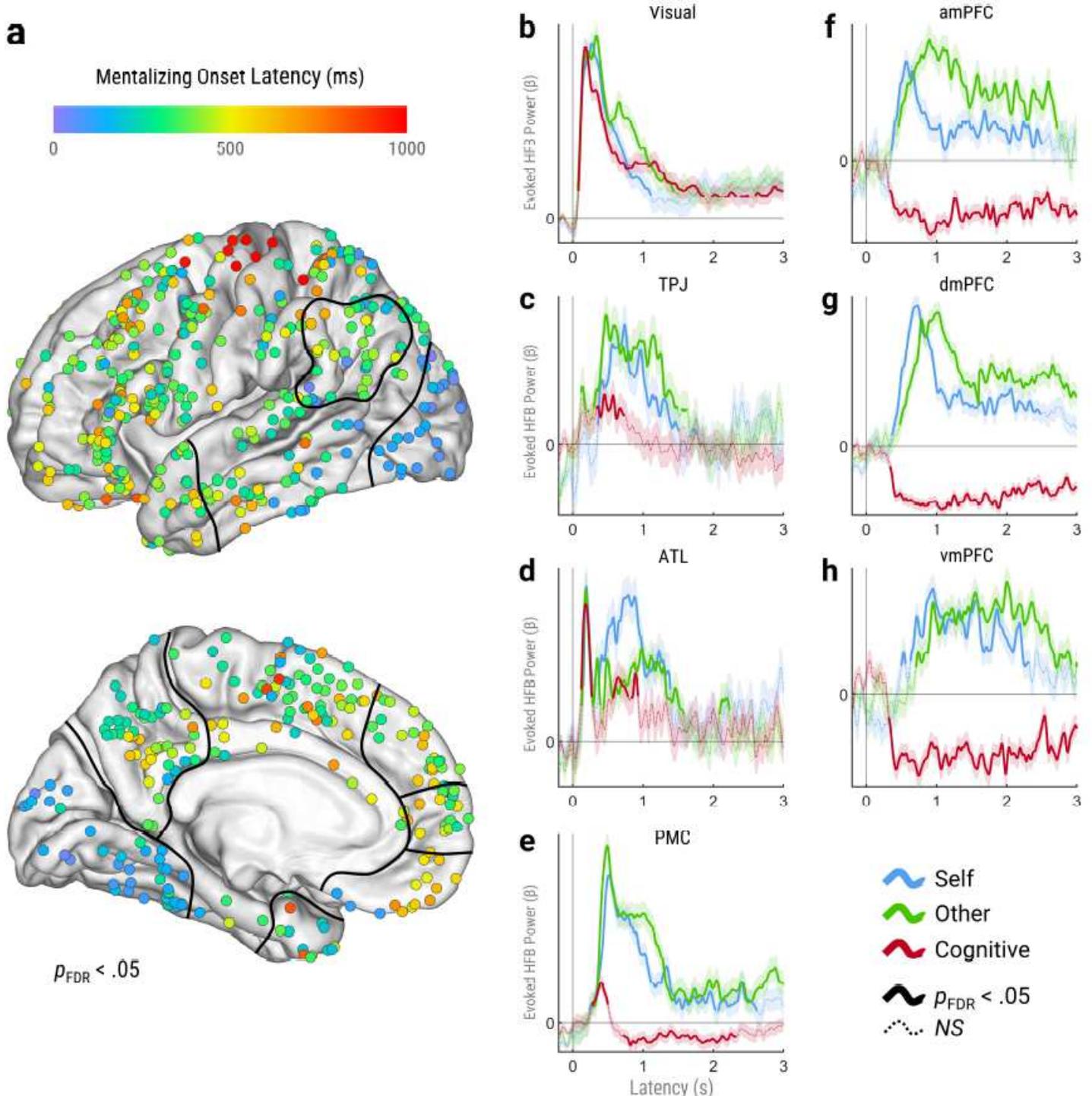


Figure 5

Single-site onset latencies and grand-average ROI timecourses. a) Mean mentalizing HFB onset latencies of mentalizing-active sites using single-trial analysis (see Figure 1E). Sites are overlaid on MNI maps with approximate ROI outlines. b-h) Grand-average HFB timecourses of ROI sites for each task condition. Thick solid lines indicate significant HFB responses relative to the pre-stimulus baseline ($p_{FDR} < .05$; corrected for number of timepoints and ROIs). Thin dashed lines indicate nonsignificant HFB responses. Shaded areas indicate standard error of β .

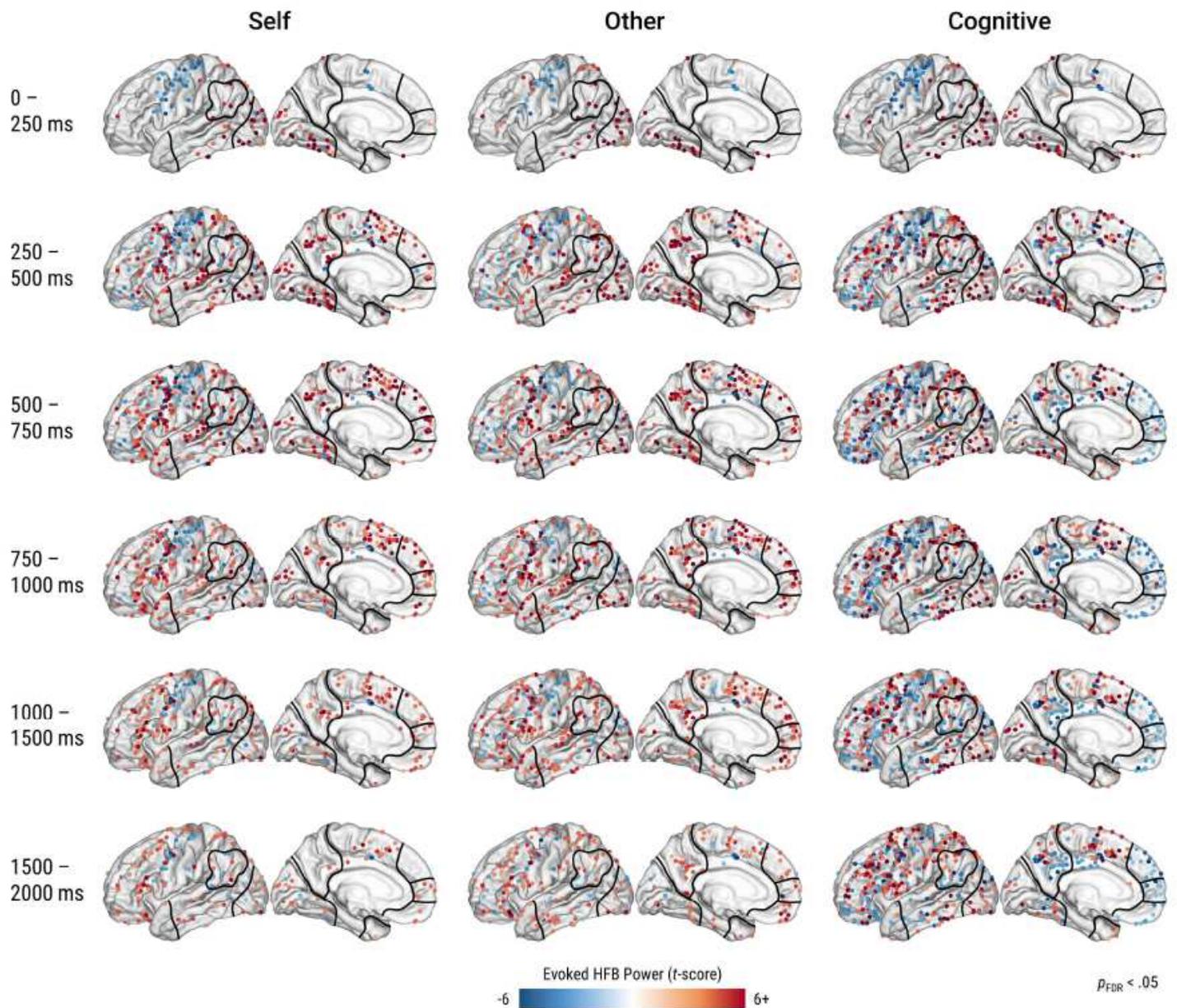


Figure 6

Summary of task-evoked neuronal responses. Sites with significant HFB responses (relative to pre-stimulus baseline) in during specific time windows ($p_{FDR} < .05$; corrected for number of sites and time windows). Sites plotted on left hemisphere with approximate ROI outlines for display purposes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TanECoGmzSuppFINAL.pdf](#)