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

Kevin M. Tan¹, Amy L. Daitch², Pedro Pinheiro-Chagas², Kieran C.R. Fox²,³, Josef Parvizi²,³, Matthew D. Lieberman¹

¹ Social Cognitive Neuroscience Laboratory, Department of Psychology, University of California, Los Angeles, CA, USA
² Laboratory of Behavioral and Cognitive Neuroscience, Department of Neurology and Neurological Sciences, Stanford University, CA, USA
³ School of Medicine, Stanford University, CA, USA

Counts:
- Title = 14
- Abstract = 150
- Main text (intro, results, discussion) = 4523
- Methods = 2000
- References = 97
Abstract

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

Humans are social by nature: our central nervous systems have evolved many mechanisms to support our rich and complex social worlds\textsuperscript{1}. Although high levels of sociality are seen throughout the animal kingdom\textsuperscript{2,3}, humans are exceptional in their capacity for mentalizing – the ability to consider the mental states of others and oneself\textsuperscript{4,5}. The field of social neuroscience seeks to understand how mentalizing and other social functions are implemented at the level of brain and biology\textsuperscript{6}. In humans, social neuroscience primarily relies on functional magnetic resonance imaging (fMRI), a neuroimaging modality with higher anatomical resolution but low temporal resolution\textsuperscript{7}. Hundreds of fMRI studies have shown that mentalizing recruits default mode network (DMN) regions – including medial prefrontal cortex (mPFC), temporoparietal junction (TPJ), and posteromedial cortex (PMC) – with remarkable consistency across countless mentalizing tasks instantiated in various sensory modalities\textsuperscript{5,8–12}. Nevertheless, the specific social cognitive functions of individual DMN regions remain unclear. When seen through fMRI, DMN regions appear to respond concurrently, yet electrophysiological studies demonstrate that critical neurocognitive dynamics occur at millisecond timescales throughout the DMN\textsuperscript{13}. Thus, the limited temporal resolution of fMRI may preclude more precise neurocognitive accounts of mentalizing and its component processes.

Several studies have investigated the fast spatiotemporal dynamics of mentalizing using source-space electroencephalography (EEG) and magnetoencephalography (MEG), neuroimaging modalities with high temporal resolution but coarser spatial resolution\textsuperscript{14}. These studies reveal a general spatiotemporal sequence of brain recruitment, starting in visual cortex, followed by mirror neuron system regions (MNS; e.g., intraparietal sulcus, premotor cortex), and lastly DMN regions\textsuperscript{15–20}. These findings exemplify the emerging consensus that visual representations are used by MNS to identify observable actions (e.g., grasping for food), which are then used by DMN to infer unobservable mental states (e.g., hunger)\textsuperscript{8,21–24}. Taken together,
MEG/EEG studies of mentalizing suggest that visual cortex, MNS, and DMN act as a hierarchical neurocognitive pathway that transforms low-level visual inputs into high-level mentalistic inferences. However, despite broad agreement at the network level, these studies report inconsistent recruitment across individual DMN regions. These inconsistencies may reflect limitations in MEG/EEG source localization, particularly in deeper regions such as mPFC and PMC, which were not sampled in many of these studies. As such, the spatiotemporal dynamics of mentalizing processing across individual DMN regions remains unclear.

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

Results

Data and design

We recruited sixteen human subjects who had electrocorticography (ECoG) electrodes neurosurgically implanted onto the cortical surface for epilepsy monitoring and treatment (Supplementary Table 2). Our behavioral task (Fig. 1a) consisted of true/false text prompts for three experimental conditions of interest: self-mentalizing (e.g. “I am honest”), other-mentalizing (e.g. “my neighbor is honest”), and a control cognitive task involving simple arithmetic (e.g. “9+86=95”). ECoG was recorded from all 2125 electrode sites in our subject cohort (Fig. 1b). Sites showing epileptic activity were excluded from primary analyses. We analyzed the high-frequency broadband signal (HFB; 70-180 Hz), which reflects the aggregate spiking of neuronal populations immediately adjacent to an electrode site.
We began by parcellating the brain into seven regions-of-interests (ROIs; Fig. 1c) using each subject’s native-space cortical surface reconstruction (see Methods). We included six DMN ROIs that are strongly implicated in mentalizing: temporoparietal junction (TPJ), anterior temporal lobe (ATL), posteromedial cortex (PMC), anteromedial prefrontal cortex (amPFC), dorsomedial prefrontal cortex (dmPFC), and ventromedial prefrontal cortex (vmPFC). Visual cortex was included as a control ROI. Out of 2125 electrode sites, 555 were included in our ROIs.

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

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

To explore the fast spatiotemporal dynamics of mentalizing processing, we first examined its functional-anatomical foundations. To this end, we began by identifying sites that showed any significant response to self- or other-mentalizing, regardless of functional specificity (Fig. 3ab). Using trial-averaged results, sites were considered ‘mentalizing-active’ (light turquoise) or ‘mentalizing-deactive’ (orange) if they produced higher or lower HFB power, respectively, relative to pre-stimulus baseline ($p_{FDR} < .05$). Sites were considered ‘mentalizing-nonresponsive’ if they produced non-significant HFB responses to mentalizing. We found
mentalizing-active sites in nearly all parts of cortex (Fig. 3a). In contrast, mentalizing-deactive sites were rarer and concentrated in somatomotor and executive regions. Overall, the majority of sites throughout the brain were mentalizing-nonresponsive (57% whole-brain; Fig. 3b).

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

A spatiotemporal sequence of mentalizing activations from visual cortex to mPFC

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

Pairwise ROI contrasts ($p_{FDR} < .05$; Fig. 3f) revealed that visual cortex produced the earliest activation onsets ($M=94$ ms, $SE=5$) of any ROI. Next, mid-latency onsets were produced...
by TPJ ($M=316$ ms, $SE=9$), ATL ($M=355$ ms, $SE=12$), and PMC ($M=367$ ms, $SE=8$), with non-significant differences between them. Later onsets were produced by amPFC ($M=486$ ms, $SE=17$) and dmPFC ($M=493$ ms, $SE=14$), with non-significant differences between them. Finally, the latest onsets were produced by vmPFC ($M=594$ ms, $SE=25$). However, three vmPFC sites on the posterior orbital surface produced very early activations (<200 ms; Figs. 5a, 2k, S7) that met outlier exclusion criteria for ROI analyses (see Methods). Peak latencies showed the same pattern of cross-ROI differences as onsets, though differences across mPFC ROIs were non-significant (Fig. 3f). Offset latencies exhibited the least differentiation across ROIs. The earliest offsets were produced by visual cortex, TPJ, and ATL, with non-significant differences between them. The latest offsets were produced by PMC, amPFC, dmPFC, and vmPFC, with non-significant differences between them. Crucially, despite these robust cross-ROI latency differences, all ROIs activated concurrently from 594-1908 ms, with even longer activation overlap between specific ROI pairs (Fig. 3e).

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

**Activation onset latency predicts mentalizing-specificity**

Thus far, we have revealed spatial gradients in the timing (Fig. 3ef) and functional specificity (Fig. 3cd) of neuronal population responses to mentalizing. To examine the correspondence between these gradients, we correlated mean onset latencies (black squares in Fig. 3f) and mentalizing-specificity (percentage of mentalizing-specific sites; Fig. 3d) across all seven ROIs. Despite this small sample, we found a near-perfect positive correlation between
onset latency and mentalizing-specificity ($r(5)=0.98, p=9.32e-5$), indicating that ROIs with later activations had greater mentalizing-specificity.

These results portray a hierarchical neurocognitive pathway\textsuperscript{28–31} that reflects both the timing and specificity of mentalizing processing across ROIs. This pathway begins in visual cortex, which featured the earliest activations and the least mentalizing-specificity. Temporoparietal DMN regions (TPJ, ATL, and PMC) appear to be intermediate stages of the pathway, as these regions featured mid-latency activations and intermediate mentalizing-specificity. The highest stages of the pathway appear to be mPFC regions (amPFC, dmPFC, and vmPFC), which featured the latest activations and overwhelming mentalizing-specificity.

\textit{Self- and other-mentalizing share a common neuroanatomical basis}

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

We found that mentalizing-active sites were overwhelmingly coactive for both mentalizing types, regardless of self/other selectivity (non-selective + self-greater + other-greater = 91% whole-brain overlap; Fig. 4b). Moreover, non-selective sites formed the largest
single category in all ROIs (range: 35-82%) and the whole-brain (67%). Among selective sites, we compared amounts of self-selective (self-only + self-greater) versus other-selective (other-only + other-greater) sites (McNemar $\chi^2$, Yates-corrected; df=1). This revealed that other-selective sites significantly ($p<.05$) outnumbered self-selective sites in visual cortex (4% self/33% other; $\chi^2=13.88$), PMC (4% self/42% other; $\chi^2=13.14$), and the whole-brain (11% self/22% other; $\chi^2=23.14$). Non-significant self/other differences were found in TPJ (26% self/29% other; $\chi^2=0$), ATL (4% self/21% other; $\chi^2=1.50$), amPFC (35% self/30% other; $\chi^2=0$), dmPFC (7% self/11% other; $\chi^2=0$), and vmPFC (18% self/45% other; $\chi^2=0.57$). In sum, self- and other-mentalizing recruited near-identical neuronal populations in a broadly non-selective manner, though selective sites were predominantly other-selective in visual cortex and PMC.

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

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

We found that self- and other-mentalizing evoked a common spatiotemporal sequence of activations across ROIs (Fig. 4c). Within this sequence, onset latencies showed non-significant Condition differences. However, significantly later peaks were evoked by other-versus self-mentalizing in PMC, amPFC, dmPFC, and vmPFC. Similarly, significantly later offsets were evoked by other-versus self-mentalizing in all DMN ROIs except TPJ. Analysis of activation duration (number of active timepoints) confirms that significantly longer activations were evoked by other-versus self-mentalizing in PMC, amPFC, dmPFC, and vmPFC (see
Supplementary Report & Fig. S1). In contrast, self/other magnitude differences were only
significant in visual cortex and PMC, where greater peak power was evoked by other- versus
self-mentalizing (Fig. 4d). Critically, LMEMs indicated that these results were dissociable from
self/other differences in RT_{Task}, among other confounds (see Methods).

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

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

Behavioral response times are best predicted by TPJ and dmPFC activity

To explore the relationship between neuronal population activity and mentalizing task
performance, we compared behavioral response times (RT_{Task}) with single-trial HFB metrics
(e.g. onsets, peaks, offsets) from mentalizing-active ROI sites using LMEMs (Table 1). Onset
latencies in visual cortex, TPJ, PMC, amPFC, and dmPFC significantly ($p<.05$) positively predicted $RT_{Task}$. Unsurprisingly, peak and offset latencies in all ROIs significantly positively predicted $RT_{Task}$. In contrast, peak power (i.e., activation magnitude) in only TPJ and dmPFC significantly predicted $RT_{Task}$ (Fig. 4e). Lastly, we examined the correlations between $RT_{Task}$ effect size ($b$-coefficients in Table 1) and mean activation latencies (Fig. 3f) across ROIs using the same correlation method as described above. This revealed that ROIs with later activations had stronger $RT_{Task}$ associations (peak latency: $r(5)=.96$, $p=.0008$; offset latency: $r(5)=.78$, $p=.038$; peak power: $r(5)=.77$, $p=.041$), though the correlation for onset latency was non-significant ($r(5)=.49$, $p=.269$). In sum, activations in later-stage ROIs had the strongest $RT_{Task}$ associations. Nevertheless, TPJ and dmPFC best predicted $RT_{Task}$ as they were the only ROIs with significant $RT_{Task}$ effects in all HFB metrics.

Summary of task-evoked neuronal population activity

To summarize the spatiotemporal dynamics of neuronal population responses to each task condition, we identified which sites produced significant HFB responses (relative to pre-stimulus baseline; $p_{FDR}<.05$) within specific time-windows (Fig. 6). From 0-250 ms, activations were largely localized to visual cortex and showed little differentiation across task conditions. From 250-500ms, activations spread beyond visual cortex to encompass temporoparietal and lateral frontal cortex. During this time window, mentalizing and the cognitive task (arithmetic) began to diverge. In temporoparietal DMN regions, self- and other-mentalizing primarily evoked activations, while arithmetic evoked a mix of activations and deactivations. Self-mentalizing also evoked some mPFC activations. From 500-750 ms, all task conditions evoked mPFC responses. Self-mentalizing evoked stronger mPFC activations than other-mentalizing. In contrast, arithmetic evoked mPFC deactivations, which continued for all successive time windows. From 750-1000 ms, self- and other-mentalizing evoked similar responses. From 1000-
1500 ms, other-mentalizing evoked more sustained activations than self-mentalizing, particularly in amPFC and dmPFC. From 1500-2000 ms, mentalizing-evoked activations weakened, except in vmPFC, which sustained strong activations for other-mentalizing.

Effects of task condition on behavioral response times

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

Discussion

Using electrocorticography (ECoG), we probed the neurocognitive substrates of mentalizing with unprecedented spatiotemporal resolution. We found that mentalizing about the self and others recruited near-identical neuronal populations (Fig. 4ab) in a common spatiotemporal sequence (Figs. 4c & 6). Within our ROIs, activations began in visual cortex, then spread to temporoparietal DMN regions (TPJ, ATL, and PMC), and lastly to mPFC regions (amPFC, dmPFC, and vmPFC; Fig. 3ef). Critically, regions with later activations exhibited greater functional specialization for mentalizing as measured by three metrics: functional specificity for mentalizing versus arithmetic (Fig. 3d), self/other differentiation in activation latencies (Fig. 4c), and prediction of behavioral response times ($RT_{Task}$; Fig. 4e & Table 1). Taken together, these results portray a common neurocognitive pathway for self- and other-mentalizing, beginning in visual cortex (low specialization), ascending through temporoparietal DMN areas (intermediate specialization), then reaching its apex in mPFC regions (high specialization).
Our results are consistent with gradient-based models of brain function, which posit that concrete sensorimotor processing in unimodal regions (e.g., visual cortex) gradually yields to increasingly abstract and inferential processing in 'high-level' trans modal regions like mPFC\textsuperscript{32,33}. We found that the strength of self/other differences in activation latencies increased along a gradient from visual cortex to vmPFC. Specifically, other-mentalizing evoked slower (Fig. 4c) and lengthier (Fig. S1) activations than self-mentalizing in successive DMN ROIs (Table 1). Thus, perhaps because we know ourselves better than others, other-mentalizing may require longer computation times at more abstract and inferential levels of processing. What might these hierarchical neurocognitive dynamics imply about fMRI findings?

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

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

We also found numerous functional distinctions between TPJ and dmPFC, which is surprising given their remarkably similar functional profiles in fMRI literature\textsuperscript{5,6,8–12,44}. Specifically, we found that TPJ produced earlier activations (Fig. 3ef) that were notably coactive.
for mentalizing and arithmetic (cognitive task; Fig. 3cd). Indeed, the onsets and offsets of TPJ activations were the earliest of any DMN ROI (Fig. 3f). In contrast, dmPFC produced significantly later activations (Fig. 3ef) that were overwhelmingly mentalizing-specific (Fig. 3cd), indicating that dmPFC sits at a higher level of mentalizing's neurocognitive pathway than TPJ. Moreover, aggregate ROI analyses revealed no significant self/other differences in TPJ (Table 1), while dmPFC showed robust self/other timing differences (Fig. 4c), suggesting that dmPFC is more sensitive to differences in mentalistic content. Furthermore, even when controlling for self/other differences, dmPFC had stronger RT<sub>Task</sub> associations than TPJ in all HFB metrics (Fig. 4e & Table 1). Strikingly, unlike TPJ, the offsets of dmPFC activations closely preceded RT<sub>Task</sub> (within ~200 ms; Fig. S3), suggesting that dmPFC is more deeply involved in the final stages of mentalistic decision-making. Taken together, while TPJ and dmPFC are both clearly crucial for mentalizing performance, dmPFC appears more specialized for mentalizing itself.

Given the marked functional differentiation between TPJ and dmPFC, what specific neurocognitive roles might they play in mentalizing? In social neuroscience, TPJ is often considered to be a functionally-specific locus for explicit belief reasoning<sup>45,36,46,37</sup>. Yet here, TPJ appears less functionally specialized relative to dmPFC (Figs. 3c-f & 4c-e). To explain this discrepancy, we suggest that TPJ provides crucial antecedents for explicit belief reasoning in dmPFC. Given TPJ's central role in automatic evaluations of thematic semantics<sup>47–55</sup>, we propose that TPJ automatically represents integrative psycho-semantic models of exemplar contexts for a given inference. In simpler terms, TPJ may help us 'see' the psycho-semantic gestalt of a given situation<sup>56</sup>. Accordingly, tasks that 'show' concrete mentalistic content (e.g., social animations, reading the mind in the eyes) reliably recruit TPJ but not dmPFC, while tasks that require mentalistic logical inferences (e.g., false belief, trait judgments) reliably recruit dmPFC in addition to TPJ<sup>10–12</sup>. Thus, when mentalistic content feels 'seeable' from perceptual processing, TPJ could generate mentalistic inferences without explicit belief reasoning. Indeed,
work on implicit and spontaneous mentalizing consistently find that TPJ (but not dmPFC) encodes an actor’s beliefs without any explicit reasoning\textsuperscript{16,57–65}. Taken together, TPJ may implicitly set the psycho-semantic stage for explicit belief reasoning that occurs later in dmPFC when necessary (e.g., the present trait judgment task; Fig. 1a).

The dmPFC may be well-suited for explicit belief reasoning\textsuperscript{66–69}. We found substantial concurrent activation between dmPFC and all other ROIs (Fig. 3ef), suggesting that dmPFC could work iteratively with lower-level regions to refine what is ‘seen’, thus providing dmPFC with increasingly-useful inputs from which to draw better inferences\textsuperscript{70,71}. Moreover, studies on strategic reasoning show that dmPFC can arbitrate between multiple mental models\textsuperscript{72} and prospective choices\textsuperscript{73} by simultaneously evaluating multiple possibilities\textsuperscript{74} through ‘fuzzy’ propositional reasoning\textsuperscript{71,75}. As such, dmPFC may arbitrate between multiple TPJ-generated exemplar contexts to help extract the most relevant and enduring semantic features for a given psychological inference. Taken together, dmPFC may integrate and refine representations throughout mentalizing’s neurocognitive pathway to strategically reason about minds.

As for mentalizing about the self or others, fMRI studies routinely suggest that amPFC is ‘self-selective’ while dmPFC is ‘other-selective’\textsuperscript{38–41}. In both regions, what underlying neuronal population dynamics could result in stronger hemodynamic responses for one mentalizing type over another? The standard assumption would be that the magnitude (i.e., intensity) of neuronal activations differs across mentalizing type. This might be seen in aggregate ROI activity, or perhaps across individual ROI sites. We tested both possibilities by examining HFB activation magnitudes (e.g., peak power) across amPFC and dmPFC. Unexpectedly, aggregate ROI analysis of peak power revealed nonsignificant self/other differences in both ROIs (Fig. 4d & Table 1). Similarly, across individual sites, we found that the vast majority of dmPFC sites showed nonsignificant self/other differences in peak power (82% non-selective), while amPFC contained near-equal amounts of self-selective sites (35%) and other-selective sites (30%; Fig. 4ef).
ab). Critically, self-selective sites and other-selective sites did not significantly outnumber each other within both ROIs. However, amPFC did contain a notable proportion of sites that only produced significant activations for self-mentalizing (25%; Fig. 4b), perhaps explaining amPFC’s ‘self-selective’ hemodynamic responses. In sum, we found that dmPFC produced equivalent activation magnitudes for both mentalizing types, which appears inconsistent with reports of ‘other-selective’ dmPFC responses in fMRI studies.

We instead found robust self/other differences in the timing of dmPFC activations. Specifically, other-mentalizing evoked slower and lengthier HFB activations compared to self-mentalizing (Fig. 4c & S1). In other words, dmPFC activity remained significantly above baseline for longer during other-mentalizing (see Figs. 2e, 5g & 6). This suggests a different account of why dmPFC produces stronger hemodynamic responses for mentalizing about others over oneself. The typical story is that dmPFC is highly specialized for thinking about other people’s minds\(^{38-41}\). However, we found that both mentalizing types recruited the same neuronal populations (100% overlap) at equivalent intensities (82% non-selective) in dmPFC (Fig. 4abd).

Alternatively, dmPFC could be sensitive to the inherently greater difficulty of other-mentalizing, which may necessitate additional processing cycles before completion; it should take longer to reason about other people’s minds. Self-mentalizing may be simplified by the rich compendium of accessible knowledge we have about ourselves, thus resulting in brief but equally intense processing. Given that standard fMRI analysis does not distinguish the intensity of response from the duration of response, it appears the latter has been mistaken for the former.

At the systems level, we revealed complex and hierarchical processing dynamics across mentalizing’s cortical pathway. Mentalizing about the self and others evoked similar spatiotemporal sequences of activation onsets, peaks, and offsets, revealing an overarching processing sequence from visual cortex to vmPFC (Fig. 4c). The sequence of onsets may depict an initial ‘feedforward sweep’ of coarser processing\(^{28,76}\) along this pathway. Indeed, onset
latencies were insensitive to self/other differences (Table 1). Onsets were followed by considerable concurrent activations across all ROIs, which were sustained until the sequence of activation offsets (Figs. 3ef, 4c, 2, 5b-e, 6 & S2). Given that concurrent activation is considered an index of recurrent processing\(^{28,77,78}\), distinct regions of mentalizing's cortical pathway may largely work together within an overarching spatiotemporal sequence. For example, PMC's lengthy activations (Fig. 3e & S1) bridged the gap between the earlier onsets of temporoparietal DMN ROIs and the later offsets of mPFC ROIs (see white squares in Fig. 3f), supporting the idea that PMC is the posterior DMN hub that helps coordinate processing between temporoparietal and prefrontal DMN regions\(^{44,79}\). Intriguingly, in aggregate analyses, self/other differences did not reach significance until concurrent activation was achieved across all ROIs (e.g., peak and offset metrics; Figs. 4cd & Table 1), perhaps signifying the importance of recurrent processes in self/other differentiation. In sum, mentalizing may be supported by a brief initial ‘feedforward sweep’ of coarser processing along the cortical pathway, followed by substantial recurrent processing that may integrate and refine representations across cortical pathway regions. These dynamics could obscure cross-regional functional distinctions in fMRI studies.

This study is not without confounds and limitations. Some of these limitations are inherent to ECoG: the use of epileptic patients, inconsistent brain coverage across subjects, and sampling bias for cortical gyri\(^{26}\). Although these limitations were mitigated to the best of our ability (see Methods), they cannot be completely ameliorated. Thus, our ECoG findings could be corroborated by examining healthy subjects with recent advances in source-space EEG/MEG, such as ultra-high density EEG\(^{80}\), optically-pumped MEG\(^{81}\), and laminar source localization\(^{82}\). Another important confound was the sparse right-hemisphere coverage of our cohort, which may limit the interpretability of our ATL and TPJ results\(^{53,83}\). However, the few right-hemisphere sites in ATL and TPJ appear functionally similar to their left-hemisphere homologues (Figs. S4-S7, Supplementary Tables 3-4). Another limitation was the use of cross-regional timing
differences to infer feedforward and recurrent processes. Future work could use effective
connectivity analyses to better reveal the directionality and causality of information flow during
mentalizing. An additional limitation was the short pre-stimulus baseline in our task (200 ms),
which sometimes contained residual activity from previous trials, thus likely resulting in
artifactual ‘deactivations’ in somatomotor sites (Fig. 6). Another confound was the greater word
count of other-mentalizing prompts (e.g., "My neighbor is...") versus self-mentalizing prompts
(e.g., "I am..."); Fig. 1a), which could conceivably result in longer or stronger activations for other-
mentalizing, although we consider this possibility unlikely outside of visual cortex (Fig. 4a-d).
Relatedly, another limitation may be the ease of our mentalizing task, which involves less
abstract mentalization than, for instance, a typical false belief task84.

Distributed hierarchical processing is a central organizing principle of neurocognitive
systems78,28,33,31. Characterizing such hierarchies has enabled incisive neuromechanistic
accounts of many psychological functions30,85. Here we provide a comprehensive
electrophysiological exploration of the human social brain, revealing that mentalizing is
characterized by complex and hierarchical neurocognitive dynamics at millisecond, millimeter,
and cross-regional scales. While many questions remain, our findings contribute to a solid
foundation upon which more conclusive neurocognitive accounts of mentalizing can be built.

Methods

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

Subjects

We employed a cohort of sixteen patients that underwent neurosurgical treatment for
drug-resistant epilepsy (Supplementary Table 2). Each patient provided written informed
consent to participate in the study, which was approved by the Stanford Institutional Review Board. As part of their presurgical evaluation, patients were implanted with ECoG at Stanford University Medical Center. The anatomical placement of electrode sites was determined according to each patient’s clinical needs. Patients were included in this study’s subject cohort if they had electrode coverage in key DMN regions: mPFC, PMC, TPJ, and ATL. Each patient was monitored in the hospital for six to ten days prior to surgery, during which the study was conducted.

Experimental Task

ECoG data was recorded while subjects performed an event-related experimental task with six conditions (trial types; Fig. 1a). Five of these conditions required true/false responses to written prompts, while one condition consisted of cued rest. Two conditions featured mentalizing prompts, either about oneself (e.g., “I am honest”) or others (e.g., “My neighbor is honest”). Subjects were instructed to select a single neighbor (current or past) as the target for other-mentalizing. Cognitive task trials consisted of basic arithmetic (e.g., “9 + 86 = 95”). Two conditions featured memory-related prompts: episodic (e.g., “I ate candy yesterday”) and self-semantic (e.g., “I eat a lot of candy”). The cued rest condition required no response and displayed a fixation crosshair for 5-10 seconds. The memory and rest conditions were not relevant to the current analyses have been reported elsewhere. Stimuli were presented in a random order and were self-paced, advancing to the next trial after the subjects’ response, or up to 15 seconds if no response. The inter-trial interval (ITI) occurred -200-0 ms before each trial. The experiment was broken into two separate runs (mean run duration = 12.50 ± 1.64 minutes). Subjects were allowed a short break in between the experimental runs. On average, each run featured 25 trials of each sentence condition, 40 cognitive trials, and 36 rest trials. Each non-rest trial contained unique prompts; prompts were not repeated within subject. Responses were
made via a handheld keypad using either the ‘1’ (true) or ‘2’ (false) key. Subjects were instructed to perform the task as accurately and as quickly as possible. All stimuli were presented in white font on a black background using Psychophysics Toolbox 3 (http://psychtoolbox.org/HomePage). Behavioral response times ($RT_{\text{task}}$) across task conditions were analyzed through a linear mixed-effects model (LMM) with Condition nested within Subject to account for subject-specific variance.

Electrocorticography data acquisition

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

Defining regions of interest (ROIs) and brain networks
Each subject's native-space cortical surface reconstruction (e.g., Fig. S2) was used to classify electrode sites into \textit{a priori} ROIs that are strongly implicated in mentalizing, with visual cortex included as a control ROI (Fig. 1c). MNI-based parcellation was avoided due to known transformation inconsistencies in ECoG\textsuperscript{93}. ROIs were defined through FreeSurfer cortical parcellation combined with visual inspection of anatomical landmarks. The ROI for 'visual cortex' consisted of occipital cortex, lingual gyrus, posterior fusiform gyrus, and posterior inferotemporal cortex. The ‘ATL’ ROI consisted of a bilateral anterior subregion of temporal cortex with precentral sulcus as the posterior bound, comprising the temporal poles and adjacent sections of entorhinal cortex and superior, middle, and inferior temporal sulci/gyri. The ‘TPJ’ ROI was a bilateral posterior subregion of inferior parietal lobule with lateral sulcus as the anterior bound, comprising angular gyrus and adjacent sections of supramarginal gyrus and superior temporal sulcus/gyrus. The 'PMC' ROI consisted of precuneus, posterior cingulate, and retrosplenial cortex. The ‘amPFC’ ROI was an mPFC subregion bounded between the ventral and dorsal reaches of corpus callosum. The ‘dmPFC’ ROI was a mPFC subregion ventrally bounded by the amPFC ROI and posteriorly bounded by the callosal rostrum. The ‘vmPFC’ ROI was an mPFC subregion dorsally bounded by the amPFC ROI and posteriorly bounded by the callosal rostrum, including the medial orbitofrontal surface.

\textbf{ECoG preprocessing}

Preprocessing was performed on a single-subject/single-electrode basis using custom routines (\url{https://github.com/LBCN-Stanford/Preprocessing_pipeline}). First, data were notch filtered for power-line noise (57-63 Hz) and harmonics (117-123 Hz, 177-183 Hz). Electrodes were discarded from further analyses if they were marked as pathological or 'noisy' by postclinical evaluation. The data was then rereferenced by subtracting the mean signal of the remaining electrodes from each electrode’s signal. The rereferenced data underwent time-
frequency decomposition into 4-200 Hz spectra in 1-10 Hz bands using 5-cycle Morlet wavelet transforms. The power of the signal in each frequency band was z-transformed across time; this partially corrects for the 1/frequency decay of neurophysiological signals and improves interpretability. Data was then epoched into trials that were time-locked to stimulus onsets, ranging from 200 ms pre-stimulus to 5000 ms post-stimulus (e.g. -200-5000 ms epochs). For each trial and frequency increment, baseline correction was performed by subtracting the mean power across the pre-stimulus baseline period (-200–0 ms) from all timepoints within a trial. To reconstruct the high-frequency broadband (HFB) signal, the primary signal of interest, z-transformed power of frequency bands within 70-180 Hz were averaged to produce a single HFB timecourse per electrode. Trials were rejected from further analyses if they featured epileptic high-frequency oscillations. Lastly, HFB signal from each electrode was low-pass filtered with a gaussian window (width=50 ms) for further analysis.

Within-site analyses

Within-site analyses (Fig. 1d-f) were performed to provide the bases for the primary multi-site analyses. Data after a behavioral response (RT_task) were discarded. All multiple comparisons corrections herein involved maintaining the False Discovery Rate (FDR) under 0.05, with p-values adjusted accordingly (p_{FDR}), using the Benjamini-Yekutieli procedure for data with any dependence structure^95. Correlations between functional specialization metrics and mean HFB latency metrics were performed using Pearson correlations.

Trial-averaged analysis

To identify sites with statistically-significant HFB responses (p_{FDR}<0.05; corrected for number of timepoints and sites within subject), linear mixed-effects models (LMEMs) were used to analyze HFB power (Fig. 1d) during each task condition. The intercept (null distribution)
consisted of timepoints within the pre-stimulus baseline (-200–0 ms). Each peri-stimulus
timepoint (0–RT\textsubscript{task} ms) was represented as a separate dummy variable. The intercept was
nested within trial to account for trial-specific variance. Restricted maximum likelihood
estimation using full-Cholesky parametrization was used to account for autocorrelation and
unequal variance between model terms\textsuperscript{96}. This LMEM specification estimates the mean HFB
response ($\beta$) for each timepoint and task condition (Fig. 1f). To dampen spikes and other noise,
timepoints were not considered significant unless $p_{\text{FDR}}<.05$ was maintained for 50 ms
consecutively. For each task condition, sites were considered ‘active’ or ‘deactive’ if evoked HFB
power was significantly higher or lower than pre-stimulus baseline; if sites produced both, the
polarity of the greatest deflection was used. Sites with nonsignificant differences from baseline
were considered ‘nonresponsive’.

\textit{Single-trial analysis}

Single-trial analysis was performed to provide four key metrics of the HFB response:
onset, peak, and offset latencies, along with peak power (Fig. 1e). For each trial, timepoints
between stimulus onset and RT\textsubscript{task} were run through a sliding window test (width: ±10 ms) to
reveal timepoints with significant HFB responses. Observations (z-scored HFB power; see Fig.
1d) in each sliding window were tested against observations from the pre-stimulus baseline via
two-sample Welch’s $t$-tests to account for unequal variances and sample sizes. This analysis
identified timepoints within individual trials that featured significant stimulus-evoked responses
(brown areas; $p_{\text{FDR}}<.05$, corrected for number of timepoints, trials, and sites) relative to the pre-
stimulus baseline (ITI; -200-0ms) preceding each trial. 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 greatest significant response. Offset latency is
the latest timepoint with a significant response (red squares). To dampen spikes and other
noise, timepoints were not considered significant unless $p_{\text{FDR}} < .05$ was maintained for 50 ms consecutively. Outlier observations were discarded if greater than three median absolute deviations (MAD) from site or six MAD from ROI (within task condition). Sites were excluded from ROI analyses if over 50% of observations exceeded outlier thresholds. Of all ROI sites, only three sites in vmPFC were excluded.

**Multi-site analyses**

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

**Functional specificity and selectivity**

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

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

Grand-average ROI timecourses

To reveal aggregate timecourses of HFB responses within ROIs (Fig. 5b-h), we examined
mean HFB timecourses (trial-averaged β-coefficients) from ROI sites that had significant
(p_{FDR} < .05; corrected for number of site and timepoints) responses for a given condition. The
intercept (null distribution) consisted of timepoints within the pre-stimulus baseline (-200–0
ms). Peri-stimulus timepoints (0–3000 ms) were represented as separate dummy variables.
The intercept was nested within Site and Subject to account for site- and subject-specific
variance.
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 ($p_{\text{FDR}}<0.05$; 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.
References


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 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 ($p_{FDR} < 0.05$; 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 ($\beta$) estimated by trial-averaged analysis. Thick solid lines indicate significant responses relative to the pre-stimulus baseline ($p_{FDR} < 0.05$, corrected for number of timepoints and sites). Thin dashed lines indicate nonsignificant HFB responses. Shaded areas indicate standard error of $\beta$. 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 ($p_{FDR} > 0.05$).
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 ($\beta$) 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 $\beta$. †excluded from ROI-level analyses due to outlier thresholds.
Figure 3: A neurocognitive pathway for mentalizing. Panels a-c show MNI maps with approximate ROI outlines. All sites plotted on left hemisphere. Panels b-d 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, deactive, 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) 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. d) 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* < .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 $\beta$. 

$p_{FDR} < .05$
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 (LMM). The middle and bottom sections show the effects of Condition (other-self) and RT within each ROI, with Condition and RT controlling for each other. All LMMs 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 = 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

<table>
<thead>
<tr>
<th>Omnibus</th>
<th>Onset Latency (ms)</th>
<th>Peak Latency (ms)</th>
<th>Offset Latency (ms)</th>
<th>Peak Power (z)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>DF&lt;sub&gt;1/2&lt;/sub&gt;</td>
<td>p</td>
<td>F</td>
</tr>
<tr>
<td>ROI</td>
<td>24.65</td>
<td>6/30</td>
<td>&lt;.001</td>
<td>3.544</td>
</tr>
<tr>
<td>Cond (Other-Self)</td>
<td>0.518</td>
<td>1/33</td>
<td>0.477</td>
<td>15.42</td>
</tr>
<tr>
<td>RT&lt;sub&gt;Task&lt;/sub&gt;</td>
<td>38.26</td>
<td>1/92</td>
<td>&lt;.001</td>
<td>352.6</td>
</tr>
<tr>
<td>ROI x Cond</td>
<td>1.681</td>
<td>6/18</td>
<td>0.183</td>
<td>2.973</td>
</tr>
<tr>
<td>ROI x RT&lt;sub&gt;Task&lt;/sub&gt;</td>
<td>9.327</td>
<td>6/30</td>
<td>&lt;.001</td>
<td>12.25</td>
</tr>
<tr>
<td>Cond x RT&lt;sub&gt;Task&lt;/sub&gt;</td>
<td>3.286</td>
<td>1/67</td>
<td>0.074</td>
<td>0.088</td>
</tr>
<tr>
<td>ROI x Cond x RT&lt;sub&gt;Task&lt;/sub&gt;</td>
<td>2.136</td>
<td>6/40</td>
<td>0.070</td>
<td>1.194</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cond in ROI</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>b</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>-5</td>
<td>10</td>
<td>0.648</td>
<td>-6</td>
<td>16</td>
<td>0.719</td>
<td>9</td>
<td>25</td>
<td>0.704</td>
<td>0.332</td>
<td>0.036</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TPJ</td>
<td>-4</td>
<td>34</td>
<td>0.901</td>
<td>-17</td>
<td>20</td>
<td>0.402</td>
<td>2</td>
<td>14</td>
<td>0.895</td>
<td>0.048</td>
<td>0.030</td>
<td>0.110</td>
</tr>
<tr>
<td>ATL</td>
<td>14</td>
<td>25</td>
<td>0.583</td>
<td>38</td>
<td>44</td>
<td>0.382</td>
<td>97</td>
<td>43</td>
<td>0.024</td>
<td>0.106</td>
<td>0.077</td>
<td>0.169</td>
</tr>
<tr>
<td>PMC</td>
<td>28</td>
<td>19</td>
<td>0.144</td>
<td>43</td>
<td>17</td>
<td>0.010</td>
<td>112</td>
<td>37</td>
<td>0.002</td>
<td>0.249</td>
<td>0.024</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>amPFC</td>
<td>21</td>
<td>26</td>
<td>0.410</td>
<td>131</td>
<td>31</td>
<td>&lt;.001</td>
<td>156</td>
<td>40</td>
<td>&lt;.001</td>
<td>0.131</td>
<td>0.099</td>
<td>0.184</td>
</tr>
<tr>
<td>dmPFC</td>
<td>45</td>
<td>33</td>
<td>0.172</td>
<td>168</td>
<td>28</td>
<td>&lt;.001</td>
<td>158</td>
<td>45</td>
<td>&lt;.001</td>
<td>0.004</td>
<td>0.027</td>
<td>0.888</td>
</tr>
<tr>
<td>vmPFC</td>
<td>31</td>
<td>34</td>
<td>0.352</td>
<td>190</td>
<td>47</td>
<td>&lt;.001</td>
<td>231</td>
<td>43</td>
<td>&lt;.001</td>
<td>0.276</td>
<td>0.162</td>
<td>0.089</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RT&lt;sub&gt;Task&lt;/sub&gt; in ROI</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>b</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>0.004</td>
<td>0.002</td>
<td>0.016</td>
<td>0.195</td>
<td>0.026</td>
<td>&lt;.001</td>
<td>0.629</td>
<td>0.033</td>
<td>&lt;.001</td>
<td>0.007</td>
<td>0.030</td>
<td>0.811</td>
</tr>
<tr>
<td>TPJ</td>
<td>0.081</td>
<td>0.019</td>
<td>&lt;.001</td>
<td>0.297</td>
<td>0.048</td>
<td>&lt;.001</td>
<td>0.655</td>
<td>0.078</td>
<td>&lt;.001</td>
<td>0.078</td>
<td>0.024</td>
<td>0.001</td>
</tr>
<tr>
<td>ATL</td>
<td>-0.009</td>
<td>0.011</td>
<td>0.407</td>
<td>0.284</td>
<td>0.052</td>
<td>&lt;.001</td>
<td>0.572</td>
<td>0.069</td>
<td>&lt;.001</td>
<td>0.056</td>
<td>0.032</td>
<td>0.075</td>
</tr>
<tr>
<td>PMC</td>
<td>0.035</td>
<td>0.014</td>
<td>0.013</td>
<td>0.261</td>
<td>0.033</td>
<td>&lt;.001</td>
<td>0.659</td>
<td>0.042</td>
<td>&lt;.001</td>
<td>0.037</td>
<td>0.027</td>
<td>0.175</td>
</tr>
<tr>
<td>amPFC</td>
<td>0.057</td>
<td>0.021</td>
<td>0.007</td>
<td>0.300</td>
<td>0.050</td>
<td>&lt;.001</td>
<td>0.670</td>
<td>0.064</td>
<td>&lt;.001</td>
<td>0.058</td>
<td>0.040</td>
<td>0.146</td>
</tr>
<tr>
<td>dmPFC</td>
<td>0.090</td>
<td>0.016</td>
<td>&lt;.001</td>
<td>0.358</td>
<td>0.033</td>
<td>&lt;.001</td>
<td>0.776</td>
<td>0.044</td>
<td>&lt;.001</td>
<td>0.080</td>
<td>0.019</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>vmPFC</td>
<td>0.045</td>
<td>0.024</td>
<td>0.065</td>
<td>0.404</td>
<td>0.087</td>
<td>&lt;.001</td>
<td>0.724</td>
<td>0.079</td>
<td>&lt;.001</td>
<td>0.072</td>
<td>0.045</td>
<td>0.108</td>
</tr>
</tbody>
</table>