

Functional Architecture of the Human Insula

Corresponding Author: Professor Josef Parvizi

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

EDITORIAL NOTE: Given the tone and critiques in Reviewer 5's comments on version 0, we asked other referees with overlapping expertise to comment on these concerns, provided editorial guidance related to them, and did not ask Reviewer 5 to re-review the paper.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)

Summary

This manuscript presents an ambitious and methodologically sophisticated investigation of the functional organization of the human insula using a rare combination of intracranial electrical stimulation, subjective phenomenology, clustering-based parcellation, and directed connectivity analyses in a neurosurgical population. The study has several notable strengths. Most importantly, it leverages a sizeable neurosurgical sample with unusually broad bilateral coverage of the insular cortex, which is a substantial asset for invasive human mapping work of this kind. In addition, the manuscript moves beyond descriptive sensory mapping by examining both “inflow” and “outflow” connectivity across insular subregions, providing an especially valuable systems-level perspective on insular organization. The combination of stimulation-evoked percepts, topographic analyses, clustering, and connectivity makes this a potentially important contribution to the literature. Overall, I found the manuscript interesting, substantial, and potentially impactful. The data are rich and the study has several standout features. My main concern is that some of the interpretive claims, particularly those concerning the interoceptive role of the insula and the functional significance of anterior insula findings, are stronger than the present data can fully support. I think the manuscript would be strengthened by a more restrained and precise framing of what the findings do and do not establish.

Major comments

1. The manuscript's central claims about interoception are currently stronger than the data appear to support. The most important issue concerns the interpretation that the findings establish specific insular subregions, particularly dorsal posterior Region I, as primary interoceptive cortex or as a primary interoceptive hub. The data clearly support an important role for the insula in bodily experience, and particularly for posterior insula involvement in thermal, nociceptive, and body-localized percepts. However, the evidence for a broader and more canonical interoceptive role is less definitive than the manuscript suggests. In particular, relatively few internal organ-system-specific sensations appear to have been elicited. Classic visceral domains such as cardiac, respiratory, gastrointestinal, and genitourinary sensations seem limited or underrepresented. This is important because the Discussion at times moves from bodily sensation broadly to interoception more specifically, without fully resolving that distinction. A more careful differentiation between somatosensory, nociceptive, vestibular/body-state, and organ-specific interoceptive percepts would improve the interpretation.

Relatedly, it would be important to clarify whether participants were specifically queried for internal sensations such as heartbeat, breathing, or visceral changes. Why were cardiac or respiratory sensations apparently so sparse? Is it possible such sensations occurred but were not explicitly queried, captured, or categorized? As currently written, the claim that these findings establish the insula's interoceptive role is somewhat overstated relative to the range of reported phenomenology, particularly in light of the functional neuroimaging literature that has consistently demonstrated insula activity in response to

cardiac interoceptive awareness during resting conditions (Schulz, 2016, *PhilTransRSocB*) and under conditions of cardiac perturbation (Hassanpour et al., 2018, *Neuropsychopharm*).

For similar reasons, the statement that “Together, these converging lines of evidence support the interpretation of dorsal posterior Region I as a primary interoceptive hub” should be softened. The study provides strong evidence that this region is important for evoked bodily percepts and is tightly embedded in local insular circuitry. However, the current findings do not as clearly establish it as a primary interoceptive cortex in the broader sense asserted by the manuscript.

2. The anterior insula interpretation is interesting, but remains somewhat speculative in its current form

One of the most intriguing aspects of the manuscript is the relative lack of subjective phenomenology evoked by stimulation of agranular anterior insula, contrasted with evidence that an anterior insular region exerts strong influence over other insular regions and is preferentially connected with prefrontal cortex. The inflow/outflow connectivity mapping is an especially strong feature of the study and adds meaningful depth beyond traditional stimulation mapping.

That said, I do not think the present evidence is yet sufficient to conclude that the anterior insula therefore functions as an “integrative hub” in the strong sense implied. The current data do not directly establish what information is being integrated, nor do they directly demonstrate integration as a computation. In this respect, the interpretation feels somewhat ahead of the evidence.

I also think there is a plausible alternative explanation for the lack of anterior insula phenomenology during stimulation. To what extent might the context of the experimental stimulation matter? The study does not appear to manipulate interoceptive expectancy, anticipation, uncertainty, or contextual forecasting of bodily signals (see Harrison et al., 2021, *Neuron*; Lovero et al., 2009 *NeuroImage*; Teckentrup et al., 2019 *NeuroImage*). If anterior insula contributes importantly to interoceptive prediction or to the contextual shaping of bodily awareness, then it may not be surprising that it failed to produce robust subjective responses under passive stimulation conditions. Put another way, anterior insula may require an interoceptive expectation or predictive context to prime responsivity and perceptual expression. This possibility seems particularly relevant given that the anterior insula reportedly showed robust responses during the saliency task, which explicitly involved a context likely to engage expectancy-related processes.

I therefore encourage the authors to discuss more explicitly whether the absence of stimulation-evoked anterior insula phenomenology should be interpreted as evidence against interoceptive function, or whether it may instead reflect the kinds of interoceptive computations that are most readily revealed by the experimental context used here.

3. The manuscript would benefit from clearer engagement with conscious access and reportability

The present study relies heavily on stimulation-evoked subjective reports and therefore, by design, focuses on phenomena that reach conscious access. This is not a weakness in itself, but it does place constraints on what can be inferred about insular function more generally. Many bodily signals processed within the insula likely remain outside conscious awareness. It would therefore be useful for the authors to clarify whether they view the present findings as reflecting the general functional organization of the insula, or more specifically the subset of insular processes that become consciously reportable.

In this respect, it may help to engage more directly with existing frameworks in the neuroscience of consciousness. According to Global Neuronal Workspace theory, conscious access is typically defined as the transition of information from nonconscious processing into a reportable state associated with large-scale propagation across distributed cortical networks (Dehaene et al., 2011, *Neuron*; Sergent et al., 2005, *Nature Neuroscience*). The present results are highly relevant to reportable bodily experience, but the Discussion should be careful not to equate this too readily with insular function as a whole.

4. The link to “as-if body loops” appears too speculative

The manuscript’s effort to bridge the present findings to the concept of “as-if body loops” seems premature. The current experiments do not explicitly assess internal simulation, contextual interoceptive forecasting, or the bypassing of the body proper in the sense implied by that framework. To make such a connection convincingly, one would ideally need an experimental design that directly probes expectation, anticipation, simulation, or prediction of bodily states, followed by assessment of how those anticipated states are instantiated neurally independent of incoming peripheral changes. As written, the bridge to “as-if body loops” reads as speculative rather than empirically established. I would therefore recommend either substantially softening this claim or removing it.

This is especially relevant in light of Damasio’s more explicit formulation of the concept in *The Feeling of What Happens*, where the “as if body loop” refers to body-related changes being created directly in sensory body maps under the control of other neural sites, thereby partially or entirely bypassing the body proper (e.g., Damasio, 1999, p. 281). The present study is valuable, but it does not directly test that type of internal simulation mechanism.

5. The structure of the datasets and statistical handling need greater clarity

The manuscript would benefit from clearer explanation of how the various participant subsets relate to one another across stimulation, clustering, connectivity, and task-based analyses. At present, it is difficult to determine whether these cohorts overlap fully, partially, or not at all. Clarifying this would strengthen the methodological transparency and improve interpretability.

Relatedly, the statistical treatment of the data requires additional clarification. The analyses appear to combine multiple electrodes within individuals, which are non-independent observations, as well as comparisons across participants, which are independent observations. It is not clear how this hierarchical structure was handled. This is important because the treatment of within-subject dependence and between-subject variability bears directly on the robustness and generalizability

of the findings.

6. The manuscript would benefit from a clearer crosswalk across anatomical frameworks

Different organizational schemes appear across the manuscript—for example, anterior/posterior distinctions in the subjective report analyses versus clustering-derived Regions I–IV in the connectivity analyses. Although each framework may be justified, the current presentation makes it somewhat difficult to integrate the results into a single coherent anatomical model. A clearer crosswalk or unifying figure linking these parcellation schemes would substantially improve coherence.

7. The rationale for the task-based approach needs to be clarified

The inclusion of the Gradual Onset Continuous Performance Task is not yet fully justified relative to the paper's main aims, since the task does not directly probe body-signal processing. At present, this component feels somewhat disconnected from the main argument, and its findings are not discussed in a way that substantially advances the central claims. If the authors wish to retain this section, they should clarify more explicitly why it is necessary and how it informs the overall model of insular organization. Otherwise, it may dilute the manuscript's focus.

Minor comments

1. Clarify general limitations of electrical cortical stimulation for sensory mapping

The Discussion would be strengthened by acknowledging the broader limitations of electrical neurosurgical stimulation for sensory mapping studies. Although this approach offers a rare and powerful opportunity to probe human brain–experience relationships, stimulation-evoked percepts do not necessarily recapitulate physiological signaling, and the resulting phenomenology may depend on stimulation parameters, local current spread, clinical context, and the structure of subject questioning. These considerations are relevant to how strongly one can infer normal functional architecture from evoked reports.

2. The study does not measure the physiological “body loop” itself

A further limitation is that, although the study evaluates insula activity and connectivity, it does not directly record the relevant physiological signals arising from the body itself—that is, the peripheral component of the “body loop” in Damasio's formulation. This matters because it remains difficult to determine whether insula stimulation may have triggered downstream autonomic or visceral changes that then contributed to subsequent sensory representations. Without concurrent measurement of physiological outputs, it is challenging to disambiguate percepts arising more directly from cortical stimulation from those that may have been mediated, amplified, or shaped by stimulation-evoked changes in the body. I think this limitation is especially important in a manuscript making claims about interoceptive processing, because the body proper is not actually measured and is therefore effectively absent from the inference chain.

3. Clarify the relationship between stimulation and connectivity analyses

Because Region IV is described as largely unresponsive to stimulation yet influential in connectivity analyses, it would be helpful to clarify whether the connectivity results were derived independently of the stimulation findings, and if so, how the authors see these two forms of evidence informing one another. This is especially important given the shift in parcellation framework across analyses.

4. Clarify within-individual variability in subjective reports

The authors note that some participants contributed to more than one category of subjective response. It would be useful to elaborate on this more fully, including the extent of within-individual variability across stimulation sites. This could offer additional insight into insular heterogeneity and may help interpret the mapping of response categories across the cortex.

5. Clarify how directionality and asymmetry were established

The manuscript would benefit from additional detail regarding how the directionality of insular connections was estimated from EEG connectivity measures. Similarly, the abstract's statement that the findings link cytoarchitecture to “directed and asymmetric electrophysiological interactions” would benefit from clarification, particularly if lateral asymmetry is being implied.

6. Improve manuscript organization and figure/reference consistency

Some parts of the manuscript could be better organized, with methodological details consolidated more clearly in the Methods and interpretive content reserved for the Discussion. In addition, the figures and references should be checked carefully for consistency, including mismatches in figure legends and inconsistency in citation formatting.

Referee #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #3

(Remarks to the Author)

This manuscript by Quabs and colleagues presents an intracranial electrophysiology study of the human insular cortex, integrating data from 87 SEEG participants within a recently developed 16-area cytoarchitectonic atlas. Three complementary methods are combined: (1) 50 Hz electrical stimulation with collection of reports of subjective feelings (in 50

participants), (2) task-evoked recordings during a Gradual Continuous Performance Task (GradCPT), and (3) cerebrocerebral evoked potentials (CCEPs) mapping intra-insular and insular-cortical connectivity.

This study is a broad characterization of the insula in terms of stimulation-evoked feelings, somatotopy of somatosensory and pain sensations, and causal mapping of the effects of stimulation within the insula and its associated cortical networks. The authors rely on the previous Quabs et al. cytoarchitectonic parcellation and group them into four zones: Region I (dorsal-posterior) encodes somatotopically organized nociceptive/thermoceptive and somatosensory signals; Region II (ventral-posterior) integrates somatic information across multiple body parts without clear somatotopy; Region III (mid-anterior) is associated with visceral sensations and anxiety; and Region IV (anterior-polar) is largely silent to stimulation (does not evoke subjective feelings) but shows robust activation to salient events (which also require cognitive control; see below) in the GradCPT. A central claim is that the functional boundaries of these zones align very well cytoarchitectonic boundaries, though I don't think this claim is directly tested nor does it need to be strictly true for the paper (and the parcellation) to have value.

In addition, CCEP connectivity mapping reveals asymmetrically strong outbound influence from Region IV to the other three regions, with strong bidirectional connections to prefrontal cortex, while Region I connects preferentially to somatosensory cortices. The authors interpret these findings as evidence for a top-down modulatory pathway from anterior to posterior insula with implications for cognitive modulation of pain and interoception.

The paper has many strengths. The functional organization of the insula is of persistent and central importance, and this is an authoritative contribution. The group has outstanding expertise in both insula neuroanatomy and iEEG mapping. There is a tradition of similar iEEG mapping studies of the insula, appropriately cited here, but this one includes a larger sample size than most and an open-ended characterization of evoked sensations. There is strong rigor in including data from reports that are evoked selectively by real (vs sham), reproducible, and dose-dependent stimulation. It is remarkable that 51% of stimulation sites evoked sensations with these criteria. The ability to compare responses across stimulation frequencies — e.g., finding that pain is not simply evoked by more intense stimulation — is a strength. Stimulation-evoked feelings are also combined with CCEP mapping, which to me is the most novel aspect of the study. The finding of asymmetrically strong outbound influence from Region IV to the other three regions, with strong bidirectional connections to prefrontal cortex, is striking. This is very meaningful when coupled with the idea that they are “silent” in the sense that stimulating this area does not evoke feelings, and responsive to the cognitive task. The figures are clear and informative. I have no doubt that the study will be widely cited and used.

I have a number of comments below, but all of these are addressable within the scope of the paper with revisions and are thus intended as suggestions for improvement and clarification.

Introduction and framing

The introduction is grounded in interesting questions, but the framing could be refined. It describes two processing principles as opposing theories, a posterior to anterior “bottom up” and a prefrontal to insular “top down” predictive system. But these do not seem to be opposing to me, as current conceptions hold that both types of processes occur in the insula and other circuits, so the “opposing theories needing reconciliation” framing seems to be a bit of a strawman. The study has a great deal of value, but it does not really provide a “causal examination of these two seemingly opposite views” or adjudicate between top-down and bottom-up influences. What we do need is more specific information about directional connectivity and flow of information, and the study provides this.

The authors state that the “main finding is striking correspondence between microstructural architecture and function”. But the results don't exactly reinforce this. It is a matter of interpretation how closely the boundaries of stimulation-evoked sensations follow the cytoarchitectural boundaries; sometimes they seem to more, sometimes not. In some cases where they clearly do not, the discrepancy is explained away as a failure of anatomical alignment (the mapping between the atlas parcels and these brains is, after all, approximate). This makes the claim of “striking correspondence” circular. However, the value of the paper to me does not hinge on this question, so I'm not sure that correspondence with cytoarchitecture is the “main finding”. If you really want to test correspondence, some more quantitative metrics and analysis would be appropriate. Some quantitative criteria might include: (1) Is the zone evoking a particular sensation contained completely within one parcel? (Or one group of parcels, but this introduces extra flexibility that would require cross-validation or similar to test). (2) Is the distribution of sensations evoked at different locations homogenous within a parcel (e.g., using a spatial poisson model) or not? (3) Is this parcellation a better description of the boundaries of functional zones (stimulation-evoked feelings) than other parcellations (e.g., using mutual information; e.g., the insular parcellation in the brainnetome atlas itself)? I don't see that existing analyses (Supp Fig 5) really address this.

Integration with human MRI/fMRI literature

The integration into the paper of previous iEEG work is understandably more complete than the integration of other work (e.g., fMRI), but there is an opportunity to relate to fMRI findings more closely. For example, it would be interesting to comment on the relationship of the findings with recently published fMRI work (<https://www.nature.com/articles/s41467-026-71568-9>), which appears to identify “domain general” anterior insular zones that may be related to the “silent”, “top-down” anterior insular zones you identify here. This study also includes a mapping between insular zones and other cortical systems that may converge or diverge with your CCEP mapping.

A really wonderful contribution would be to share MNI-space coordinates of stimulation sites and evoked sensations (I note that this seems to be included in the data sharing statement, which is great), and/or maps of smoothed locations of various types of sensations (e.g., in MNI-space nifti image files). This would allow for closer quantitative comparisons with other

studies. I think this is at the authors' discretion, but it would really enhance the value of the work.

Sampling of self-reported feelings

An issue to consider in future studies is how the sampling of feelings is done. My understanding is that patients were asked how they feel in an open-ended way. This has strengths, but it can also lead to biases in the kinds of feelings patients expect, attend to, and prepared to report. If one feeling is present it may mask other feelings that are present but would only be reported if patients are asked specifically about them. That is why emotion researchers usually sample categories systematically (e.g., asking about multiple feelings in response to a stimulation), and medical checklists ask patients to reflect on each of a series of categories one at a time. Those methods also have issues but I wonder about the kinds of biases introduced by the methods here.

In addition, I wonder if the location of stimulation sites was randomized across repeated trials. If not, there is potential for confirmation bias. e.g., with stimulation of site x, the patient says, "I feel tingling in my right leg". The researchers then stimulate the same site, and the patient is primed to report the same sensation, so data are not independent. This may not be an issue, but more clarity on potential caveats/issues would be helpful.

More broadly, a curious feature of the findings is that stimulation evoked feelings quite frequently, but only evoked a very small subset of the possible range of feelings humans experience. e.g., I don't think there are any reports of emotions (no disgust?), no itch (?), and no reports of social cognitive judgments or perceptual distortions, etc. Was this related to the scope of the instructions and the way questions were asked of patients, or do you think this reflects a truth about insular function — i.e., these other feelings are more indirectly related?

Last comment here: Could you provide a more detailed breakdown somewhere about what visceral sensations mean? e.g., any nausea? Rotation/vestibular sense? Stomach churning, etc?

CPT task

As I understand the CPT task, it is discussed as a "salience detection" task, but may conflate novelty and inhibitory demand with "salient" events. The rare (mountain) events require response inhibition in a series of rapid "go" (response) trials. If so, acknowledging the range of psychological processes that could be evoking neural activity would be helpful. An inhibitory account would fit well with other work identifying anterior insula in cognitive control (e.g., the multi-domain paper cited above). Anterior insular activity in inhibition and cognitive control tasks is a very consistent finding in early fMRI literature, though insular locations are sometimes (mis-)labeled as ventrolateral PFC. See, e.g., Sylvester et al. 2003 and Wager et al. 2005 from the Jonides lab, work by Thompson-Shill at Penn, and work by Aron and Poldrack on response inhibition (stop signal task). The "reversed instructions" helps to some degree, but is only in 6 participants, and I don't think this rules out a role for cognitive control (or even response inhibition specifically).

Clustering and justification for reclassification of Area Id4

The four-zone structure relies on k-means clustering using the elbow method for choosing 4 clusters. This is OK, but is not generally considered a rigorous way to choose clusters; bootstrap and permutation tests with, e.g., mean silhouette values provide a more principled and modern metric. This is not a big deal, as the choice of number of clusters is inherently somewhat arbitrary in my opinion. K-means also makes strong assumptions about the shapes of clusters (multivariate Gaussian), and different methods are sensitive to different shapes. It might be good to just acknowledge that the four-cluster solution is a heuristic solution rather than a hard truth.

K-means clustering assigned Id4 to Region II (the somatosensory-dominant cluster), consistent with Id4's own profile, which the authors describe as "predominantly somatosensory rather than pain and temperature" (line 227). The authors then manually reclassified Id4 to Region I (the pain/temperature-dominant cluster) because Id4's responses were localized to a single body part (the mouth), completing Region I's somatotopic gradient.

What the need for this reclassification reveals is that the original feature set may have lacked a functionally meaningful dimension -- somatotopic specificity (i.e., whether responses are localized to a single body part versus distributed across multiple parts). The clustering could not capture this because it only encoded response category frequencies, not the somatotopic precision of responses. What would happen if the clustering were done on a more complete set of features, including somatotopy? If Id4 groups with Region I it would strengthen the justification for the current scheme. If the results differ in other ways, you might note that there are alternate groupings.

The broader underlying issue is how we are supposed to think of the organization. Clustering assumes that each cytoarchitectonic parcel belongs to one and only one group. It's a useful simplification, but why can't the underlying organization instead consist of broad gradients, with the zone show somatotopy being probabilistic, and partially overlapping (but partially diverging from) probabilistic zones showing visceral sensations, etc.

CCEP interpretation

There seems to be a methodological inconsistency in how CCEP data are interpreted. The F1 metric captures the early, N1-like component indicative of feedforward connectivity, and the claim of Region IV's dominant outflow rests on this metric. However, the finding that Regions I and II show stronger influence over Region III in the "late low-frequency range" (lines 399–402; interpreted as feedback) derives from spectrogram inspection without a formally defined metric, specified time windows, or statistical tests. The authors should either define a complementary metric for this late component with equivalent rigor or acknowledge the feedback finding as exploratory.

One interpretation of the findings that I do really like is that insular Zone IV provides a bridge between prefrontal (lateral and ventromedial) systems involved in meaning-making (see Koban et al. 2021 Nat Rev Neurosci) and cognitive maps (e.g., Behrens 2018; Constantinescu and Behrens 2016) and somatic systems including pain, i.e., PFC -> Zone IV -> Zone I -> sensorimotor cortex. Several labs have demonstrated pathways in rodents from anterior cingulate to Zone I regions (dorsal posterior insula) and Zone I regions to spinal cord to control nociception (Tan and Kuner 2017 Nat Neurosci is a good example).

Requests for clarification/etc

- Were these patients included in previous publications, or is this a new cohort? How much of dataset is previously published? If patients have appeared in previous publications it is important to note because the data here are not fully independent of those.
- Some places say N = 50 with stimulation, some N = 47, some N = 56. It seems to depend on the type of stimulation/locations. Clarify briefly?
- The figures are very nice, but it would be nice to also see the mapping of your insular diagrams onto MNI-space MRI images somewhere. It's otherwise hard to infer how the stim sites match, except by way of the atlas regions.
- Are there sites that evoke multiple sensations? Yes, there are, but the data are presented in Fig 1 as though there is always only 1 sensation evoked, and later multiple sensations are discussed. Clarify that the multiple sensations are reports of multiple sensations in response to a single stimulus (and not an aggregation over different participants/stimulations?)
- More granularity on the types of sensations would be helpful if possible. What is the relationship between the maps for pain and somatosensation? There is some discussion, but it's purely qualitative — "some sites" responded to warmth and pain, "other sites" somatosensations without pain. What is the quantitative joint distribution? Are any of the visceral sensations painful? "Pain and temperature" is one category, but can you distinguish between painful and non painful temperature sensations? "Somatosensory" includes pressure...was the pressure ever painful? Were any positive emotional responses ever elicited? Or feelings of sadness or anergia/apathy/fatigue?
- "We reassigned area Id4 from Region II to Region I". OK, but it's confusing to have Fig 1 with one scheme and then Fig 2 with a different scheme. I think this will confuse readers later.
- "We included parcels represented in at least 3 participants for each insular region" Can you show a map of which parcels are covered and which we simply do not have information about? This is important to interpret the results.
- Can you more clearly label motor (M1/area 4), sensory (S1), and premotor regions on the brainnetome parcels in the CCEP analysis?

Typos and other minor comments

you might tone down the claim that the study identifies "an electrophysiological pathway for cognitive modulation of pain and bodily perception". The study includes no condition testing pain modulation (e.g., no attentional or cognitive manipulation during noxious stimulation) or otherwise demonstrates this directly. It like this hypothesis, but it would be helpful to identify the claim as a hypothesis rather than a demonstrated finding.

microscopical

the finding of anxiety responses from the stimulation of left (but not right) anterior insula...and then "the lateralization of anxiety responses to the right hemisphere needs...."

The anxiety group showed higher mean Beck Anxiety Inventory scores -> this makes it sound like there is a group of patients with anxiety, but I think "group" refers to electrodes?

The drop from 87 to 50 participants is unexplained. Mentioned that 87 participants underwent SEEG, but only 50 received electrical stimulation mapping. No reason is given for why the other 37 were excluded.

Line 189: MDS is not a validation of the cluster solution. The elbow method alone determined $k = 4$. The MDS stress value (0.01) indicates the 3D projection faithfully preserves pairwise distances, not that four clusters is the correct number. Citing MDS as "validation" of the quadripartite scheme is incorrect. Proper validation would be e.g., silhouette analysis or permutation tests for cluster stability.

Lines 238–240: The "sharp transition" between Regions I and II is illustrated with one participant (four examples from Supplementary Figure 2 come from a single participant's brain). The paper does not report whether this sharp transition was systematically observed across all participants with electrodes traversing the Region I/II border. Group-level statistics showing different response proportions between regions are not equivalent to demonstrating sharp borders within individuals.

Line 449: an incomplete sentence? "The most anterior insula emerges as a hub, lacking direct access to subjective percepts, but exerting strong outbound influence across other insular networks and bilateral electrophysiological relationship with prefrontal, including the anterior cingulate, networks."

Lines 506–509: The "as-if body loop" link is a large inferential gap. Connecting their findings to Damasio's concept requires demonstrating that Region IV's interactions with prefrontal cortex enable generation of bodily state representations without peripheral input. The CCEP data show that Region IV <-> prefrontal connectivity exists, but CCEPs reveal nothing about the content of information being transmitted. The leap from "Region IV is connected to prefrontal cortex" to "Region IV supports as-if body loops" seems to be speculative.

Ideas for extensions (later)

Is it possible to characterize sensations evoked by sham stimulation (placebo)? Sensations be real but context-dependent and therefore not reproducible.

Referee #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #5

(Remarks to the Author)

In this paper, the authors use intracranial recordings from human epilepsy patients and also intracranial stimulation to map different functions of the insula during pain and interoceptive experience. Even though the intracranial dataset is a rich dataset, there are several issues such as mathematical and statistical errors, absence of any novelty, and failure to give proper credit to prior researchers, which precludes its publication in Nature or Nature-branded journals. I outline them below.

First of all, the title is misleading:

“Functional Architecture of the Human Insula Revealed by Causal Intracranial Mapping”

The authors are mostly mapping during lower order cognition such as sensory processing, not higher level cognition. The title should be modified to reflect accordingly.

In the Introduction, the authors mention:

“Previous studies, while offering important insights, have been limited in their approach: Most prior studies have examined the insula through isolated anatomical, functional, or connectivity lenses. Here, we unify these dimensions using a high-resolution and multimodal approach that enables an unprecedented characterization of its functional architecture using electrophysiological causal measures of its function and connectivity with not only the broader cortical mantle but also among its different subregions.”

These comments are simply factually wrong. The authors have omitted references to several other pioneers and their contributions to the comprehensive mapping of the anatomical, functional, or connectivity lenses of the human insula, including the different subregions of the insula.

I think it's highly unethical to omit a plethora of prior references which have comprehensively mapped the functions of the insula and then claim the already known findings of the insula as novel findings.

There are other factually incorrect or misleading comments throughout the manuscript, where the authors claim to have made a contribution, however, the data presented in the manuscript is not enough to support these claims. For example, the following sentence is again simply not true:

“In a cohort of 87 human participants implanted with intracranial electrodes, our current work aims to understand the fundamental organizing functional principles of insular cortex in the human brain in light of the new evidence about its microscopical anatomical architecture.”

To unravel the fundamental organizing functional principles of insular cortex, insula mapping using higher order cognitive tasks such as attention, memory, etc, are necessary, which is not what the authors are showing in their manuscript.

The paper is also riddled with major mathematical errors and impossible statistics. Specific comments are below.

Significant omission of prior literature:

(i) The authors talk about salience and higher order cognition, but the entire human insula intracranial EEG literature is omitted, see for example, work coming from the Nitin Tandon group or the Robert Knight group.

(ii) Mapping subjective states in the insula using brain stimulation is not new and known for decades. See for example, Ostrowsky et al., Cerebral Cortex, also including authors' own previously published work, for example, Duong et al., Brain Stimulation.

(iii) Entire animal literature pertaining to the insula is also missing. For example, see Caruana et al., Current Biology.

(iv) The authors talk about insula stimulation and connectivity, however, the work coming from the lab of Philippe Ryvlin is missing:

(v) The entire cytoarchitectonic literature of the human insula using fMRI/sMRI/dMRI is also missing.

Major mathematical errors:

(i) The use of common average referencing rather than bipolar referencing is wrong, see work coming from Robert Knight lab.

(ii) "The remaining electrodes were balanced across the hemispheres (234 left, 244 right) and anterior-posterior (237 anterior, 241 posterior) subdivisions."

This is almost impossible. It is a near impossible task to get the same number of channels in both the hemispheres, since epilepsy predominantly affects the left hemisphere. Similarly, anterior insula will have more number of channels compared to posterior insula.

I have never seen a single insula paper published in literature which has the same number of electrodes in both the left and right hemisphere. Similarly, I have never seen a single insula paper published in literature which has the same number of electrodes in both the anterior and posterior insula.

Honestly speaking, it's really shocking that the right hemisphere has more electrodes (244) compared to the left hemisphere (234). It's similarly very very concerning that the posterior insula has more electrodes (241) compared to the anterior insula (237).

"Collectively, these distributions demonstrate balanced sampling across both hemispheres and anatomical subdivisions of the insular cortex."

This is hilarious because this never happens.

This also raises serious ethical concerns regarding whether the electrodes were actually implanted based on clinical, rather than, research purposes only. The authors mention that the electrode implantations were driven by clinical rather than research needs, however, based on the statistics that they show, it's now unclear to me whether proper ethics were followed related to electrode implantations.

(iii) "In the stimulation experiments, we only included participants' subjective reports if they passed our strict quality control metrics: Reports were (1) evoked only during real, but not sham, stimulations, (2) reproducible when the same site was repeatedly stimulated, and (3) dose-dependent (i.e., effect intensity increased as stimulation intensity increased)."

I am sorry that these are not quality control measures, rather these are cherry-picking of massive proportion. The reason is that the human brain is highly heterogeneous and heterogeneous responses are expected when we apply identical stimuli to the brain, a very well-known phenomenon known for decades now. In the above so called "quality control" metrics, the authors are deliberately picking the things that are aligned with their hypothesis and throwing away the data that are not aligned with their hypothesis. Using this logic, literally any hypothesis can be proven to be correct by just choosing the trials/electrodes that are consistent with someone's hypothesis.

"(1) evoked only during real, but not sham, stimulations" – Reports need to be included regardless of whether there was a evoked response during real or sham stimulations. "(2) reproducible when the same site was repeatedly stimulated" – Reports need to be included regardless of whether the same response was reproducible or not. "(3) dose-dependent (i.e., effect intensity increased as stimulation intensity increased)" – Reports need to be included regardless of the effect intensity value as the stimulation intensity increased.

(iv) "Subjective responses were evoked by electrical stimulation in 51% of insular sites, with a higher elicitation rate in the posterior (67%) compared to the anterior insula (32%)"

This is very surprising since I would expect higher elicitation rate in the anterior compared to the posterior insula because of the relevance of the anterior insula in many higher order brain functions.

Similar issues remain from lines 214-273, where the authors have simply dumped a lot of statistics and response profiles, however, there is no effort to reconcile these findings to an integrated theory of the insula function. Again, similar issues from lines 274-307.

Again similar critiques apply to lines 309-344. There is a very solid literature when it comes to the salience detection function of the anterior insula. In addition to omitting a very large number of important references, the authors are again just showing some statistics without a clear theory or hypothesis. Just reporting the percentage of significant electrodes statistics is not very useful.

Similar critiques apply to lines 346 to 402. It's really impossible to make anything out of what the authors are presenting in the context of insular functions.

(v) "Stimulation consisted of alternating square-wave pulses at 50 Hz frequency with 0.2 ms pulse width, delivered for 1–4 seconds per trial."

50 Hz stimulation frequency is typically applied when studying higher order brain functions such as attention and memory, since gamma frequency is more relevant for attention and memory, see work from Nanthia Suthana's lab. It's unclear what the rationale is to apply 50 Hz gamma frequency stimulation in the insula which is not an attention/memory brain region, and it's also unclear what the relevance of 50 Hz brain stimulation is when the authors are primarily studying lower order processes such as sensory processing.

(vi) "Electrodes were labeled as significant when the significant bin count exceeded 2."

This selection looks very arbitrary.

(vii) "Then, common average re-referencing was applied, with channels excluded from the reference if they: (a) showed variance exceeding five times or falling below one-fifth of the cross-channel median variance; (b) contained spikes with a rate exceeding 1 spike per second, where spikes were defined as inter-sample voltage changes $\geq 80 \mu\text{V}$; (c) exhibited pathological activity with pathological high-frequency oscillations (Liu & Parvizi, 2019); or (d) displayed abnormal power spectral characteristics upon visual inspection."

This is again cherry-picking on a big scale. It's unclear how many electrodes/trials were excluded using these criteria.

(viii) "We conducted independent-samples t-tests to compare the mean HFA power amplitude between target and baseline trials within a specific temporal window of 400–1200 ms post-stimulus (relative to the onset of the image fade-in)."

Again, the choice of this window selection seems arbitrary.

(ix) "Electrodes were classified as target-responsive if the difference in mean HFA power between these two conditions reached a two-tailed significance threshold of $p < 0.05$, after False Discovery Rate (FDR) correction was applied to account for the total number of electrodes within each subject."

Was this corrected for across subjects too since there are a relatively large number of subjects?

(x) "Preprocessing included notch filtering at 60, 120, and 180 Hz, followed by exclusion of noisy channels and trials. Noisy channels were identified based on extreme raw amplitude (>5 SD), frequent spikes ($>3\times$ median spike rate), high mean absolute voltage (>4 SD), or elevated trial variance (>3.5 SD). Trials with mean amplitudes exceeding 4 standard deviations were also discarded."

This is again cherry-picking. How many channels and trials were excluded like this?

(xi) "To examine directional asymmetries across the full time-frequency spectrum beyond the early high-frequency F1 response, we calculated contrast spectrograms between paired electrodes for each regional comparison. For matched electrode pairs, power spectrograms were computed in each direction, averaged across pairs and trials, and subtracted to yield directional contrast maps (forward minus backward). Cluster-based permutation testing identified statistically significant spatiotemporal clusters. Candidate clusters were defined as contiguous time-frequency regions where point-wise t-tests across electrode pairs showed significant directional differences ($\alpha = 0.05$). The sum of t-values within each cluster was computed. Significance was assessed via 5,000 permutations where direction labels were randomly shuffled, contrast spectrograms recalculated, and maximum/minimum cluster t-sums recorded to form a null distribution. Observed clusters exceeding 95% of the null distribution were deemed significant."

This is mathematically wrong (Figure 5). To assess directional connectivity, please use a directional connectivity measure such as the Granger causality, see the Anil Seth Granger causality toolbox.

Other major issues:

(i) "These findings reveal a four-fold structural-functional organization of the human insula based on electrically evoked response profiles of microstructural areas."

It's unclear what the relevance of four-fold here is. There is no theory to support only a four-fold structural-functional organization of the human insula.

(ii) "Stimulation of the posterior and dorsal insula reliably elicited pain and thermosensory percepts – as well as somatosensory changes as detailed above. However, we observed an intriguing meso-scale functional division among anatomically adjacent neuronal populations with distinct yet related roles: while stimulation of some sites evoked pain alone, stimulation of other sites within the same region produced purely thermosensory sensations, stimulation of a third set of sites

elicited both warmth and pain, and stimulation of yet another set of sites elicited somatosensory sensations in the absence of pain or temperature percepts. It should be noted that the average electrical current eliciting pain and temperature was comparable with all other responses (pain and temperature: median 3 mA, IQR 2-4mA/ all other responses: median 4 mA, IQR 2-6) (Supplementary Figure 1). Stimulation at no site elicited a sensation of cold."

The authors are reporting a lot of results here. However, it's hard to reconcile these results in a coherent theory. Obviously, when we apply intracranial stimulation, there will always be some channels which will respond more compared to some other channels, however, without a tight link to the anatomy or theory of the insula it's unclear how these findings are useful. These results are highly variable, so what do all these results mean in the context of the insula theory?

(iii) There is also no control region where brain stimulation is applied. So it's unclear whether the effects are also present brain-wide or specifically localized to the insula. If stimulating other brain areas also evokes similar response profiles to the insula, then the authors cannot make causal claims specifically related to the insula.

(iv) "DATA AVAILABILITY

We have mapped the four insular regions in MNI152 ICBM 2009c Nonlinear Asymmetric space and will make it publicly available for download upon publication of our work. Electrophysiological connectivity data from our study along with the exact coordinates of the stimulated or recorded sites in MNI space will also be shared upon publication."

It is suggested to make the raw unprocessed data available for ethical reasons and reproducibility. Holding on to data and without making it available for other researchers hinders the progress of science and reproducibility. These days, data can be easily anonymized and made publicly available, see publicly available datasets from the Eddy Chang's group at UCSF or Rob Knight's group at UC Berkeley.

(v) Lyu, D. et al. Mapping human thalamocortical connectivity with electrical stimulation and recording. Nature neuroscience 28 (2025). <https://doi.org/10.1038/s41593-025-02009-x>

The above reference is cited twice, once as reference 44 and then again as reference 72. I do not understand this at all.

Concerns on figures:

(i) The insula in Figure 1A right panel and Figure 1B, and in all the subsequent figures looks very very weird and strange. I have never seen an insula structure like that. Something is going very wrong in the registration algorithms.

(ii) It's unclear why there are a large number of non-responsive electrodes in the anterior insula subdivisions during the salience task in Figure 4.

(iii) Figure 5 is mathematically wrong like I mentioned above.

Overall, the manuscript is very poorly written and poorly organized, even though the dataset itself is a very rich dataset. With so many issues, mathematical and statistical errors, absence of any novelty, and failure to give proper credit to prior researchers, I am not sure whether the paper is even publishable as presented. This also precludes its publication in Nature or Nature-branded journals.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)

Summary

I was generally convinced by the authors' response and the revised manuscript. The authors have taken the prior comments seriously and, in my view, have substantially improved the paper accordingly. The revision is stronger in its framing, clearer in several of its methodological explanations, and more appropriately calibrated in its interpretation of the findings.

At this stage, I only have a small number of remaining comments, all of which are relatively minor and primarily concern conceptual precision, clarification of cohort structure, statistical wording, and figure labeling.

Minor comments

1. The historical framing of interoception in the Introduction should be corrected

The Introduction currently attributes to Sherrington a definition of interoception as "the sensing of internal bodily states, distinct from exteroception and proprioception." This is not quite accurate historically. Sherrington's contribution was to

distinguish classes of physiological stimulation according to whether they arose from outside the body, within the body, or from the musculoskeletal apparatus; that is, he classified stimuli as exteroceptive, interoceptive, and proprioceptive. The more explicit formulation of interoception as the sensing of internal bodily states is a more contemporary conceptualization rather than Sherrington's original definition.

I would recommend revising this sentence for historical accuracy and adding one or more appropriate contemporary references to support the modern definition being used in the manuscript.

2. The description of cohort structure is improved, but one point remains unclear

The revised manuscript is clearer than the prior version in detailing the composition of the different data subsets and the degree of overlap between analyses. That said, I am still somewhat confused about the distribution of the Stanford and Italian cohorts in relation to the Somatosensory Evoked Potentials analysis. It would help to make this point fully explicit, ideally in one concise sentence or a simple schematic/table indicating which cohort contributed to which analysis.

This is now a relatively minor clarity issue, but resolving it would make the workflow easier to follow.

3. The wording of the statistical approach for the Saliency Task should be clarified

Regarding the Saliency Task, the manuscript refers to independent-samples t-tests conducted within each participant. This surprised me, since measurements obtained within the same participant are ordinarily treated as dependent rather than independent. I may be misunderstanding what the authors intended here, but the current wording is potentially confusing.

It would therefore be helpful for the authors to clarify exactly what comparison was performed and why this statistical formulation is appropriate. If the analysis was in fact based on within-subject comparisons, the terminology should be revised accordingly.

4. At least some figure elements may still require checking

Some figures may still need another careful review for labeling clarity. For example, in Panel A of Figure 1, the "responsive" group does not seem to correspond transparently to a visible portion of the bar plots. In particular, the meaning of the grey segment is not immediately clear. The authors should clarify this in the figure and/or legend.

Referee #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #3

(Remarks to the Author)

I have reviewed the responses to all of the prior reviews and revisions to the manuscript. I believe this is an important paper that substantially advances the field, and I support publication in Nature. The authors have done a thoughtful and thorough job in responding to all of the comments raised.

The combination of techniques and findings presented here, which constitute over a decade of work collecting data from these rare patients, provide something qualitatively new that builds on previous work, but also adds new insight. The CCEP findings are a good example of novel and important new findings, as they show directional connectivity between the frontal cortex (and other cortical areas) and insula, and within the insula, that tells us something new about how 'top-down' influences and predictive processing works. In addition, the paper provides fascinating, foundational data on the types of feelings, evoked by insular stimulation, which are surprisingly common but also surprisingly limited to particular classes of interoceptive, somatosensory, and pain-related sensations. The combined findings on anterior insular area 4 — that it is devoid of evoked feelings, projects back to influence posterior areas 1 and 2, and receives substantial frontal projections — provides substantive new data that challenges and updates now-classic views of anterior insular function and will lead to new conceptual models. These data will also be shared, which will be a boon for the field. In short, the authors have satisfied the concerns raised.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reply to Co-Referees 1 and 2

Summary: This manuscript presents an ambitious and methodologically sophisticated investigation of the functional organization of the human insula using a rare combination of intracranial electrical stimulation, subjective phenomenology, clustering-based parcellation, and directed connectivity analyses in a neurosurgical population. The study has several notable strengths. Most importantly, it leverages a sizeable neurosurgical sample with unusually broad bilateral coverage of the insular cortex, which is a substantial asset for invasive human mapping work of this kind. In addition, the manuscript moves beyond descriptive sensory mapping by examining both “inflow” and “outflow” connectivity across insular subregions, providing an especially valuable systems-level perspective on insular organization. The combination of stimulation-evoked percepts, topographic analyses, clustering, and connectivity makes this a potentially important contribution to the literature.

Overall, I found the manuscript interesting, substantial, and potentially impactful. The data are rich and the study has several standout features. My main concern is that some of the interpretive claims, particularly those concerning the interoceptive role of the insula and the functional significance of anterior insula findings, are stronger than the present data can fully support. I think the manuscript would be strengthened by a more restrained and precise framing of what the findings do and do not establish.

We thank both referees for their constructive engagement with our work and greatly appreciate the considerable time they devoted to reviewing the manuscript. We are encouraged that they found the work rigorous, and we have taken their concerns regarding interpretive overreach seriously. In response, we have revised and moderated our interpretations at several points throughout the manuscript as detailed below.

Major comments

1. The manuscript's central claims about interoception are currently stronger than the data appear to support. The most important issue concerns the interpretation that the findings establish specific insular subregions, particularly dorsal posterior Region I, as primary interoceptive cortex or as a primary interoceptive hub. The data clearly support an important role for the insula in bodily experience, and particularly for posterior insula involvement in thermal, nociceptive, and body-localized percepts. However, the evidence for a broader and more canonical interoceptive role is less definitive than the manuscript suggests.

In particular, relatively few internal organ-system-specific sensations appear to have been elicited. Classic visceral domains such as cardiac, respiratory, gastrointestinal, and genitourinary sensations seem limited or underrepresented. This is important because the Discussion at times moves from bodily sensation broadly to interoception more specifically, without fully resolving that distinction. A more careful differentiation between somatosensory, nociceptive, vestibular/body-state, and organ-specific interoceptive percepts would improve the interpretation.

In full agreement with the Reviewers, we have revised the manuscript throughout and now refer to Region I as “a region of the insula with core involvement in somatotopically localized bodily experiences, including thermal, nociceptive, and tactile sensations.” Likewise, we now refer to Region II as “a region of the insula with core involvement in bodily experiences encompassing

broader regions of the body in the thermal, nociceptive, and tactile domains.” We have also modified the wording of the Introduction to focus on the insular involvement in body related signals, including visceral sensations, pain, and temperature.

In addition, we acknowledge that the subjective reports elicited by posterior insular stimulation primarily involved pain, temperature, and touch-related experiences, rather than “visceral interoception” per se. By contrast, as noted in the original submission and clarified further in the revised manuscript, the mid-anterior insula (Cluster III) appears to be the insular “region with core involvement in bodily experiences related to internal bodily states, including visceral sensations”.

We have also revised the Discussion to acknowledge that the role of the insula in interoception may be modulated at the multi-regional level, with posterior insular cortices contributing predominantly to bodily experiences related to thermal, nociceptive, and tactile processing, whereas the mid-anterior insula appears to be more closely engaged in internal bodily and visceral functions.

Relatedly, it would be important to clarify whether participants were specifically queried for internal sensations such as heartbeat, breathing, or visceral changes. Why were cardiac or respiratory sensations apparently so sparse? Is it possible such sensations occurred but were not explicitly queried, captured, or categorized? As currently written, the claim that these findings establish the insula’s interoceptive role is somewhat overstated relative to the range of reported phenomenology, particularly in light of the functional neuroimaging literature that has consistently demonstrated insula activity in response to cardiac interoceptive awareness during resting conditions (Schulz, 2016, PhilTransRSocB) and under conditions of cardiac perturbation (Hassanpour et al., 2018, Neuropsychopharm).

We have now added two new paragraphs in Results highlighting the visceral changes reported following stimulation of the mid-anterior insula and addressing the referees’ comments:

New Text: Electrical stimulation of the mid-anterior insular Region III elicited a rich and diverse range of visceral experiences across participants. The most frequently reported sensations involved inside the chest and abdomen, at times associated with a feeling of nausea or near-emesis (e.g., “almost had a dry heave,” “queasy, like the beginning phase that you’re going to throw up, but you’re not going to throw up”), pressure or fullness in the chest and rising or spreading visceral sensations described as energy or waves moving through the trunk. Several participants reported disturbances including sensation of a wave inside their body as if “being on a boat”, or a sense of the whole body moving or vibrating. Respiratory and pharyngeal percepts included sensations in the throat, sinuses, and chest cavity, as well as changes in vocal projection and a feeling of breathlessness. A subset of participants reported craniofacial and cephalic sensations, including wave-like or ripple sensations spreading across the scalp.

...

It is important to note that our observation of changes induced by insular stimulations was primarily based on spontaneous subjective reports rather than targeted questions or quantified measures of heartbeat, breathing, or visceral functions. It is therefore possible that more instances of cardiac or respiratory sensations could have been captured if they were explicitly queried or measured quantitatively.

For similar reasons, the statement that “Together, these converging lines of evidence support the interpretation of dorsal posterior Region I as a primary interoceptive hub” should be softened. The study provides strong evidence that this region is important for evoked bodily

percepts and is tightly embedded in local insular circuitry. However, the current findings do not as clearly establish it as a primary interoceptive cortex in the broader sense asserted by the manuscript.

In agreement with the Referees, as we noted earlier, we have now refrained from referring to Region I as the “primary interoceptive hub” describe it as “a cortical region tightly embedded in local insular circuitry with core involvement in somatotopically localized bodily experiences, including thermal, nociceptive, and tactile sensations.”

2. The anterior insula interpretation is interesting but remains somewhat speculative in its current form. One of the most intriguing aspects of the manuscript is the relative lack of subjective phenomenology evoked by stimulation of agranular anterior insula, contrasted with evidence that an anterior insular region exerts strong influence over other insular regions and is preferentially connected with prefrontal cortex. The inflow/outflow connectivity mapping is an especially strong feature of the study and adds meaningful depth beyond traditional stimulation mapping.

That said, I do not think the present evidence is yet sufficient to conclude that the anterior insula therefore functions as an “integrative hub” in the strong sense implied. The current data do not directly establish what information is being integrated, nor do they directly demonstrate integration as a computation. In this respect, the interpretation feels somewhat ahead of the evidence.

In agreement with the Referees, we have now removed any mention of “hub” or “integrate” unless we refer to other studies suggesting such functions. We use neutral language such as “involved in” when we refer to our findings. Whenever our statement crosses from evidence to a postulate, we highlight it as such to clarify that future tests are needed to validate our hypothetical claim.

I also think there is a plausible alternative explanation for the lack of anterior insula phenomenology during stimulation. To what extent might the context of the experimental stimulation matter? The study does not appear to manipulate interoceptive expectancy, anticipation, uncertainty, or contextual forecasting of bodily signals (see Harrison et al., 2021, *Neuron*; Lovero et al., 2009 *NeuroImage*; Teckentrup et al., 2019 *NeuroImage*). If anterior insula contributes importantly to interoceptive prediction or to the contextual shaping of bodily awareness, then it may not be surprising that it failed to produce robust subjective responses under passive stimulation conditions. Put another way, anterior insula may require an interoceptive expectation or predictive context to prime responsivity and perceptual expression. This possibility seems particularly relevant given that the anterior insula reportedly showed robust responses during the saliency task, which explicitly involved a context likely to engage expectancy-related processes.

I therefore encourage the authors to discuss more explicitly whether the absence of stimulation-evoked anterior insula phenomenology should be interpreted as evidence against interoceptive function, or whether it may instead reflect the kinds of interoceptive computations that are most readily revealed by the experimental context used here.

We fully agree with the Reviewers and consider their comment highly important. Indeed, their formulation captures the point more clearly than our own wording. Therefore, we have taken the liberty of incorporating their comments literally in the following paragraph that is now added into the Discussion:

We are mindful that the absence of subjective responses during passive stimulation of Region IV does not preclude the possibility that stimulation of this region could induce behavioral or conscious perceptual changes in different experimental contexts incorporating expectancy, anticipation, uncertainty, or predictive processing of bodily signals. A plausible explanation for the lack of subjective changes with insular Region IV stimulation might be the experimental context in which stimulations occurred. In our procedure we did not manipulate interoceptive expectancy, anticipation, uncertainty, or contextual prediction of bodily signals that are known to be highly associated with the engagement of this region of the insula [e.g., Harrison et al., 2021, Lovero et al., 2009, Teckentrup et al., 2019].

3. The manuscript would benefit from clearer engagement with conscious access and reportability. The present study relies heavily on stimulation-evoked subjective reports and therefore, by design, focuses on phenomena that reach conscious access. This is not a weakness in itself, but it does place constraints on what can be inferred about insular function more generally. Many bodily signals processed within the insula likely remain outside conscious awareness. It would therefore be useful for the authors to clarify whether they view the present findings as reflecting the general functional organization of the insula, or more specifically the subset of insular processes that become consciously reportable.

In this respect, it may help to engage more directly with existing frameworks in the neuroscience of consciousness. According to Global Neuronal Workspace theory, conscious access is typically defined as the transition of information from nonconscious processing into a reportable state associated with large-scale propagation across distributed cortical networks (Dehaene et al., 2011, *Neuron*; Sergent et al., 2005, *Nature Neuroscience*). The present results are highly relevant to reportable bodily experience, but the Discussion should be careful not to equate this too readily with insular function as a whole.

We thank the reviewers for highlighting this very important issue. We have now added a closing paragraph to remind the reader about this important issue. The new paragraph reads as follows:

*In closing, we emphasize that stimulation-evoked subjective experiences likely capture only a subset of the functions of the stimulated brain region. As detailed elsewhere (Siddiqi, Kording, Parvizi, & Fox, 2022), while electrical stimulation provides valuable causal evidence linking a region to conscious perceptual or behavioral phenomena, the likelihood of eliciting such experiences depend on electrical discharges delivered during stimulation engaging neural processes underlying access consciousness (Block, 1990, *Behavioral Brain Science*; Dehaene et al., 2011, *Neuron*; Sergent et al., 2005, *Nature Neuroscience*). Thus, our findings should be interpreted as reflecting only a subset of insular processes that become consciously accessible and reportable, rather than those remaining out of reach of consciousness.*

4. The link to “as-if body loops” appears too speculative The manuscript’s effort to bridge the present findings to the concept of “as-if body loops” seems premature. The current experiments do not explicitly assess internal simulation, contextual interoceptive forecasting, or the bypassing of the body proper in the sense implied by that framework. To make such a connection convincingly, one would ideally need an experimental design that directly probes expectation, anticipation, simulation, or prediction of bodily states, followed by assessment of how those anticipated states are instantiated neurally independent of incoming peripheral changes. As written, the bridge to “as-if body loops” reads as speculative rather than empirically established. I would therefore recommend either substantially softening this claim or removing it. This is especially relevant in light of Damasio’s more explicit formulation of the

concept in *The Feeling of What Happens*, where the “as if body loop” refers to body-related changes being created directly in sensory body maps under the control of other neural sites, thereby partially or entirely bypassing the body proper (e.g., Damasio, 1999, p. 281). The present study is valuable, but it does not directly test that type of internal simulation mechanism.

We followed the reviewers’ suggestion and removed the paragraph about Damasio’s as-if-body loop.

5. The structure of the datasets and statistical handling need greater clarity.

The manuscript would benefit from clearer explanation of how the various participant subsets relate to one another across stimulation, clustering, connectivity, and task-based analyses. At present, it is difficult to determine whether these cohorts overlap fully, partially, or not at all. Clarifying this would strengthen the methodological transparency and improve interpretability.

Relatedly, the statistical treatment of the data requires additional clarification. The analyses appear to combine multiple electrodes within individuals, which are non-independent observations, as well as comparisons across participants, which are independent observations. It is not clear how this hierarchical structure was handled. This is important because the treatment of within-subject dependence and between-subject variability bears directly on the robustness and generalizability of the findings.

We thank the reviewers for their important methodological observation. In the revised manuscript, we have addressed both issues of i) sample/cohort transparency (which participants contributed to which analyses, and how those subsets relate to each other) as well as ii) the statistical handling of dependent observations within individual.

As a result, we incorporated new statistical analyses and revised the Methods and Result sections. Importantly, these revised analyses, which are more rigorous than the original ones, yielded results fully consistent with those originally reported. This convergence strengthens confidence in the robustness of the findings.

Issue I: We have added a paragraph to the Methods section detailing the composition of each cohort and the degree of overlap between them. Additionally, we provide a new Supplementary Table presenting a complete overlap matrix across all four cohorts, allowing the reader to directly assess which participants contributed to which experimental components. We note that intracranial recordings and stimulations targeting the insular cortex remain considerably less common than those in medial temporal lobe structures. The full dataset was accordingly accumulated over 15 years of our clinical practice. Because all studies were conducted in a clinical setting, participation in each experimental component depended on factors including electrode placement, patient availability during the hospitalization window, and task suitability and patient consent, all of which resulted in different number of participants across analyses.

Issue II: We would like to explain how we revised our analyses to account for the hierarchical structure of our data, i.e., multiple electrodes nested within subjects:

Subjective responses and regional comparisons: For all analyses of stimulation-evoked responses, we had originally used binomial analysis. In the revised manuscript, we now use Generalized Estimating Equations (GEE) with an exchangeable working correlation structure

and subject identity as the clustering variable. This approach explicitly models the non-independence of electrodes recorded within the same participant while still allowing inference at the population level. Specifically, GEE was applied to: (i) the comparison of overall response rates between anterior and posterior insula; (ii) lateralization analyses in bilaterally implanted participants; and (iii) pairwise regional comparisons of response profiles across all four insular clusters. The use of GEE with a binomial family for binary outcomes (response/no response) and exchangeable correlation structure is a well-established approach for clustered binary data and directly addresses the within-subject dependence as the reviewers have raised.

Salience Task (regional t-value comparisons): For the salience task analyses, electrode-level t-values were derived from independent-samples t-tests conducted within each participant, comparing high-frequency activity power between target and baseline trials, thereby normalizing each electrode's response relative to that subject's own baseline and noise estimate. This preprocessing step absorbs the dominant source of between-subject variability, rendering the resulting t-values substantially more comparable across subjects and electrodes. Regional differences were then assessed across these normalized t-values using Kruskal-Wallis H tests with post-hoc permutation tests and additionally validated via subject-level paired permutation tests aggregating electrode t-values per participant within each cluster. The convergence of both electrode-level and subject-level permutation results confirms that the reported regional differences are robust to residual within-subject electrode clustering.

Connectivity Strength after CCEP: For the directional asymmetry analysis of cerebro-cerebral evoked potential (CCEP) connectivity, we used permutation testing with 10,000 iterations. Critically, to preserve the within-subject dependence structure, sign-flipping was applied at the subject level rather than the electrode-pair level — meaning all pairs from the same participant were flipped together as a unit on each permutation. This approach respects the hierarchical structure of the data and ensures that the null distribution reflects between-subject rather than between-electrode variability, providing valid inference at the population level. We note that the connectivity metric itself was already normalized at the subject level prior to this analysis, meaning that subject-level shuffling constitutes a conservative double-correction for within-subject dependence. We acknowledge this and consider it a strength rather than a limitation: any significant effects reported are therefore robust to this conservative treatment, and non-significant results should be interpreted with this additional conservatism in mind.

We have clarified these points and revised the statistical data in the revised manuscript to make the statistical treatment of the hierarchical data structure fully transparent. This is reflected in the updated Results and Methods sections, as well as in the modified Supplementary Figure 5.

New Text:

Results

Subjective responses were evoked by electrical stimulation in 51% of insular sites, with a higher elicitation rate in the posterior (67%) compared to the anterior insula (32%) (generalized estimating equations [GEE] accounting for within-subject clustering, Wald $\chi^2 = 8.78$, $df = 1$, OR = 2.74, 95% CI [1.41-5.35], $p = 0.003$) (Figure 1).

We explored the lateralization of the induced effects by analyzing data from only participants who had bilateral insular implementations. Our data suggested that the effects from anterior (13 participants, 30.6% left, 31% right, generalized estimating equations, Wald $\chi^2 = 0.14$, $df = 1$, OR = 1.23, 95% CI [0.42-3.59], $p = 0.706$) or posterior (11 participants, 74.1% left, 87.4% right, generalized estimating equations, Wald $\chi^2 = 2.89$, $df = 1$, OR = 3.07, 95% CI [0.84-11.17], $p = 0.089$) sites were not significantly different.

Methods:

Functional mapping:

To compare stimulation-evoked response rates between anterior and posterior insula, and between left and right hemispheres — we used Generalized Estimating Equations (GEE) with a binomial family and exchangeable working correlation structure, with subject identity as the clustering variable. This approach explicitly accounts for the non-independence of multiple electrodes recorded within the same participant, providing valid population-level inference while appropriately handling the hierarchical structure of the data. Odds ratios (OR) with 95% confidence intervals and Wald χ^2 statistics are reported for all GEE comparisons.

We note that our claim of anatomical specificity of response profiles was quantitatively validated using two complementary approaches, both employing generalized estimating equations (GEE) models accounting for within-subject clustering of electrodes. First, we assessed deviations from spatial randomness by comparing region-wise response frequencies against null distributions generated by random sampling from the full insular electrode population. Second, we performed pairwise regional comparisons between all four insular regions (Supplementary Figure 5 with statistics).

Building on the previous experimental approach in combination with a high-resolution microstructural atlas of the insula, a pattern of four distinct insular regions emerged with unique functional and structural properties. To validate the anatomical specificity of response profiles in each region (pain/temperature, somatosensory, visceral, anxiety, salience, body-part representation, and no response to stimulation), we employed two complementary statistical approaches. Given the nominal quality of our data and non-independence of electrodes — as multiple electrodes were sampled per subject — we used generalized estimating equations (GEE) with a binomial family and exchangeable working correlation structure to account for within-subject clustering.

Salience task

To control for inter-subject variability, we additionally conducted subject-level permutation tests by aggregating electrode t-values per participant within each cluster, then randomly flipping the sign of within-participant cluster differences to create null distributions under the hypothesis of no systematic effect.

CCEP

To test for directional asymmetry while accounting for the non-independence of multiple electrode pairs within participants, we used permutation testing (10,000 iterations). On each iteration, we randomly flipped the direction labels (forward ↔ backward) for all pairs from each participant together as a unit, then calculated the mean difference across all pairs for that permuted dataset. This process generated a null distribution comprising 10,000 mean differences, representing the range of values expected by chance in the absence of true directional asymmetry. The observed mean difference was considered statistically significant ($p < 0.05$) if it was more extreme than 95% of the values in this null distribution.

6. The manuscript would benefit from a clearer crosswalk across anatomical frameworks. Different organizational schemes appear across the manuscript—for example, anterior/posterior distinctions in the subjective report analyses versus clustering-derived Regions I–IV in the connectivity analyses. Although each framework may be justified, the current presentation makes it somewhat difficult to integrate the results into a single coherent anatomical model. A clearer crosswalk or unifying figure linking these parcellation schemes would substantially improve coherence.

We thank the Reviewers for highlighting this issue and believe our revisions substantially improve anatomical coherence across the manuscript, allowing the reader to integrate findings across both frameworks.

Throughout this manuscript, two complementary anatomical frameworks are used. The anterior/posterior distinction refers to the division at the central sulcus of the insula, a boundary commonly used in the clinical and functional literature, illustrated in both the gross anatomical and flattened views in Figure 1A. The data-driven clustering analysis (Figure 1c) yielded four regions (I–IV), which map onto this framework as follows: Regions I and II correspond to anterior insula, while Regions III and IV correspond to posterior insula. Where results are described using anterior/posterior terminology, the corresponding cluster identities are noted parenthetically. We have now added this clarification to the legends of Figures 1-3.

7. The rationale for the task-based approach needs to be clarified. The inclusion of the Gradual Onset Continuous Performance Task is not yet fully justified relative to the paper's main aims, since the task does not directly probe body-signal processing. At present, this component feels somewhat disconnected from the main argument, and its findings are not discussed in a way that substantially advances the central claims. If the authors wish to retain this section, they should clarify more explicitly why it is necessary and how it informs the overall model of insular organization. Otherwise, it may dilute the manuscript's focus.

We thank the Reviewers for this important critique. We agree that the rationale for including the GradCPT requires clearer articulation. We respectfully wish to retain this component for two reasons that we believe are central to the manuscript's broader contribution.

First, the anterior polar insular cortex was a relatively silent region in the context of direct cortical stimulation, rarely producing any overt responses compared to more posterior regions. The GradCPT results demonstrate that when probed with a directed, ecologically sensitive experimental paradigm, this region yields robust and interpretable functional signals, even though it was “silent” and did not induce any reportable conscious experience during restful electrical stimulations. Our findings speak directly to the Reviewers' earlier remark about how the experimental task design and setting and expectation matter critically for uncovering the functional repertoire of regions that appear silent to stimulations and quiescent under standard conditions.

Second, and relatedly, the GradCPT experiment engages psychological processes, including cognitive control, sustained attention, and response selection and inhibition, that are canonically attributed to prefrontal cortex. The recruitment of anterior insular Region IV during these processes therefore extends our model of insular organization beyond bodily perception alone and supports our CCEP connectivity findings that Region IV is positioned in the interface between the prefrontal cortex and more posterior insular regions. This reframing is, we believe, one of the more conceptually significant implications of our present work.

Together, these two points, i.e., the methodological argument for directed task probing of putatively “silent” cortex, and the functional argument for insular involvement in prefrontal-like

cognitive processes, make us believe that the GradCPT findings can be integral rather than peripheral component of the manuscript.

We have made this reasoning explicit in the revised Discussion.

Minor comments

Clarify general limitations of electrical cortical stimulation for sensory mapping

The Discussion would be strengthened by acknowledging the broader limitations of electrical neurosurgical stimulation for sensory mapping studies. Although this approach offers a rare and powerful opportunity to probe human brain–experience relationships, stimulation-evoked percepts do not necessarily recapitulate physiological signaling, and the resulting phenomenology may depend on stimulation parameters, local current spread, clinical context, and the structure of subject questioning. These considerations are relevant to how strongly one can infer normal functional architecture from evoked reports.

We have now rephrased the final paragraph of the Discussion to include the points raised by the Reviewers as follows:

In closing, we emphasize that stimulation-evoked subjective experiences likely capture only a subset of the functions of the stimulated brain region. As detailed elsewhere (Siddiqi et al., 2022), while electrical stimulation provides valuable causal evidence linking a region to conscious perceptual or behavioral phenomena, the likelihood of eliciting such experiences depends on whether electrical discharges engage the neural processes underlying access consciousness (Block, 1990; Dehaene & Changeux, 2011; Sergent, Baillet, & Dehaene, 2005). Thus, our findings should be interpreted as reflecting only a subset of insular processes that become consciously accessible and reportable, rather than those remaining out of reach of consciousness. We therefore interpret our stimulation results as informative about the functional repertoire of insular subregions, while recognizing that they reflect responses to an electrical perturbation rather than direct readouts of physiological function. Furthermore, we acknowledge the general limitations of electrical cortical stimulation as a tool for sensory and functional mapping. Although direct cortical stimulation offers a rare and privileged window into human brain and conscious experience relationships, bypassing the inferential constraints of correlational studies (Siddiqi et al., 2022), stimulation-evoked percepts do not necessarily recapitulate physiological signaling under natural conditions. The phenomenology elicited by stimulation may be shaped by factors extrinsic to a region's normal functional role, including stimulation parameters (frequency, amplitude, pulse width, and duration), local current spread to adjacent or connected regions, the clinical context of the recording environment, and the structure and framing of subject questioning, which may influence whether and how patients report subjective experiences. It remains to be determined to what extent these considerations limit the strength of inferences one can draw from stimulation-induced reports about the normal functional architecture of a region. At the same time, the convergence of our stimulation findings with independent evidence from task-based recordings and connectivity analyses, each carrying distinct methodological assumptions, provides a degree of triangulation that partially mitigates these concerns.

The study does not measure the physiological “body loop” itself

A further limitation is that, although the study evaluates insula activity and connectivity, it does not directly record the relevant physiological signals arising from the body itself—that is, the peripheral component of the “body loop” in Damasio’s formulation. This matters because it remains difficult to determine whether insula stimulation may have triggered downstream autonomic or visceral changes that then contributed to subsequent sensory representations. Without concurrent measurement of physiological outputs, it is challenging to disambiguate percepts arising more directly from cortical stimulation from those that may have been mediated,

amplified, or shaped by stimulation-evoked changes in the body. I think this limitation is especially important in a manuscript making claims about interoceptive processing, because the body proper is not actually measured and is therefore effectively absent from the inference chain.

We have now added the following paragraph in the Discussion

In regard to our interpretation of posterior insular Regions I and II, we acknowledge an important limitation in our study, namely that concurrent with electrical stimulations, we did not record physiological signals in the body itself. Therefore, it remains to be known if cortical stimulation of the insula may have triggered downstream bodily changes that in turn shaped subsequent sensory representations, making it difficult to fully disambiguate percepts arising directly from cortical perturbation from those mediated by stimulation-evoked peripheral responses. Future studies combining intracranial stimulation with concurrent physiological monitoring will be important for closing this inferential gap.

Clarify the relationship between stimulation and connectivity analyses

Because Region IV is described as largely unresponsive to stimulation yet influential in connectivity analyses, it would be helpful to clarify whether the connectivity results were derived independently of the stimulation findings, and if so, how the authors see these two forms of evidence informing one another. This is especially important given the shift in parcellation framework across analyses.

We appreciate the Reviewers' suggestion. To clarify, the connectivity and stimulation analyses were conducted independently, using distinct datasets and methods, and across partially overlapping but not identical cohorts of participants. The connectivity analysis was, however, interpreted within the anatomical framework established by the functional stimulation mapping, that is, the quadripartite parcellation derived from stimulation-evoked response profiles served as the anatomical frame within which connectivity results are reported. While the connectivity results are therefore interpreted within the stimulation-derived parcellation, the two analyses remain methodologically independent: the connectivity patterns were not used to define the parcellation, nor did the stimulation results constrain the connectivity findings. Their convergence thus reflects two independent lines of evidence pointing toward the same conclusion about Region IV's functional role.

As for the second point raised by the Reviewers, we now explain in the revised manuscript, that the apparent paradox of Region IV being largely unresponsive to stimulation yet influential in connectivity analyses is reconciled by its position within large-scale cortical networks. In our prior work (Fox et al., Nature Human Behaviour, 2020), we demonstrated that the probability of eliciting subjective experiences with electrical stimulation decreases along a cortical gradient, from highest rates in focal sensorimotor regions to lowest rates in highly distributed, integrative association networks such as the prefrontal cortex. Region IV's connectivity profile, characterized by strong coupling with prefrontal cortices, places it squarely within this integrative network architecture, suggesting that its functional contribution emerges from large-scale network interactions rather than locally generated, consciously accessible percepts. The stimulation silence of Region IV and its connectivity fingerprint are therefore not contradictory but mutually reinforcing: both point to a region whose function is distributed, prefrontally embedded, and unlikely to be captured by focal stimulation-evoked phenomenology alone.

Clarify within-individual variability in subjective reports:

The authors note that some participants contributed to more than one category of subjective response. It would be useful to elaborate on this more fully, including the extent of within-

individual variability across stimulation sites. This could offer additional insight into insular heterogeneity and may help interpret the mapping of response categories across the cortex.

We thank the Reviewers for raising this point and apologize that our original description was not sufficiently clear. To clarify: individual participants could contribute to more than one response category depending on their electrode coverage, but each individual electrode contact yielded responses belonging to a single category. For example, a participant with contacts in both Region I and Region III would contribute somatic reports from the former and visceral reports from the latter. We have now added this clarification in the text:

some participants contributing to more than one category of responses (depending on the number of insular electrodes and distribution of these electrodes across multiple insular regions each of which contributed to a different category of responses).

Clarify how directionality and asymmetry were established

The manuscript would benefit from additional detail regarding how the directionality of insular connections was estimated from EEG connectivity measures. Similarly, the abstract's statement that the findings link cytoarchitecture to "directed and asymmetric electrophysiological interactions" would benefit from clarification, particularly if lateral asymmetry is being implied.

We recognize that our original statement was insufficiently clear. Directionality in our CCEP framework is inherently defined by the stimulation paradigm itself: by systematically stimulating electrode A while recording at electrode B, and then reversing this (stimulating B while recording at A), we obtain two independent, empirically-measured connectivity estimates for each direction (A to B vs B to A). This forward/backward design provides a direct, physiologically grounded measure of directed connectivity that does not rely on inferential techniques such as Granger causality or dynamic causal modelling, but rather on the causal logic of direct cortical stimulation.

To establish asymmetry in the connectivity of insular Regions with each other, we employed three complementary approaches at different levels of data aggregation: electrode-pair level permutation testing (10,000 iterations) with direction labels flipped at the participant level to preserve within-subject dependence; cluster-based permutation testing on directional contrast spectrograms across the full time-frequency spectrum (5,000 iterations); and a subject-level sensitivity analysis averaging electrode pairs per participant, which is independent of electrode-level clustering effects.

We also clarify that the term 'asymmetric' in the abstract refers to directional asymmetry in connectivity strength between insular subregions, specifically, that the magnitude of evoked responses differs depending on which region is stimulated versus which is recorded, rather than lateral hemispheric asymmetry. These issues are clarified in the revised Methods accordingly.

Improve manuscript organization and figure/reference consistency

Some parts of the manuscript could be better organized, with methodological details consolidated more clearly in the Methods and interpretive content reserved for the Discussion. In addition, the figures and references should be checked carefully for consistency, including mismatches in figure legends and inconsistency in citation formatting.

We thank the Reviewers for highlighting this. We have addressed these issues and believe the revisions have improved the overall clarity and presentation of the manuscript.

Reply to Reviewers 3 and 4

This manuscript by Quabs and colleagues presents an intracranial electrophysiology study of the human insular cortex, integrating data from 87 SEEG participants within a recently developed 16-area cytoarchitectonic atlas. Three complementary methods are combined: (1) 50 Hz electrical stimulation with collection of reports of subjective feelings (in 50 participants), (2) task-evoked recordings during a Gradual Continuous Performance Task (GradCPT), and (3) cerebrocerebral evoked potentials (CCEPs) mapping intra-insular and insular-cortical connectivity.

This study is a broad characterization of the insula in terms of stimulation-evoked feelings, somatotopy of somatosensory and pain sensations, and causal mapping of the effects of stimulation within the insula and its associated cortical networks. The authors rely on the previous Quabs et al. cytoarchitectonic parcellation and group them into four zones: Region I (dorsal-posterior) encodes somatotopically organized nociceptive/thermoceptive and somatosensory signals; Region II (ventral-posterior) integrates somatic information across multiple body parts without clear somatotopy; Region III (mid-anterior) is associated with visceral sensations and anxiety; and Region IV (anterior-polar) is largely silent to stimulation (does not evoke subjective feelings) but shows robust activation to salient events (which also require cognitive control; see below) in the GradCPT. A central claim is that the functional boundaries of these zones align very well cytoarchitectonic boundaries, though I don't think this claim is directly tested nor does it need to be strictly true for the paper (and the parcellation) to have value.

We thank the Reviewers for their careful and accurate summary of our work. Regarding the GradCPT findings and their interpretation, we have taken this feedback to heart and added substantial clarification, as detailed in our replies below. Regarding the claim about cytoarchitectonic boundaries, we agree with the Reviewers' observation that the alignment between functional and cytoarchitectonic boundaries is not directly tested in the present study. Importantly, as the Reviewer notes, the value of the parcellation and the strength of our findings do not depend on this claim being strictly true. We have therefore removed the mention of cytoarchitectonic matching from the manuscript, and we believe the paper is stronger and more precise for it.

In addition, CCEP connectivity mapping reveals asymmetrically strong outbound influence from Region IV to the other three regions, with strong bidirectional connections to prefrontal cortex, while Region I connects preferentially to somatosensory cortices. The authors interpret these findings as evidence for a top-down modulatory pathway from anterior to posterior insula with implications for cognitive modulation of pain and interoception.

We again thank the Reviewers for their insightful interpretation of our work.

The paper has many strengths. The functional organization of the insula is of persistent and central importance, and this is an authoritative contribution.

The group has outstanding expertise in both insula neuroanatomy and iEEG mapping. There is a tradition of similar iEEG mapping studies of the insula, appropriately cited here, but this one includes a larger sample size than most and an open-ended characterization of evoked sensations. There is strong rigor in including data from reports that are evoked selectively by real (vs sham), reproducible, and dose-dependent stimulation. It is remarkable that 51% of stimulation sites evoked sensations with these criteria. The ability to compare responses across

stimulation frequencies — e.g., finding that pain is not simply evoked by more intense stimulation — is a strength. Stimulation-evoked feelings are also combined with CCEP mapping, which to me is the most novel aspect of the study. The finding of asymmetrically strong outbound influence from Region IV to the other three regions, with strong bidirectional connections to prefrontal cortex, is striking. This is very meaningful when coupled with the idea that they are “silent” in the sense that stimulating this area does not evoke feelings, and responsive to the cognitive task. The figures are clear and informative. I have no doubt that the study will be widely cited and used.

We are grateful to the Reviewers 3 and 4 for their thoughtful and encouraging assessment of our work. It is highly gratifying to have the significance of our study recognized. We also sincerely appreciate the Reviewers’ acknowledgment of the rigor of our stimulation criteria (i.e., including only selective, reproducible, and dose-dependent responses) which were deliberately designed to ensure the validity and interpretability of the subjective report data. We are pleased that these methodological considerations were recognized and valued.

I have a number of comments below, but all of these are addressable within the scope of the paper with revisions and are thus intended as suggestions for improvement and clarification.

We have carefully addressed each of the Reviewer's suggestions in the detailed replies below, and we believe the revisions have meaningfully strengthened the manuscript. We hope the revised version meets with the Reviewer's approval.

Introduction and framing

The introduction is grounded in interesting questions, but the framing could be refined. It describes two processing principles as opposing theories, a posterior to anterior “bottom up” and a prefrontal to insular “top down” predictive system. But these do not seem to be opposing to me, as current conceptions hold that both types of processes occur in the insula and other circuits, so the “opposing theories needing reconciliation” framing seems to be a bit of a strawman. The study has a great deal of value, but it does not really provide a “causal examination of these two seemingly opposite views” or adjudicate between top-down and bottom-up influences. What we do need is more specific information about directional connectivity and flow of information, and the study provides this.

We agree entirely with the Reviewers that the framing of our work can be presented in a better way. We have now revised the Introduction according to the Reviewers suggestion. Rather than positioning the study as adjudicating between top-down and bottom-up accounts, we now frame our contribution around what the study genuinely provides, namely direct empirical characterization of the directional flow of information across insular subregions and between the insula and other cortical regions. This framing is both more accurate and, we believe, more compelling as the directional connectivity findings stand on their own merits without requiring an artificially polarized theoretical backdrop.

We thank the Reviewer for guiding us toward a stronger and more precise framing of the work. The new Introduction reads as follows:

While the prevailing model of insular functions, based on the notion of interoception, emphasizes a largely bottom-up (i.e. posterior to anterior) flow of information processing, another complementary framework of “interoceptive inference”(Seth, 2013) proposes an active ‘top-down’ processing within the framework of predictive coding(Barrett, Mesquita, Ochsner, & Gross, 2007; Barrett & Simmons, 2015). In this framework the insular cortex, with strong top down inputs from the

prefrontal cortices, supports error-based learning of feeling states and uncertainty (Singer, Critchley, & Preuschoff, 2009) and responding to interoceptive mismatches that underlie anxiety (Paulus & Stein, 2006). As both bottom-up and top-down processes may operate concurrently within insular circuits, a better understanding of insular functional organization requires characterizing not only the regional specialization of insular processes but also the directionality and asymmetry of the information flow within and across the insula and between insula and other brain regions that support related functions. The present study was designed to address this need directly, providing causal, electrophysiologically grounded evidence for the directional flow of information across insular subregions and between the insula and the brain.

The authors state that the “main finding is striking correspondence between microstructural architecture and function”. But the results don’t exactly reinforce this. It is a matter of interpretation how closely the boundaries of stimulation-evoked sensations follow the cytoarchitectural boundaries; sometimes they seem to more, sometimes not. In some cases where they clearly do not, the discrepancy is explained away as a failure of anatomical alignment (the mapping between the atlas parcels and these brains is, after all, approximate). This makes the claim of “striking correspondence” circular. However, the value of the paper to me does not hinge on this question, so I’m not sure that correspondence with cytoarchitecture is the “main finding”.

We thank the Reviewers for their comment. We agree that the claim of “striking correspondence” between cytoarchitectonic boundaries and functional zones was unnecessarily overstated and have now removed this characterization from the manuscript. As the Reviewers correctly note, the correspondence is variable around the border zones – an observation that led us blame this on the failure of anatomical alignment between native anatomical space and standardized atlas space. However, we agree with the Reviewers that the paper’s central contributions do not hinge on this correspondence. There are more compelling and novel aspects of our work that could be highlighted as main findings of the study. We have revised the Abstract, Introduction, and Discussion to foreground these contributions more accurately, and we believe the manuscript text reads more strongly and less controversially as a result.

REVIEWERS: If you really want to test correspondence, some more quantitative metrics and analysis would be appropriate. Some quantitative criteria might include: (1) Is the zone evoking a particular sensation contained completely within one parcel? (Or one group of parcels, but this introduces extra flexibility that would require cross-validation or similar to test). (2) Is the distribution of sensations evoked at different locations homogenous within a parcel (e.g., using a spatial poisson model) or not? (3) Is this parcellation a better description of the boundaries of functional zones (stimulation-evoked feelings) than other parcellations (e.g., using mutual information; e.g., the insular parcellation in the brainnetome atlas itself)? I don’t see that existing analyses (Supp Fig 5) really address this.

We thank the Reviewers for these valuable methodological suggestions. The three proposed quantitative approaches of containment analysis, spatial homogeneity testing, and parcellation comparison via mutual information are elegant and would indeed provide a more rigorous formal test of cytoarchitectonic correspondence.

We respectfully ask, however, that these analyses be considered beyond the scope of the current manuscript. Implementing a full quantitative cytoarchitectonic validation framework would constitute a substantially different and expanded study, one that would require dedicated methodology, additional data, and a level of treatment that cannot be adequately compressed into a revision. Such detailed linkage could be addressed in a future study.

The integration into the paper of previous iEEG work is understandably more complete than the integration of other work (e.g., fMRI), but there is an opportunity to relate to fMRI findings more closely. For example, it would be interesting to comment on the relationship of the findings with recently published fMRI work (<https://www.nature.com/articles/s41467-026-71568-9>), which appears to identify “domain general” anterior insular zones that may be related to the “silent”, “top-down” anterior insular zones you identify here. This study also includes a mapping between insular zones and other cortical systems that may converge or diverge with your CCEP mapping.

We thank the Reviewers for drawing our attention to the study by Kwon et al. We have now added the following paragraph to the Discussion to emphasize the complementary nature of our work and that of Kwon et al:

Our findings converge with a recent large-scale fMRI mega-analysis (Kwon et al. 2026, Nature Communications), which identified a domain-general convergence zone in bilateral dorsal anterior insula responding across pain, appetitive, aversive, and cognitive control domains, surrounded by domain-selective zones in more posterior and ventral insular regions. The location and functional profile of this domain-general zone corresponds closely to our anterior-polar Region IV, a subregion largely silent to direct stimulation yet robustly recruited during the GradCPT and strongly coupled with prefrontal cortex in our CCEP analyses. The domain-selective zones for pain and aversive processing in posterior and ventral anterior insula - identified in this fMRI study- correspond well to our Regions I–II and III respectively, and the cytoarchitectonic mappings in both studies converge on the same Julich-Brain atlas parcels. Critically, our CCEP mapping addresses the notion of causal effective connections, demonstrating that Region IV exerts asymmetrically strong outbound causal influence over posterior insular regions while receiving comparatively weaker indirect inputs. Together, the two studies provide complementary evidence: fMRI establishes the domain-general functional profile of the anterior polar insula at the population level, while our intracranial approach reveals the directed electrophysiological pathways through which this region exerts its influence.

A really wonderful contribution would be to share MNI-space coordinates of stimulation sites and evoked sensations (I note that this seems to be included in the data sharing statement, which is great), and/or maps of smoothed locations of various types of sensations (e.g., in MNI-space nifti image files). This would allow for closer quantitative comparisons with other studies. I think this is at the authors’ discretion, but it would really enhance the value of the work.

We are pleased to confirm that MNI-space coordinates of stimulation sites and associated evoked sensations will be included in our data sharing. We agree that these maps will substantially facilitate quantitative comparisons with other datasets, including fMRI studies, and we are glad to make them available to the community as part of our data release.

Sampling of self-reported feelings

An issue to consider in future studies is how the sampling of feelings is done. My understanding is that patients were asked how they feel in an open-ended way. This has strengths, but it can also lead to biases in the kinds of feelings patients expect, attend to, and prepared to report. If one feeling is present it may mask other feelings that are present but would only be reported if patients are asked specifically about them. That is why emotion researchers usually sample categories systematically (e.g., asking about multiple feelings in response to a stimulation), and medical checklists ask patients to reflect on each of a series of categories one at a time. Those methods also have issues but I wonder about the kinds of biases introduced by the methods here.

We appreciate the Reviewers' suggestion and agree entirely with it. The Reviewers are correct that our procedure used open-ended questioning, in which patients were asked to describe their experience following stimulation without being prompted with specific categories. We utilize this approach as it has genuine strengths, it minimizes response demand characteristics and reduces the risk that patients report sensations they did not actually experience simply because they were suggested by the experimenter but, as the Reviewers point out, it carries its own, namely that spontaneously reported sensations may underrepresent feelings that are present but below the threshold of spontaneous report, or that patients would not think to name without prompting.

In the future we will use systematic category sampling, as used in structured emotion research and clinical checklists, and we agree with the Reviewers that it would indeed provide a more complete picture of the phenomenological space associated with each insular subregion, and might reveal sensations that are masked or suppressed under open-ended conditions. We acknowledge this as a meaningful limitation of the current methodology and have added a brief note to the Methods as follows:

The open-ended nature of our questioning procedure minimized demand characteristics and reduced the risk of patients reporting sensations they did not genuinely experience, but we are mindful that it may underrepresent feelings that are present yet below the threshold of spontaneous report or that patients would not independently volunteer without direct prompting. Systematic category sampling approaches, such as those utilized in emotion research, where patients are asked to rate the presence or absence of each specific sensation from a predefined list, would provide a more exhaustive characterization of the phenomenological space associated with each insular subregion and may reveal sensations masked under open-ended conditions.

In addition, I wonder if the location of stimulation sites was randomized across repeated trials. If not, there is potential for confirmation bias. e.g., with stimulation of site x, the patient says, "I feel tingling in my right leg". The researchers then stimulate the same site, and the patient is primed to report the same sensation, so data are not independent. This may not be an issue, but more clarity on potential caveats/issues would be helpful.

We thank the Reviewers for highlighting this important issue and have clarified this issue in the Methods as follows:

Moreover, we clarify that stimulation sites were selected in a variable order across mapping sessions and patients were not informed of either the site being stimulated or the stimulation intensity on any given trial. This blinding reduced the potential for site-specific response priming, as patients could neither anticipate which location would be stimulated next nor adjust their reports based on perceived stimulation strength. Further safeguards were provided by our strict reproducibility criteria: responses were included only if they were absent during sham stimulations and showed dose-dependence with stimulation intensity, parameters to which patients were fully blinded. Together, these procedural controls substantially mitigate the potential for possible confirmation bias.

More broadly, a curious feature of the findings is that stimulation evoked feelings quite frequently, but only evoked a very small subset of the possible range of feelings humans experience. e.g., I don't think there are any reports of emotions (no disgust?), no itch (?), and no reports of social cognitive judgments or perceptual distortions, etc. Was this related to the scope of the instructions and the way questions were asked of patients, or do you think this reflects a

truth about insular function — i.e., these other feelings are more indirectly related?

We thank the Reviewers for this very intriguing question, and we must be candid that we do not have a fully satisfying explanation for the restricted range of feelings evoked by insular stimulation in our dataset, and were ourselves surprised by our findings. The absence of reports of discrete emotions such as disgust, social cognitive experiences, perceptual distortions, or itch, all of which have been associated with insular activity in neuroimaging studies, was striking, but we are confident that they were not due to the methodological artifact of how questions were posed to patients. We have now added the following short paragraph in the Discussion about this issue, as we believe it represents one of the more thought-provoking unresolved issues raised by our present study – We thank again the Reviewers for raising it:

It is worth noting that the restricted phenomenological range we observed here contrasts markedly with findings from our prior intracranial stimulation work in other cortical regions where we followed exactly the same procedures. For instance, as we have already published, stimulation of the precuneus (Lyu et al., 2023), dorsal anterior cingulate cortex (Parvizi, Rangarajan, Shirer, Desai, & Greicius, 2013), and hypothalamus (Parvizi et al., 2022) in some of the same patients elicited remarkably complex and high-order experiences including self-depersonalization, dissociative states, and volitional phenomena such as a compelling “will to persevere” that are considerably more elaborate than anything we encountered with insular stimulation. The fact that such complex phenomenology can be evoked by direct stimulation elsewhere in the brain, but not in the insula, makes the restricted range of insular responses all the more puzzling and suggests this is not simply a methodological artifact of how questions were posed to patients. One interpretation, consistent with our connectivity findings, is that the insula's contributions to complex emotional, social, and self-referential experience emerge principally from its participation in large-scale distributed networks rather than from focal insular activity alone. Stimulating the insula in isolation may be insufficient to recruit the broader network dynamics required to generate fully formed emotional or social experiences.

Last comment here: Could you provide a more detailed breakdown somewhere about what visceral sensations mean? e.g., any nausea? Rotation/vestibular sense? Stomach churning, etc?

We have now added more details in the legend for Figure 3 (Region III)

While anxiety responses were reported directly as “feeling of as nervousness” or “feeling anxious”, the visceral sensations were over all in one of the following categories: Nausea and GI sensations (e.g., dry heave, squirly feeling through stomach, queasy feeling); Chest and thoracic sensations (pressure or sudden rush in chest); Rising sensations (rising sensation from tailbone up through abdomen, taking breath away, bottom up sensation through center of body, rolling sensation down in stomach); Wave sensations (like being on a boat, whole body wave, body vibrating, falling backwards, sensations going awry in head, ripple going across head like rock hitting flat water); Throat and neck sensations (sensation in throat and base of neck, sinuses, lip, and chest/lung area pushing down sensation, like wanting to talk but mumbling, unable to talk).

Clustering and justification for reclassification of Area Id4

The four-zone structure relies on k-means clustering using the elbow method for choosing 4 clusters. This is OK but is not generally considered a rigorous way to choose clusters; bootstrap and permutation tests with, e.g., mean silhouette values provide a more principled and modern

metric. This is not a big deal, as the choice of number of clusters is inherently somewhat arbitrary in my opinion. K-means also makes strong assumptions about the shapes of clusters (multivariate Gaussian), and different methods are sensitive to different shapes. It might be good to just acknowledge that the four-cluster solution is a heuristic solution rather than a hard truth.

We thank the Reviewers for this methodological point and fully agree. We have now updated the cluster selection procedure to include silhouette score and Davies-Bouldin score alongside the elbow method, all three of which converge on K=4. We explicitly acknowledge in the revised text that the four-cluster solution is a heuristic, i.e., alternative solutions such as K=5 are mathematically plausible, and that the final choice of K=4 reflects the best balance between mathematical optimality and biological interpretability. Figure 1 has been updated as follows:

Results:

Unsupervised k-means clustering identified four functional regions (Figure 1C). The optimal cluster number was determined by convergence of three complementary metrics: silhouette score, Davies-Bouldin score, and elbow method, all pointing to K=4. We acknowledge that the four-cluster solution represents a heuristic rather than a hard truth, i.e., alternative solutions such as K=5 are mathematically plausible and cannot be formally excluded. The selection of K=4 reflects the best balance between mathematical optimality across all three metrics and the biological plausibility of the resulting clusters. Nevertheless, the solution was robust across metrics and validated by 3D multidimensional scaling (stress < 0.01). Region I (dorsal-posterior insula: Ig1, Ig2, Ig3, Id2) exhibited predominantly pain and temperature responses. Region II (ventral-posterior insula: Id3, Id4, Id5, Ia1, Ia2) showed primarily somatosensory responses. Region III (mid-anterior insula: Id6, Id7) demonstrated the highest frequency of visceral responses, and Region IV (anterior pole: Id8, Id9, Id10) was largely unresponsive to direct electrical stimulation. These findings reveal a four-fold structural-functional organization of the human insula based on electrically evoked response profiles of microstructural areas.

Methods:

We used standardized feature vectors and applied K-means clustering with cluster number optimization determined by convergence of three complementary metrics: silhouette score, Davies-Bouldin score, and the elbow method. For visualization, we performed multidimensional scaling (MDS) to project the high-dimensional feature space into 3D Euclidean space, enabling spatial interpretation of functional similarity between areas. Stress was calculated as a goodness-of-fit measure between the resulting MDS configuration and the underlying data, quantifying how well pairwise distances are preserved in the lower-dimensional representation.

K-means clustering assigned Id4 to Region II (the somatosensory-dominant cluster), consistent with Id4's own profile, which the authors describe as "predominantly somatosensory rather than pain and temperature" (line 227). The authors then manually reclassified Id4 to Region I (the pain/temperature-dominant cluster) because Id4's responses were localized to a single body part (the mouth), completing Region I's somatotopic gradient. What the need for this reclassification reveals is that the original feature set may have lacked a functionally meaningful dimension -- somatotopic specificity (i.e., whether responses are localized to a single body part versus distributed across multiple parts). The clustering could not capture this because it only encoded response category frequencies, not the somatotopic precision of responses. What

would happen if the clustering were done on a more complete set of features, including somatotopy? If Id4 groups with Region I it would strengthen the justification for the current scheme. If the results differ in other ways, you might note that there are alternate groupings.

The Reviewers are correct that somatotopic specificity represents a functionally meaningful dimension that was not captured by response category frequencies alone. However, we would like to clarify that single body part localization alone was not sufficient to distinguish Region I from Region II. Both regions can have sites exhibiting responses localized to a single body part. The critical difference lies in the somatotopic “order” of those responses: in Region I, single body part responses follow a meaningful and spatially coherent somatotopic gradient, whereas in Region II, responses do not follow a consistent or interpretable somatotopic organization and co-occur with a higher proportion of multi-localized responses. This distinction is precisely what makes Id4 functionally closer to Region I: its mouth-localized responses fit meaningfully into the somatotopic *order* in Region I (face, arms/ trunk and leg).

We fully acknowledge that capturing this somatotopic precision as a formal clustering feature will be methodologically challenging. We were unable to identify a sensible metric that would encode meaningful somatotopic “order” without over-complicating and overinflating the feature space in ways that would introduce arbitrary choices and reduce interpretability. We therefore opted for the more transparent descriptive approach, with the reclassification of Id4 explicitly justified on biological rather than algorithmic grounds.

We wish to note that our preliminary analysis suggests that the cytoarchitectonic and receptor-architectonic of Id4 is similar to the sub-areas in Region I than Region II. As noted, an ongoing dedicated follow-up study is underway by the Julich team, which we hope will provide a more rigorous quantitative basis for resolving this open question.

We hope the following new clarification in Figure II legend will clarify our position better:

It should be noted that the location of the mouth region (light blue lines) corresponds to parcel Id4 in the Julich atlas as it is microstructurally a transitional area of its own, and different from the most granular dorsal posterior insula. Although our initial clustering analysis placed Id4 in Region II (Figure 1C), we reassigned it to Region I for two reasons. First, stimulation of Id4 consistently evoked sensations localized to a single discrete body part, i.e., the mouth. Second, and more importantly, these mouth-localized responses fit coherently into the somatotopic organization of Region I, where body part representations follow a spatially ordered progression from leg responses posteriorly through trunk, arm, and face responses anteriorly. This somatotopic “order” is what distinguishes Region I functionally, and Id4's mouth responses align with it naturally. The mouth responses in Region II fit the responses in the rest of GI system in Region III abutting the Id4 area. We are mindful that future studies may position id4 as a transitional zone of its own binding the face area of Region I with the areas in Region III responsible for the rest of the GI system – descending from the oral cavity and down to stomach and beyond.

The broader underlying issue is how we are supposed to think of the organization. Clustering assumes that each cytoarchitectonic parcel belongs to one and only one group. It's a useful simplification, but why can't the underlying organization instead consist of broad gradients, with the zone show somatotopy being probabilistic, and partially overlapping (but partially diverging from) probabilistic zones showing visceral sensations, etc.

We thank the Reviewers for raising this important conceptual point. The question of whether cortical organization is better described by discrete boundaries or overlapping gradients is widely debated in both cytoarchitectonics and functional neuroimaging, and we appreciate the opportunity to clarify our position.

Based on extensive review of histological sections at the single-subject level (by the Julich co-authors of this manuscript - J.Q. and S.C.), our strong impression is that cytoarchitectonic boundaries within the insula are genuinely discontinuous within individuals with clear borders being identifiable at the level of individual brains, and the apparent gradients visible in MNI space largely reflect the averaging of individually sharp boundaries across subjects during spatial normalization, rather than true biological continuity.

A concrete example illustrates why this discrete substrate matters functionally. The granular posterior insula (Region I), characterized in our study by a pronounced granular input layer 4, predominantly elicits pain, temperature, and somatosensory experiences upon stimulation. The mid-anterior insula (Region III), which elicits visceral responses, has a fundamentally different cytoarchitecture — almost no cells in layer 4 but a highly differentiated layer 5. It is therefore more parsimonious to interpret these regional differences as reflecting distinct functional roles rather than as points along a continuous gradient. The cytoarchitectonic discontinuity between these regions supports a discrete model at the level of resolution available here.

We fully acknowledge, however, that discrete parcellation and gradient organization are not mutually exclusive: sharp cytoarchitectonic boundaries may coexist with graded functional transitions, and the degree to which functional zones respect or transgress cytoarchitectonic borders is an empirical question that the present dataset is not optimally positioned to resolve. Our dedicated follow-up study will directly quantify the relationship between cytoarchitectonic measurements and stimulation-evoked responses, with the explicit aim of formally testing whether functional organization is better described by discrete boundaries, overlapping gradients, or a combination of both. We believe this represents one of the most important open questions in insular organization and we look forward to addressing it rigorously.

There seems to be a methodological inconsistency in how CCEP data are interpreted. The F1 metric captures the early, N1-like component indicative of feedforward connectivity, and the claim of Region IV's dominant outflow rests on this metric. However, the finding that Regions I and II show stronger influence over Region III in the "late low-frequency range" (lines 399–402; interpreted as feedback) derives from spectrogram inspection without a formally defined metric, specified time windows, or statistical tests. The authors should either define a complementary metric for this late component with equivalent rigor or acknowledge the feedback finding as exploratory.

We thank the Reviewers for another very important methodological observation. We apologize that our original text did not include formal quantification equivalent to the F1 metric. We have now made this available in the revised manuscript. As we have detailed in our publication (Lyu et al, Nature Neuroscience 2025), F2 feature corresponds to the delayed low-frequency spectral change that is widely considered in the CCEP literature as a feedback signal. With statistical testing using cluster-informed permutation tests accounting for electrode non-independence, we now report these metrics with full statistics, confirming that significant directional asymmetry was observed exclusively between Region III and I and II, with no significant asymmetry for any other region pair. The relevant passage has been updated as follows:

Notably, the low-frequency asymmetry was observed only between the posterior regions (I and II) and the mid-anterior Region III. The same pattern reached significance exclusively for these

region pairs when comparing the late low-frequency components F2 with no significant asymmetry observed for any other region pair (F2: III→I = 0.49 ± 0.32 , I→III = 0.61 ± 0.25 , within-individual pair-level $p = 0.012$, $d = -0.377$, 95% CI [-0.715, -0.130]; III→II = 0.64 ± 0.19 , II→III = 0.68 ± 0.18 , within-individual pair-level $p = 0.012$, $d = -0.283$, 95% CI [-0.490, -0.094]).

One interpretation of the findings that I do really like is that insular Zone IV provides a bridge between prefrontal (lateral and ventromedial) systems involved in meaning-making (see Koban et al. 2021 Nat Rev Neurosci) and cognitive maps (e.g., Behrens 2018; Constantinescu and Behrens 2016) and somatic systems including pain, i.e., PFC → Zone IV → Zone I → sensorimotor cortex. Several labs have demonstrated pathways in rodents from anterior cingulate to Zone I regions (dorsal posterior insula) and Zone I regions to spinal cord to control nociception (Tan and Kuner 2017 Nat Neurosci is a good example).

We are grateful to the Reviewers for this elegant interpretive framework, and we agree that it captures something important about the functional architecture our data reveal. The proposed pathway of PFC → Region IV → Region I is strikingly consistent with our CCEP connectivity findings, which demonstrate asymmetrically strong outbound influence from Region IV to posterior insular regions, combined with robust bidirectional coupling between Region IV and prefrontal cortices including anterior cingulate cortex. Our data therefore provide direct electrophysiological support for precisely the kind of hierarchical modulatory pathway the Reviewer describes. We have incorporated this interpretive framework into the revised Discussion, citing Koban et al. (2021), Behrens (2018), and Constantinescu and Behrens (2016) in the context of prefrontal cognitive mapping and meaning-making. We believe this framing substantially enriches the mechanistic interpretation of our results and are very grateful to the Reviewer for pointing us toward it.

The following paragraph was added in the Discussion:

Our findings are in line with the extant literature proposing that the dorsal anterior insula (Region IV in our study) serves as a bridge between prefrontal meaning-making and cognitive mapping systems on one hand, and somatic and nociceptive systems on the other (e.g., (Koban, Gianaros, Kober, & Wager, 2021) (Behrens et al., 2018; Constantinescu, O'Reilly, & Behrens, 2016) (Tan & Kuner, 2021)). This theoretical framework, supported by imaging and animal studies, helps explain one of the more puzzling features of our dataset, namely why Region IV is largely silent to direct stimulation yet robustly activated during the GradCPT and so strongly positioned to influence posterior insular function (based on CCEP data). If Region IV's role is fundamentally to translate PFC-derived predictions and contextual representations into modulations of somatic and nociceptive processing, its function would be inherently relational and network-dependent and therefore unlikely to be captured by focal stimulation alone.

Requests for clarification/etc

- Were these patients included in previous publications, or is this a new cohort? How much of dataset is previously published? If patients have appeared in previous publications it is important to note because the data here are not fully independent of those.

Yes some of the patients in this study were included in prior studies:

Specifically:

- Two patients were included in Parvizi et al Neuron publication regarding dorsal anterior cingulate stimulation

- One patient was included in Parvizi et al 2022
- 17 patients were included in Duong et al 2023 publication.

- Some places say N = 50 with stimulation, some N = 47, some N = 56. It seems to depend on the type of stimulation/locations. Clarify briefly?

The varying participant numbers reflect the fact that different experimental components were available for different subsets of participants. We have now added a Supplementary Table with detailed information about the cohorts for each study element and to clarify this directly, the following new paragraph was added to the Methods section:

Of note, due to the long duration of data collection spanning more than a decade, and given the clinical setting of our experiments, not all experiments were conducted for every participant, and therefore not all participants contributed data to every experimental component of the study. The functional stimulation cohort comprised 50 participants, the salience task cohort 35 participants, the intra-insular connectivity cohort 47 participants, and the insulo-cortical connectivity cohort 56 participants. An exact overview of cohort overlap is provided in Supplementary Table 1.

- The figures are very nice, but it would be nice to also see the mapping of your insular diagrams onto MNI-space MRI images somewhere. It's otherwise hard to infer how the stim sites match, except by way of the atlas regions.

We thank the Reviewers for this suggestion. The MNI-space MRI images mapping our insular parcellation onto standard brain templates have been published in our companion cytoarchitectonic atlas paper (Quabs et al., *NeuroImage* 2022; DOI: 10.1016/j.neuroimage.2022.119453), where detailed volumetric maps of all 16 cytoarchitectonic areas in MNI space are provided. The 3D maps in the MNI space within the Julich Brain Atlas are as part of the EBRAINS Human Brain Atlas, where everyone can view and download the respective maps in different spaces.

To facilitate cross-referencing, we have added a citation to this paper at the relevant point in the manuscript and directed readers explicitly to the relevant figures therein. If the Reviewer feels it would further aid interpretability, we are happy to reproduce a summary MNI-space panel as a supplementary figure in the current manuscript.

- Are there sites that evoke multiple sensations? Yes, there are, but the data are presented in Fig 1 as though there is always only 1 sensation evoked, and later multiple sensations are discussed. Clarify that the multiple sensations are reports of multiple sensations in response to a single stimulus (and not an aggregation over different participants/stimulations?)

Each stimulation site evoked one predominant sensation category, and the multiple sensations discussed in the manuscript reflect responses from different sites and not multiple distinct sensations evoked simultaneously by a single stimulation of a single site. Figure 1 therefore accurately represents the response profile of individual sites. The only exception to this single-sensation-per-site pattern occurred in Region I, where stimulation of individual sites occasionally evoked either pain alone, temperature alone, or a combination of both pain and temperature sensations within a single stimulation sites. This co-occurrence at individual sites is precisely the reason we combined pain and temperature into a single category for the purposes of our regional analysis, as we have addressed in detail elsewhere in our replies. Outside of this specific exception, the presentation in Figure 1 faithfully reflects the within-site response structure.

- More granularity on the types of sensations would be helpful if possible. What is the relationship between the maps for pain and somatosensation? There is some discussion, but it's purely qualitative — “some sites” responded to warmth and pain, “other sites” somatosensations without pain. What is the quantitative joint distribution?

Are any of the visceral sensations painful?

None of the visceral sensations were painful, but they were aversive (we added details in Figure legend 3)

“Pain and temperature” is one category, but can you distinguish between painful and non painful temperature sensations?

We thank the Reviewers for pointing this out. Indeed, many participants were able to distinguish between warm sensations without painful qualities and hot or burning sensations with a clearly painful dimension, suggesting that patients are capable of making this phenomenological distinction when directly probed. We acknowledge that the combination of pain and temperature sensations may obscure potentially meaningful anatomical distinctions between the two response types. Informally, we did not observe a clear spatial separation between painful and non-painful temperature responses within the posterior insular Region I and II, suggesting they may share a common regional substrate consistent with the known properties of the granular posterior insula as a thermosensory and nociceptive convergence zone. A dedicated study with larger sample sizes and systematic probing of temperature quality would be needed to formally resolve this question, and we have noted it as a direction for future work.

“Somatosensory” includes pressure...was the pressure ever painful?

Pressure and vibration were categorized as somatosensory and therefore in no case painful.

Were any positive emotional responses ever elicited?

In one single patient, and in one single instance, a sense of euphoria was reported but we could not verify this in any other patient.

Or feelings of sadness or anergia/apathy/fatigue?

No feelings of anergia/apathy/fatigue were evoked

- “We reassigned area Id4 from Region II to Region I”. OK, but it's confusing to have Fig 1 with one scheme and then Fig 2 with a different scheme. I think this will confuse readers later.

We agree with the Reviewers that presenting two different regional assignment schemes across figures could be confusing to readers. We have now updated the legend for Figure 2 to explain the reassignment of area Id4 from Region II to Region I.

The new text in Figure legend 2 reads as follows

It should be noted that the location of the mouth region (light blue lines) corresponds to parcel Id4 in the Julich atlas as it is microstructurally a transitional area of its own, and different from the most granular dorsal posterior insula. Although our initial clustering analysis placed Id4 in Region II (Figure 1C), we reassigned it to Region I for two reasons. First, stimulation of Id4 consistently evoked sensations localized to a single discrete body part, i.e., the mouth. Second, and more importantly, these mouth-localized responses fit coherently into the somatotopic organization of Region I, where body part representations follow a spatially ordered progression from leg responses posteriorly through trunk, arm, and face responses anteriorly. This somatotopic “order” is what distinguishes Region I functionally, and Id4's mouth responses align with it naturally. The mouth responses in Region II fit the responses in the rest of GI system in Region III abutting the Id4 area. We are mindful that future studies may position id4 as a transitional zone of its own binding the face area of Region I with the areas in Region III responsible for the rest of the GI system – descending from the oral cavity and down to stomach and beyond.

- “We included parcels represented in at least 3 participants for each insular region” Can you show a map of which parcels are covered and which we simply do not have information about? This is important to interpret the results.

This information was already provided in Figure 1 (bottom right panel). The two grey areas (Id1 and Ia3) are the ones we had insufficient coverage.

- Can you more clearly label motor (M1/area 4), sensory (S1), and premotor regions on the brainnetome parcels in the CCEP analysis?

We have now revised the figure with labels of sensory motor regions

Typos and other minor comments

you might tone down the claim that the study identifies “an electrophysiological pathway for cognitive modulation of pain and bodily perception”. The study includes no condition testing pain modulation (e.g., no attentional or cognitive manipulation during noxious stimulation) or otherwise demonstrates this directly. It like this hypothesis, but it would be helpful to identify the claim as a hypothesis rather than a demonstrated finding.

We modified the text from “revealing an electrophysiological pathway for cognitive modulation of pain and bodily perception” to

“suggesting a possible electrophysiological pathway for cognitive modulation of pain and bodily perception”

the finding of anxiety responses from the stimulation of left (but not right) anterior insula...and then “the lateralization of anxiety responses to the right hemisphere needs....”

We apologize for the mistake. We have now corrected the text to

One exception was the finding of anxiety responses from the stimulation of left (but not right) anterior insula. However, only three of the five participants had bilateral electrodes and the lateralization of anxiety responses needs future validation.

The anxiety group showed higher mean Beck Anxiety Inventory scores -> this makes it sound like there is a group of patients with anxiety, but I think “group” refers to electrodes?

The group refers to patients in whom the stimulation of the insula induced anxiety

The drop from 87 to 50 participants is unexplained. Mentioned that 87 participants underwent SEEG, but only 50 received electrical stimulation mapping. No reason is given for why the other 37 were excluded.

As we have stated above, the varying participant numbers reflect the fact that different experimental components were available for different subsets of participants. We have now added a Supplementary Table with detailed information about the cohorts for each study element and to clarify this directly, the following new paragraph was added to the Methods section:

Of note, due to the long duration of data collection spanning more than a decade, and given the clinical setting of our experiments, not all experiments were conducted for every participant, and therefore not all participants contributed data to every experimental component of the study. The functional stimulation cohort comprised 50 participants, the salience task cohort 35 participants, the intra-insular connectivity cohort 47 participants, and the insulo-cortical connectivity cohort 56 participants. An exact overview of cohort overlap is provided in Supplementary Table 1.

Line 189: MDS is not a validation of the cluster solution. The elbow method alone determined $k = 4$. The MDS stress value (0.01) indicates the 3D projection faithfully preserves pairwise distances, not that four clusters is the correct number. Citing MDS as "validation" of the quadripartite scheme is incorrect. Proper validation would be e.g., silhouette analysis or permutation tests for cluster stability.

Please see our earlier response to "Clustering and justification for reclassification of Area Id4"

Lines 238–240: The "sharp transition" between Regions I and II is illustrated with one participant (four examples from Supplementary Figure 2 come from a single participant's brain). The paper does not report whether this sharp transition was systematically observed across all participants with electrodes traversing the Region I/II border. Group-level statistics showing different response proportions between regions are not equivalent to demonstrating sharp borders within individuals.

We wish to clarify that the four examples shown in Supplementary Figure 2 were drawn from four different participants, not a single individual. We apologize if this was not sufficiently clear in the figure legend and have revised it accordingly.

Regarding the broader question of whether sharp transitions between Regions I and II were systematically observed across participants, we agree with the Reviewers that group-level statistics demonstrating different response proportions between regions are not equivalent to demonstrating sharp within-individual borders, and that this distinction deserves explicit treatment. In participants with oblique electrode trajectories traversing the border between two different regions, which represents a subset of our cohort given the clinical constraints on electrode placement, sharp transitions in response category were consistently observed at the single-subject level, with adjacent electrode contacts separated by only a few millimeters showing qualitatively different response profiles.

We acknowledge that the sampling of sharp functional boundaries requires a systematic quantitative analysis of within-subject boundary sharpness in a larger number of subjects with sufficient coverage of all border zones, a direction we have note for future dedicated studies.

Line 449: an incomplete sentence? “The most anterior insula emerges as a hub, lacking direct access to subjective percepts, but exerting strong outbound influence across other insular networks and bilateral electrophysiological relationship with prefrontal, including the anterior cingulate, networks.”

We have now corrected the sentence to

The most anterior insula emerges as a cortical area lacking direct access to subjective percepts but exerting strong outbound influence across other insular networks, with bilateral electrophysiological relationship with prefrontal network, including the anterior cingulate.

Lines 506–509: The “as-if body loop” link is a large inferential gap. Connecting their findings to Damasio's concept requires demonstrating that Region IV's interactions with prefrontal cortex enable generation of bodily state representations without peripheral input. The CCEP data show that Region IV <-> prefrontal connectivity exists, but CCEPs reveal nothing about the content of information being transmitted. The leap from “Region IV is connected to prefrontal cortex” to “Region IV supports as-if body loops” seems to be speculative.

Please see our reply to Point 4 raised by Reviewers 1 and 2. We have now removed the mention of as-if body loop from the manuscript. We confirm that the surrounding text still holds together after the removal of as-if body statement.

Ideas for extensions (later)

Is it possible to characterize sensations evoked by sham stimulation (placebo)? Sensations be real but context-dependent and therefore not reproducible.

We thank the Reviewers for this insightful comment. The question of whether sham-evoked sensations reflect genuine but context-dependent perceptual experiences is an important one. We would point the Reviewer to our prior empirical work directly addressing this issue (Fox & Parvizi, Brain Stimulation, 2021), in which we systematically assessed the fidelity of first-person reports following sham intracranial electrical stimulation in a large sample of neurosurgical patients.

In that study, we administered 159 sham stimulations across 44 patients and found an overall false positive rate of only 6.9%, and critically, 75% of patients never produced a single false positive report. Furthermore, the few false positive responses that were observed could largely be attributed to lingering effects from preceding veridical stimulation rather than genuine placebo-induced percepts. Responses to sham stimulations only occurred if the sham stimulation was delivered shortly after the non-sham stimulation. No cases of complex confabulation or psychogenic motor effects were observed. These findings demonstrated that sham intracranial stimulation only rarely triggers any experiential response, and that patient first-person reports under these conditions are highly reliable.

We therefore believe that context-dependent placebo responses are unlikely to represent a significant body of data, given the established low false positive rate of sham stimulations.

Reply to Reviewer 5

In this paper, the authors use intracranial recordings from human epilepsy patients and also intracranial stimulation to map different functions of the insula during pain and interoceptive experience. Even though the intracranial dataset is a rich dataset, there are several issues such as mathematical and statistical errors, absence of any novelty, and failure to give proper credit to prior researchers, which precludes its publication in Nature or Nature-branded journals. I outline them below.

We thank the Reviewer for acknowledging that our intracranial dataset is rich. However, we must respectfully but firmly take issue with the three characterizations offered in this opening statement, as each of them is based on misinterpretation and misrepresentation of our manuscript.

- 1) Mathematical and statistical errors: As we demonstrate in our point-by-point responses below, the Reviewer's specific technical objections either mischaracterize established methodological practice in the field (e.g., labeling common average referencing as simply "wrong", or excluding sham stimulations as "cherry picking"), conflate methodological choices with errors (e.g., our use of cluster-based permutation testing for directional connectivity), or raise questions that are already addressed in our Methods section. We have not been shown a single calculation in our manuscript that is incorrect.
- 2) Absence of any novelty: Our study is the first to simultaneously integrate high-density intracranial stimulation mapping, task-based electrophysiology, cytoarchitectonic parcellation using a newly developed 16-area atlas, bidirectional causal connectivity mapping and somatosensory evoked potentials — in a relatively large cohort of participants. Four of five reviewers, including experts in this field, assessed our work as containing significant novel contributions. The Reviewer's assertion to the contrary is not substantiated by specific arguments in their report.
- 3) Failure to give proper credit to prior researchers: We acknowledge that our original submission had gaps in its citation of the prior literature, and we have corrected this in the revised manuscript by adding references from the Tandon group, Knight group, and Rylvlin lab and the cytoarchitectonic neuroimaging literature, among others. We note, however, that several of the references the reviewer claims were omitted, including Ostrowsky et al. (Cerebral Cortex, 2002, original reference 15) and our own prior work (Duong et al., Brain Stimulation, 2023, original reference 30), in fact were cited in the original submission. The Reviewer's characterization of our omissions as "highly unethical" goes well beyond what the peer review record supports, and we reject that characterization. Importantly, as we have explained in details in our replies, none of the added references alter the core novelty of our contribution, which rests on the simultaneous integration of cytoarchitectonic parcellation, causal stimulation mapping, and bidirectional electrophysiological connectivity in a single large human cohort, a combination that no prior study has achieved.

We address each of the reviewer's specific points in detail below.

First of all, the title is misleading:

“Functional Architecture of the Human Insula Revealed by Causal Intracranial Mapping”

The authors are mostly mapping during lower order cognition such as sensory processing, not higher level cognition. The title should be modified to reflect accordingly.

We respectfully disagree with this comment, which rests on a fundamental conceptual mischaracterization of our work. The Reviewer states that we are "mapping during lower order cognition such as sensory processing." This is incorrect on two counts.

First, electrical stimulation of the cortex does not constitute “lower order sensory processing” in any meaningful sense of the term.

Second, our study extends well beyond the stimulation mapping component. It includes: (1) task-based electrophysiological recordings during salience detection and action mode selection, which relies heavily on cognitive control – as other Reviewers have acknowledged to be higher-order cognitive processes; (2) bidirectional causal connectivity mapping between insular subregions and prefrontal, anterior cingulate, and orbitofrontal cortices canonically associated with cognition, emotion regulation, and executive function; and (3) a cytoarchitecturally grounded framework linking microstructural organization to directed electrophysiological interactions across the insular hierarchy.

We believe the title "Functional Architecture of the Human Insula Revealed by Causal Intracranial Mapping" accurately and precisely describes this body of work. We have retained it accordingly

In the Introduction, the authors mention:

“Previous studies, while offering important insights, have been limited in their approach: Most prior studies have examined the insula through isolated anatomical, functional, or connectivity lenses. Here, we unify these dimensions using a high-resolution and multimodal approach that enables an unprecedented characterization of its functional architecture using electrophysiological causal measures of its function and connectivity with not only the broader cortical mantle but also among its different subregions.”

These comments are simply factually wrong. The authors have omitted references to several other pioneers and their contributions to the comprehensive mapping of the anatomical, functional, or connectivity lenses of the human insula, including the different subregions of the insula.

I think it's highly unethical to omit a plethora of prior references which have comprehensively mapped the functions of the insula and then claim the already known findings of the insula as novel findings.

We respectfully but firmly disagree that our introductory statement is "factually wrong," and we again reject the characterization of our citation practice as "highly unethical."

The Reviewer has misread our claim. We did not state that prior studies have failed to examine the insula anatomically, functionally, or from a connectivity perspective. We stated that most prior studies have examined these dimensions in isolation, that is, separately rather than in combination. This is a statement about methodological integration, not about the existence or quality of prior work. It is accurate and defensible.

To be precise: prior anatomical studies have mapped cytoarchitecture without concurrent functional mapping. Prior functional stimulation studies have mapped evoked responses without concurrent connectivity analysis. Prior connectivity studies have examined insular networks without concurrent cytoarchitectonic parcellation or causal stimulation mapping. What our study contributes — and what no prior study has achieved - is the simultaneous integration of all three dimensions using causal electrophysiological measures. This is not a claim that prior researchers failed; it is a claim about what our multimodal design adds.

We have nonetheless revised this passage in the Introduction to make this distinction more explicit. The revised text reads:

"Previous studies have each provided important insights into the anatomy, function, and connectivity of the insula. Here, we build on and extend this foundational work by unifying these dimensions using a high-resolution and multimodal approach that enables an unprecedented characterization of its functional architecture using electrophysiological causal measures of its function and connectivity with not only the broader cortical mantle but also among its different subregions."

Regarding the ethics accusation: we have addressed this in our opening response and will not repeat that response here beyond noting that the claim remains, in our view, without foundation.

There are other factually incorrect or misleading comments throughout the manuscript, where the authors claim to have made a contribution, however, the data presented in the manuscript is not enough to support these claims. For example, the following sentence is again simply not true:

"In a cohort of 87 human participants implanted with intracranial electrodes, our current work aims to understand the fundamental organizing functional principles of insular cortex in the human brain in light of the new evidence about its microscopical anatomical architecture."

To unravel the fundamental organizing functional principles of insular cortex, insula mapping using higher order cognitive tasks such as attention, memory, etc, are necessary, which is not what the authors are showing in their manuscript.

Again, we respectfully disagree with this criticism on multiple grounds.

First, the reviewer imposes an arbitrary and unsupported definition of what constitutes "fundamental organizing functional principles." The claim that uncovering fundamental organizing principles of the insula requires mapping attention and memory is the Reviewer's personal theoretical preference, not an established methodological standard. There is no consensus in the field — and the reviewer cites none — that attention and memory tasks are necessary preconditions for characterizing insular functional organization.

Second, the reviewer appears to have overlooked substantial portions of our manuscript. Our study does not consist solely of stimulation mapping of sensory responses. It includes: Task-based electrophysiological recordings during the GradCPT, a validated paradigm for salience detection, sustained attention, and action mode selection — processes the reviewer themselves would classify as higher-order cognitive functions as well as bidirectional causal connectivity mapping between insular subregions and bidirectional causal connectivity mapping between insula and the cortical mantle. The claim that we are not examining higher-order cognition is therefore factually incorrect as applied to our actual manuscript.

Third, the word "aims" in the sentence that the reviewer quotes is doing important work that the reviewer has ignored. We wrote that our work aims to understand the fundamental organizing functional principles" — not that it definitively resolves them.

This is standard scientific language expressing the scope and motivation of a study. No single study, including those the reviewer apparently believes would be sufficient, could claim to have fully resolved the organizing principles of an entire cortical region. We make no such claim.

Finally, on the broader accusation of "factually incorrect or misleading comments": the reviewer makes this accusation sweepingly but supports it with only this single example, which as we

have shown above does not hold up to scrutiny. We ask the reviewer and the editor to note that strong language — "factually wrong," "simply not true," "highly unethical" - deployed repeatedly without adequate evidentiary support is not a substitute for scientific argument.

The paper is also riddled with major mathematical errors and impossible statistics. Specific comments are below.

We take any allegation of mathematical or statistical errors seriously and have carefully examined every specific technical point raised by the Reviewer. The statement that our paper is "riddled with major mathematical errors and impossible statistics" is an extraordinarily serious allegation that, if accurate, would indeed preclude publication. Having reviewed every specific point the reviewer raises in support of this claim, we find that not one of them identifies an actual mathematical error in our calculations.

In our view, a mathematical error is a miscalculation, an incorrect formula, a violation of statistical assumptions, or an invalid inference from a result. The reviewer identifies none of these. We ask the editor to weigh this carefully when evaluating the reviewer's overall characterization of our work.

Significant omission of prior literature:

(i) The authors talk about salience and higher order cognition, but the entire human insula intracranial EEG literature is omitted, see for example, work coming from the Nitin Tandon group or the Robert Knight group.

We have already addressed the broader literature omission concern in our response to the reviewer's opening comments. Here we respond specifically to the two examples cited.

Regarding the Tandon group, the Tandon lab's insula-related work is represented primarily by Woolnough et al. (2019, eLife), which maps insular activity during speech production, non-speech orofacial movements, and speech listening. This is a valuable paper, but its scientific content, speech sensorimotor processing, is entirely orthogonal to the focus of our study. Our manuscript did not address speech production.

Regarding the Knight group: Their insula-related work addresses reward prediction error coding, verbal working memory decision and response monitoring, and auditory deviance detection, all of which are valuable contributions to understanding the insula's role in cognitive processing, but again scientifically orthogonal to our study's focus.

We wish to make a broader point here. The reviewer invokes these two groups as examples of an "entire literature" we have omitted, yet neither group works on the questions our paper addresses. This suggests that the reviewer's citation demands are driven not by genuine assessment of scientific relevance but by other considerations.

We have added citations that Reviewers 1,2 and 3,4 had kindly alerted us about and have done our best to include citations throughout our revised manuscript where omissions were scientifically genuine and relevant.

(ii) Mapping subjective states in the insula using brain stimulation is not new and known for decades. See for example, Ostrowsky et al., Cerebral Cortex, also including authors' own previously published work, for example, Duong et al., Brain Stimulation.

This criticism is directly contradicted by our manuscript. We did not claim to be the first to map the insula using electrical stimulation. Our Introduction contains sixteen consecutive citations (references 15–30) documenting the prior stimulation literature in detail, including the full body of work from the Mazzola, Isnard, Mauguière, Ostrowsky, and related groups.

More specifically, both papers the Reviewer cites as omitted: Ostrowsky et al. (Cerebral Cortex) and Duong et al. (Brain Stimulation), were cited in our original submission as references 15 and 30 respectively.

Our novelty claim does not rest on priority in stimulation mapping. It rests on the simultaneous integration of stimulation-evoked phenomenology with a newly developed 16-area cytoarchitectonic atlas, task-based electrophysiology, and bidirectional causal connectivity mapping within a large cohort yielding a novel quadripartite functional architecture within the insula – which we hope will guide numerous future studies for the times to come.

(iii) Entire animal literature pertaining to the insula is also missing. For example, see Caruana et al., Current Biology.

We agree with the Reviewer that there is a large body of literature from animal studies on the insula. This literature spans decades, multiple species, and hundreds of publications covering taste, visceral function, pain, autonomic regulation, social behavior, and beyond. It is neither possible nor appropriate to cite this literature comprehensively in a focused human mapping paper. No published paper in this field, including those the reviewer cites approvingly elsewhere in their report, meets this standard. What is expected, and what we have done, is to cite animal work that directly informs our specific findings and interpretations.

(iv) The authors talk about insula stimulation and connectivity, however, the work coming from the lab of Philippe Ryvlin is missing:

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/hbm.22366>

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/hbm.22549>

We have added both papers to the revised manuscript. However, we wish to be precise about what these papers contribute and how they differ from our work, as the differences are substantial.

The two cited papers, Almashaikhi et al. (2014, Human Brain Mapping, hbm.22366 and hbm.22549), used CCEPs in 11 patients with 39 insular electrode pairs. Their connectivity analysis was organized according to the five insular gyri, finding differences in connectivity rate across gyri, with the posterior short gyrus showing greater connectivity than middle short and long gyri.

Our work differs from these studies in several major dimensions:

- 1) Scale: Our study includes 87 participants with 413 stimulation pairs and 751 recording pairs. This is not a minor quantitative difference; it is the difference between preliminary observations and statistically robust population-level conclusions.
- 2) Anatomical framework: The Ryvlin studies organized their findings according to the five macroscopic insular gyri, which we now show is a coarse anatomical scheme and thus not fully captures the functionally relevant microarchitectonic organization our study integrates.
- 3) Directionality: The Ryvlin lab examined connectivity rates without characterizing directional asymmetry. Our study's central finding is precisely the directional asymmetry of connectivity.
- 4) Integration: The Ryvlin lab examined connectivity in isolation. Our study integrates connectivity findings with several other independent metrics and experiments.
- 5) Lastly, but perhaps most importantly, CCEP methodology: The Ryvlin lab used a conventional CCEP approach based on evoked potential amplitude. We used an

extended methodology that quantifies the full spectrotemporal richness of evoked responses including inter-trial phase coherence, providing richer characterization of connectivity strength and directionality (Lyu et al Nature Neuroscience 2025).

In summary, the Ryvlin studies represent important early work that we now cite and build upon. They do not, however, diminish the novelty of our contribution, which advances beyond their findings in sample size, anatomical resolution, directional analysis, methodological sophistication, and multimodal integration.

(v) The entire cytoarchitectonic literature of the human insula using fMRI/sMRI/dMRI is also missing.

This criticism is again directly contradicted by the content of our manuscript. The cytoarchitectonic framework that forms the entire anatomical backbone of our study is built upon the Julich Brain Atlas - a comprehensive, quantitative, 3D cytoarchitectonic parcellation of the human brain developed by some of the co-authors of our current study.

Our manuscript cites the foundational papers underlying Julich atlas extensively, including Amunts & Zilles (2015), Amunts et al. (2020, Science), Kurth et al. (2010, Cerebral Cortex), and the insula-specific cytoarchitectonic papers from our own group (Quabs et al., 2022, NeuroImage; Quabs et al., 2025, Human Brain Mapping) — references 4, 34, 35, 36, 37, 38, 79, and 80 in our manuscript. These are not peripheral citations; they define the 16-area microstructural parcellation onto which all of our functional and connectivity findings are mapped.

Furthermore, the first author of our paper (J.Q.) is a specialist in the cytoarchitectonic organization of the human insula and a co-developer of the very atlas we use. The suggestion that we are unaware of or have failed to engage with the cytoarchitectonic literature is not only factually wrong — it is unfortunately the opposite of the truth.

Major mathematical errors:

(i) The use of common average referencing rather than bipolar referencing is wrong, see work coming from Robert Knight lab.

We reject the characterization of this as a "mathematical error." Common average referencing (CAR) is a widely used, well-validated approach in intracranial EEG analysis, employed in landmark studies across top tier journals. The choice between CAR and bipolar referencing involves genuine trade-offs that the field has debated extensively, and neither approach is simply "wrong."

Critically, the literature is clear about precisely when CAR becomes problematic: its known vulnerability is the introduction of spurious correlations in phase synchrony measures and coherence computation.

We performed none of these analyses. Our measures of inter-trial phase coherence computed relative to stimulation onset across trials and not across channels. None of these measures are susceptible to the cross-channel reference contamination that motivates the preference for bipolar referencing in phase-based functional connectivity analyses.

The reviewer's concern, while valid in its proper domain, is simply inapplicable to our methodology.

We have added the following brief justification of our choice of CAR to the Methods:

Common average referencing (CAR) is a widely used, well-validated approach in intracranial EEG analysis, but it may introduce spurious correlations in phase synchrony measures and coherence computation. It should be noted that our measures of inter-trial phase coherence

was computed relative to stimulation onset across trials and not across channels. None of these measures are susceptible to the cross-channel reference contamination that motivates the preference for bipolar referencing in some SEEG analyses.

(ii) "The remaining electrodes were balanced across the hemispheres (234 left, 244 right) and anterior-posterior (237 anterior, 241 posterior) subdivisions."

This is almost impossible. It is a near impossible task to get the same number of channels in both the hemispheres, since epilepsy predominantly affects the left hemisphere. Similarly, anterior insula will have more number of channels compared to posterior insula.

I have never seen a single insula paper published in literature which has the same number of electrodes in both the left and right hemisphere. Similarly, I have never seen a single insula paper published in literature which has the same number of electrodes in both the anterior and posterior insula.

Honestly speaking, it's really shocking that the right hemisphere has more electrodes (244) compared to the left hemisphere (234). It's similarly very very concerning that the posterior insula has more electrodes (241) compared to the anterior insula (237).

"Collectively, these distributions demonstrate balanced sampling across both hemispheres and anatomical subdivisions of the insular cortex."

This is hilarious because this never happens.

This also raises serious ethical concerns regarding whether the electrodes were actually implanted based on clinical, rather than, research purposes only. The authors mention that the electrode implantations were driven by clinical rather than research needs, however, based on the statistics that they show, it's now unclear to me whether proper ethics were followed related to electrode implantations.

We respectfully disagree with the characterization of our electrode counts as a "mathematical error." No miscalculation has occurred; these are straightforward counts of contacts across participants. The Reviewer's concern appears to reflect an expectation based on smaller cohorts rather than an error in our reporting. Comparable electrode distributions in 500+ patients have been reported by Dr. Avanzini (a co-author on this manuscript), providing independent precedent for our numbers.

Our Methods explicitly state that individual participants had either left-only (19 participants), right-only (10), or bilateral (21) coverage, with a mean hemispheric asymmetry of only 3.38 ± 3.97 electrodes per bilateral participant. Approximate balance at the group level is therefore the expected statistical outcome of aggregating across a large and clinically diverse cohort of 87 participants drawn from 15+ years of clinical epilepsy monitoring by our group. Epilepsy lateralization varies considerably across patients, temporal, frontal, parietal, and insular epilepsies affect both hemispheres, and our cohort reflects this diversity. The Reviewer's claim that right hemisphere coverage exceeding left is "really shocking" and that posterior exceeding anterior is "very very concerning" reflects an assumption about epilepsy lateralization that does not hold across a large mixed cohort.

The Reviewer then makes an extraordinary leap — from a statistical distribution they find unexpected to an allegation that our electrode implantations may not have been clinically motivated. We address this allegation here with the seriousness it warrants. Our Methods state explicitly: "Electrode implantation sites were determined solely based on clinical criteria by a

multidisciplinary team including specialists in epilepsy, neurosurgery, psychiatry, neuropsychology, and radiology. No electrodes were implanted solely for research purposes." All implantations were performed under Stanford IRB protocol and full documentation is available upon request to the editor.

Making a research ethics allegation against a group of clinical investigators on the basis of nothing more than electrode counts that the Reviewer personally finds surprising is, in our view, an irresponsible use of the peer review process.

We ask the editor to take note of this.

(iii) "In the stimulation experiments, we only included participants' subjective reports if they passed our strict quality control metrics: Reports were (1) evoked only during real, but not sham, stimulations, (2) reproducible when the same site was repeatedly stimulated, and (3) dose-dependent (i.e., effect intensity increased as stimulation intensity increased)."

I am sorry that these are not quality control measures, rather these are cherry-picking of massive proportion. The reason is that the human brain is highly heterogeneous and heterogeneous responses are expected when we apply identical stimuli to the brain, a very well-known phenomenon known for decades now. In the above so called "quality control" metrics, the authors are deliberately picking the things that are aligned with their hypothesis and throwing away the data that are not aligned with their hypothesis. Using this logic, literally any hypothesis can be proven to be correct by just choosing the trials/electrodes that are consistent with someone's hypothesis.

"(1) evoked only during real, but not sham, stimulations" – Reports need to be included regardless of whether there was a evoked response during real or sham stimulations. "(2) reproducible when the same site was repeatedly stimulated" – Reports need to be included regardless of whether the same response was reproducible or not. "(3) dose-dependent (i.e., effect intensity increased as stimulation intensity increased)" – Reports need to be included regardless of the effect intensity value as the stimulation intensity increased.

We strongly disagree with this characterization, which conflates methodological rigor with scientific misconduct. Our three criteria are not cherry-picking. They are the operational definition of a valid stimulus-evoked percept, and they are entirely standard in the human intracranial stimulation literature. Here is what each criterion does and why it is necessary:

Criterion 1 — Evoked only during real, not sham stimulation: Our protocol employed a blinded design in which sham trials (0 mA) were randomly interspersed with real stimulation trials, rendering the two conditions indistinguishable to participants. Any report that occurs with equal frequency during sham and real stimulation cannot, by definition, be attributed to the stimulation itself - it reflects a spontaneous event, a placebo response, or random noise. Treating such reports as stimulation-evoked percepts does not enhance scientific rigor; it fundamentally undermines it. To illustrate: if a patient presents with a headache, and subsequently reports a headache following 0 mA stimulation of the insula, there is no principled basis for counting that response as evidence of a stimulation-evoked effect as the patient's headache was present prior to the start of the stimulation.

Criterion 2 — Reproducible upon repeated stimulation: Reproducibility is a foundational principle of experimental science. A percept that occurs once and never again upon identical stimulation of the same site cannot be attributed to that site with any confidence. The reviewer argues that heterogeneous responses are expected and therefore irreproducible responses should be included. This misunderstands the distinction between biological variability, which we

fully acknowledge and which is captured in our response rate statistics, and the requirement that a claimed effect be reliably elicitable. We are not claiming that every stimulation of a site produces a response; we are claiming that when a site does reliably produce a response, that response reflects the functional properties of that site.

Criterion 3 — Dose-dependence: A dose-response relationship is one of the most fundamental criteria for establishing a causal relationship between a stimulus and an effect, recognized as such across all of biomedical science. A percept whose intensity does not decrease or increase with decreasing or increasing stimulation intensity is not demonstrably linked to the stimulation. Including such reports would conflate stimulation effects with spontaneous fluctuations in awareness.

The reviewer's argument — that including all reports regardless of their relationship to the stimulus would be more scientifically valid, would, if followed, make it impossible to distinguish signal from noise in any stimulation study. By the reviewer's logic, a placebo-controlled drug trial should include patients who responded to the placebo in the treatment group, because excluding them constitutes "cherry-picking." This is not a standard that any scientific field accepts.

We further note that our entire stimulation protocol was designed with multiple independent safeguards against bias: participants were blinded to stimulation parameters and site; no investigators were in the participant's line of sight; sham trials were indistinguishable from real trials; and transcripts were prepared by an external rater naive to study aims. These are the hallmarks of a rigorously controlled experiment, not of cherry-picking.

Finally, the reviewer's claim that "literally any hypothesis can be proven correct" using our approach is simply wrong. Our criteria are entirely agnostic to the content of responses as they apply equally regardless of what the participant reports. A visceral sensation, a pain response, or no response at all would be included or excluded by identical criteria. There is no mechanism by which these criteria could systematically favor one type of response over another.

(iv) "Subjective responses were evoked by electrical stimulation in 51% of insular sites, with a higher elicitation rate in the posterior (67%) compared to the anterior insula (32%)"

This is very surprising since I would expect higher elicitation rate in the anterior compared to the posterior insula because of the relevance of the anterior insula in many higher order brain functions.

This is not a mathematical error by any definition. The Reviewer offers no calculation, identifies no methodological flaw, and points to no error in our statistics. What the Reviewer expresses is surprise that our finding does not match their prior expectation. That is not a basis for labeling a result a "major mathematical error."

More importantly, the Reviewer's expectation is contradicted by a well-established principle in the human intracranial stimulation literature that they appear to be unaware of. Fox et al. (Nature Human Behaviour, 2020) - cited as reference 54 in our manuscript - demonstrated directly that the probability of eliciting subjective experiences with electrical cortical stimulation is inversely related to the degree of network complexity and distributional connectivity of the stimulated region. Focal regions with discrete sensorimotor or interoceptive representations yield high stimulation response rates precisely because their function is locally instantiated. Highly distributed transmodal association regions, such as the anterior polar insula with its broad prefrontal connectivity, yield low response rates because their computational contributions emerge from large-scale network interactions rather than focal cortical activity. Stimulating such

regions with local electrical current is analogous to pulling one thread in a complex distributed network: the network-level computation is not engaged.

We cite and discuss this principle explicitly in our manuscript: "In our prior work, we have demonstrated that the probability of eliciting subjective experiences with electrical stimulation decreases along a cortical gradient — highest in focal sensorimotor regions to lowest rates when highly distributed, integrative association networks are stimulated." Our finding that the anterior polar Region IV has the lowest stimulation response rate of all four insular regions is therefore not surprising - it is precisely what this established framework predicts, and it constitutes one of the most theoretically coherent and well-grounded findings in our paper.

The Reviewer's expectation that higher-order regions should show higher elicitation rates inverts this well-documented relationship. Had we found the opposite of what we report, that would have been the surprising result requiring explanation.

Similar issues remain from lines 214-273, where the authors have simply dumped a lot of statistics and response profiles, however, there is no effort to reconcile these findings to an integrated theory of the insula function. Again, similar issues from lines 274-307.

Again similar critiques apply to lines 309-344. There is a very solid literature when it comes to the salience detection function of the anterior insula. In addition to omitting a very large number of important references, the authors are again just showing some statistics without a clear theory or hypothesis. Just reporting the percentage of significant electrodes statistics is not very useful.

Similar critiques apply to lines 346 to 402. It's really impossible to make anything out of what the authors are presenting in the context of insular functions.

The Reviewer asserts that our Results section lacks theoretical integration and that "just reporting the percentage of significant electrodes is not very useful." We respectfully disagree. Reporting precise statistics is not a weakness, but a requirement of rigorous scientific communication. The Results section of an empirical paper is the appropriate place to present data; theoretical reconciliation is the function of the Discussion, where we do exactly what the Reviewer claims is absent.

(v) "Stimulation consisted of alternating square-wave pulses at 50 Hz frequency with 0.2 ms pulse width, delivered for 1–4 seconds per trial."

50 Hz stimulation frequency is typically applied when studying higher order brain functions such as attention and memory, since gamma frequency is more relevant for attention and memory, see work from Nanthia Suthana's lab. It's unclear what the rationale is to apply 50 Hz gamma frequency stimulation in the insula which is not a attention/memory brain region, and it's also unclear what the relevance of 50 Hz brain stimulation is when the authors are primarily studying lower order processes such as sensory processing.

The claim that 50 Hz stimulation is specifically reserved for attention and memory is factually incorrect and reflects a fundamental confusion between two entirely different uses of 50 Hz in neuroscience.

The reviewer appears to be conflating two distinct applications of this frequency:

1. High-frequency electrical stimulation for functional mapping (what we do): 50 Hz biphasic pulse trains delivered through intracranial electrodes to map cortical function by transiently

perturbing local neural activity. This is the standard frequency used in clinical and research cortical mapping worldwide since the time of Penfield.

2. Gamma-frequency oscillatory entrainment (what the reviewer is describing): A completely different paradigm in which rhythmic stimulation at gamma frequencies is used to entrain or modulate endogenous neural oscillations, studied in the context of memory and cognition by groups including Suthana's lab. This paradigm has entirely different parameters, goals, electrode configurations, and theoretical justifications from functional mapping stimulation.

(vi) "Electrodes were labeled as significant when the significant bin count exceeded 2."
This selection looks very arbitrary.

Here the Reviewer is referring to our somatosensory evoked potential analysis of the Italian cohort (led by co-author Dr. Avanzini from Milan) – This is the method they used to determine whether a given electrode showed a significant response to peripheral nerve stimulation.

The criterion "significant bin count exceeded 2" means that an electrode was classified as responsive only if more than 2 adjacent time bins out of the analyzed time-frequency bins (60) showed statistically significant gamma-band power increases following nerve stimulation. This is a threshold designed to avoid calling an electrode responsive based on a single spurious significant bin, and factually is intended at reducing the false positive ratio.

The reviewer calls this "arbitrary" but it is not. This criterion comes directly from the published methodology of Avanzini et al. (2016, PNAS; 2018, NeuroImage; 2021, Brain; 2025 Current Biology), with the manuscript testing somatotopy (Avanzini et al, 2018) cited as reference 40 in your manuscript. It is their method, published and peer-reviewed, and we applied it consistently.

(vii) "Then, common average re-referencing was applied, with channels excluded from the reference if they: (a) showed variance exceeding five times or falling below one-fifth of the cross-channel median variance; (b) contained spikes with a rate exceeding 1 spike per second, where spikes were defined as inter-sample voltage changes $\geq 80 \mu\text{V}$; (c) exhibited pathological activity with pathological high-frequency oscillations (Liu & Parvizi, 2019); or (d) displayed abnormal power spectral characteristics upon visual inspection."

This is again cherry-picking on a big scale. It's unclear how many electrodes/trials were excluded using these criteria.

We respectfully disagree with the Reviewer. This is not cherry-picking. These are standard preprocessing steps used universally in intracranial EEG research to remove pathological, artifactual, or electrically noisy channels from the common average before analysis. We did not selectively exclude trials or data from our results. We only constructed a "common average signal" without noisy and pathological channels. Majority of published intracranial EEG study, including those the reviewer cites approvingly elsewhere in their report, applied equivalent exclusion criteria. Including channels with pathological high-frequency oscillations, excessive spiking activity, or extreme variance would not make our analysis more rigorous; it would contaminate the reference common average signal. The characterization of principled artifact rejection as "cherry-picking on a big scale" misunderstands what cherry-picking means. Cherry-picking is the selective reporting of results that support a hypothesis while suppressing those that do not.

(viii) "We conducted independent-samples t-tests to compare the mean HFA power amplitude

between target and baseline trials within a specific temporal window of 400–1200 ms post-stimulus (relative to the onset of the image fade-in)."

Again, the choice of this window selection seems arbitrary.

The 400–1200 ms window was adopted directly from the established and published GradCPT analysis protocol because each stimulus starts fading at around 400ms. This method is identical to our previous publication (Kucyi et al. 2020, Nature Communications), cited as reference 61 in our manuscript, which specifically developed and validated the electrophysiological analysis approach for this task. Our Methods state explicitly that we followed "established GradCPT analytic protocols".

The reviewer's pattern of labeling every methodological parameter as "arbitrary" without identifying what an appropriate alternative would be or why does not constitute a scientific critique.

(ix) "Electrodes were classified as target-responsive if the difference in mean HFA power between these two conditions reached a two-tailed significance threshold of $p < 0.05$, after False Discovery Rate (FDR) correction was applied to account for the total number of electrodes within each subject."

Was this corrected for across subjects too since there are a relatively large number of subjects?

The reviewer has misunderstood our statistical design. In our analysis we computed a t-statistic for each electrode comparing HFA power between target and baseline trials, and FDR correction was applied across all electrodes within that subject. This is the appropriate unit of analysis because the question we are asking is whether a given electrode is responsive - a within-subject, electrode-level classification question. Critically, we did not perform a single omnibus test across all electrodes from all subjects simultaneously, which is what would require cross-subject FDR correction.

Moreover, the subsequent between-region comparisons where we ask whether response rates differ across insular regions were handled using Generalized Estimating Equations (GEE) with subject identity as the clustering variable, explicitly accounting for the non-independence of multiple electrodes within the same participant and for between-subject variability. This is a more sophisticated and appropriate approach than simple cross-subject FDR correction, as it models the hierarchical structure of the data directly.

(x) "Preprocessing included notch filtering at 60, 120, and 180 Hz, followed by exclusion of noisy channels and trials. Noisy channels were identified based on extreme raw amplitude (>5 SD), frequent spikes ($>3\times$ median spike rate), high mean absolute voltage (>4 SD), or elevated trial variance (>3.5 SD). Trials with mean amplitudes exceeding 4 standard deviations were also discarded."

This is again cherry-picking. How many channels and trials were excluded like this?

We have addressed the "cherry-picking" accusation in our response to point (vii) above and will not repeat that response in full here. Notch filtering at line noise frequencies, and exclusion of channels and trials with extreme amplitudes, excessive spiking, or pathological variance are not cherry-picking. They are universally applied, completely standard preprocessing steps in

intracranial EEG research. Every published intracranial EEG paper applies equivalent procedures. The reviewer's repeated application of the term "cherry-picking" to standard artifact rejection reveals a misunderstanding of what that term implies.

(xi) "To examine directional asymmetries across the full time-frequency spectrum beyond the early high-frequency F1 response, we calculated contrast spectrograms between paired electrodes for each regional comparison. For matched electrode pairs, power spectrograms were computed in each direction, averaged across pairs and trials, and subtracted to yield directional contrast maps (forward minus backward). Cluster-based permutation testing identified statistically significant spatiotemporal clusters. Candidate clusters were defined as contiguous time-frequency regions where point-wise t-tests across electrode pairs showed significant directional differences ($\alpha = 0.05$). The sum of t-values within each cluster was computed. Significance was assessed via 5,000 permutations where direction labels were randomly shuffled, contrast spectrograms recalculated, and maximum/minimum cluster t-sums recorded to form a null distribution. Observed clusters exceeding 95% of the null distribution were deemed significant."

This is mathematically wrong (Figure 5). To assess directional connectivity, please use a directional connectivity measure such as the Granger causality, see the Anil Seth Granger causality toolbox.

We reject the characterization that our approach is "mathematically wrong," and we must point out that the Reviewer's proposed alternative Granger causality is not applicable to CCEP data and ought not to be applied to it. Granger causality is a method designed for continuous, spontaneous time series data while CCEP data is evoked and contains the artifact of electrical stimulation and the range of frequencies affected by the evoked responses, both of which will interfere with the GC analysis and may result in spurious results. The mathematical assumptions underlying Granger causality are violated by design in CCEP recordings

In our CCEP-based directional connectivity analysis, directionality is established through the experimental design itself: we stimulate site A and record at site B, then stimulate site B and record at site A – within the same brain – and compare the resulting evoked responses in both directions. This bidirectional stimulation-recording design is the gold standard for causal connectivity assessment in human intracranial research, it is more directly causal than Granger causality, which merely infers temporal precedence from correlational structure in spontaneous signals. Our cluster-based permutation testing on directional contrast spectrograms then identifies which time-frequency components show statistically significant asymmetry between the two directions, a rigorous, nonparametric, assumption-free approach well suited to the characteristics of CCEP data.

Other major issues:

(i) "These findings reveal a four-fold structural-functional organization of the human insula based on electrically evoked response profiles of microstructural areas."

It's unclear what the relevance of four-fold here is. There is no theory to support only a four-fold structural-functional organization of the human insula.

The Reviewer has misunderstood the nature of our finding. The four-fold organization is not a theoretical claim, but it is an empirical result. We did not begin with a theory that the insula

should have four subdivisions and then design our analysis to confirm it. We applied unsupervised k-means clustering to our stimulation response profiles across 16 cytoarchitectonic areas and asked the data how many meaningful groups emerged. The answer was four.

Our manuscript is explicit about this. We state: "The optimal cluster number was determined by convergence of three complementary metrics: silhouette score, Davies-Bouldin score, and elbow method, all pointing to K=4." We further "acknowledge that the four-fold functional architecture of the insula described here is derived from recordings at hundreds of insular sites across several dozen participants... the four-cluster solution represents a heuristic rather than a hard truth, i.e., alternative solutions such as K=5 are mathematically plausible and cannot be formally excluded. The selection of K=4 reflects the best balance between mathematical optimality across all three metrics and the biological plausibility of the resulting clusters."

The Reviewer's objection that "there is no theory to support only a four-fold organization" is therefore not a critique of our analysis but it is a description of why the finding is novel.

(ii) "Stimulation of the posterior and dorsal insula reliably elicited pain and thermosensory percepts – as well as somatosensory changes as detailed above. However, we observed an intriguing meso-scale functional division among anatomically adjacent neuronal populations with distinct yet related roles: while stimulation of some sites evoked pain alone, stimulation of other sites within the same region produced purely thermosensory sensations, stimulation of a third set of sites elicited both warmth and pain, and stimulation of yet another set of sites elicited somatosensory sensations in the absence of pain or temperature percepts. It should be noted that the average electrical current eliciting pain and temperature was comparable with all other responses (pain and temperature: median 3 mA, IQR 2-4mA/ all other responses: median 4 mA, IQR 2-6) (Supplementary Figure 1). Stimulation at no site elicited a sensation of cold."

The authors are reporting a lot of results here. However, it's hard to reconcile these results in a coherent theory. Obviously, when we apply intracranial stimulation, there will always be some channels which will respond more compared to some other channels, however, without a tight link to the anatomy or theory of the insula it's unclear how these findings are useful. These results are highly variable, so what do all these results mean in the context of the insula theory?

The Reviewer acknowledges that we are "reporting a lot of results" but objects that they are difficult to reconcile with a coherent theory. We respectfully disagree on both counts.

First, our findings are explicitly linked to anatomy. The functional divisions we describe, i.e., pain alone, thermosensory alone, combined warmth and pain, and somatosensory without pain or temperature, are mapped onto the 16-area cytoarchitectonic atlas of the insula and shown to correspond to distinct microstructural regions within the granular posterior insula. This is precisely the "tight link to anatomy" the reviewer claims is absent.

Second, our findings are theoretically meaningful and directly relevant to longstanding debates in the pain and interoception literature. The dissociation between pain and thermosensory responses at adjacent stimulation sites speaks to the question of whether nociceptive and thermoceptive signals share a common cortical substrate or are processed in partially segregated populations, a question with direct relevance to the neuroscience of pain and to clinical targets for neuromodulation. The observation that no site elicited cold sensation is itself a theoretically significant negative finding, consistent with the known asymmetry between heat and cold processing in the insular cortex.

Third, the reviewer's argument that variability in stimulation responses renders findings uninterpretable would, if accepted, invalidate the entire human intracranial stimulation literature.

Functional heterogeneity at the meso-scale is not noise, it is signal. Documenting it precisely, as we have done, is the contribution.

(iii) There is also no control region where brain stimulation is applied. So it's unclear whether the effects are also present brain-wide or specifically localized to the insula. If stimulating other brain areas also evokes similar response profiles to the insula, then the authors cannot make causal claims specifically related to the insula.

We address this on two levels.

First, our prior publications provide directly relevant whole-brain control data. For instance, Selimbeyoglu & Parvizi (2010) comprehensively summarized approximately 100 years of published electrical stimulation data across all cortical and subcortical regions, providing a brain-wide phenomenological reference map against which insular responses can be contextualized. Fox et al. (Nature Human Behaviour, 2020) systematically reported stimulation-evoked observations across the entire cortical mantle in overlapping patient cohorts using identical stimulation parameters and questioning procedures. Critically, stimulation of prefrontal, orbitofrontal, cingulate, and other regions in our hands yielded qualitatively and categorically distinct phenomenological profiles from anything we encountered with insular stimulation. This rules out a methodological floor effect and confirms that the restricted somatovisceral profile of insular responses reflects genuine functional specificity rather than a limitation of our approach.

Second, our dataset contains a powerful internal control structure that is in some respects more rigorous than a single external control region. Each insular cytoarchitectonic region served as an anatomical control for the others — providing three independent internal comparisons rather than one. Posterior granular regions reliably evoked somatosensory and nociceptive experiences; mid-anterior regions evoked visceral and anxious sensations; and the anterior polar region was largely silent. This cross-subject consistency of regionally specific and categorically distinct response profiles constitutes strong evidence of genuine functional specialization. Had stimulation produced non-specific effects, one would expect either uniform responses across all subregions or random variability across participants — neither of which we observed.

(iv) "DATA AVAILABILITY"

We have mapped the four insular regions in MNI152 ICBM 2009c Nonlinear Asymmetric space and will make it publicly available for download upon publication of our work.

Electrophysiological connectivity data from our study along with the exact coordinates of the stimulated or recorded sites in MNI space will also be shared upon publication."

It is suggested to make the raw unprocessed data available for ethical reasons and reproducibility. Holding on to data and without making it available for other researchers hinders the progress of science and reproducibility. These days, data can be easily anonymized and made publicly available, see publicly available datasets from the Eddy Chang's group at UCSF or Rob Knight's group at UC Berkeley.

As we have replied to Reviewer 3 and 4, we are pleased to confirm that MNI-space coordinates of stimulation sites and associated evoked sensations will be included in our data sharing. We agree that these maps will substantially facilitate quantitative comparisons with other datasets, including fMRI studies, and we are glad to make them available to the community as part of our data release.

(v) Lyu, D. et al. Mapping human thalamocortical connectivity with electrical stimulation and recording. Nature neuroscience 28 (2025). <https://doi.org/10.1038/s41593-025-02009-x>

The above reference is cited twice, once as reference 44 and then again as reference 72. I do not understand this at all.

We thank the Reviewer for identifying this and have corrected it in the revised manuscript.

Concerns on figures:

(i) The insula in Figure 1A right panel and Figure 1B, and in all the subsequent figures looks very very weird and strange. I have never seen an insula structure like that. Something is going very wrong in the registration algorithms.

What the reviewer finds unfamiliar is the flat map representation of the insular cortex, a format that may appear unusual at first glance precisely because it differs from standard three-dimensional brain renderings. This representation is not a registration artifact; it is a deliberate and scientifically motivated visualization choice. The insular outline in our figures follows the cytoarchitectonic surface reconstruction published in Quabs et al. (2022, NeuroImage) and is part of the Jubrain/EBRAINS platform developed within the Human Brain Project — a publicly available, peer-reviewed, and widely used resource in the human brain mapping community. The reviewer's concern that "something is going very wrong in the registration algorithms" is based solely on visual unfamiliarity with this format and does not identify any specific anatomical or technical error.

(ii) It's unclear why there are a large number of non-responsive electrodes in the anterior insula subdivisions during the salience task in Figure 4.

The non-responsive electrodes visible in the anterior insula subdivisions during the GradCPT task in Figure 4 simply reflect electrodes that did not reach our predetermined statistical significance threshold of $p < 0.05$ after FDR correction. This is expected and entirely unremarkable. Not every electrode within a functionally active region will show statistically significant responses, and this reflects normal biological variability across recording sites, differences in electrode placement relative to the most active neuronal populations, and the statistical threshold we applied. The presence of non-significant electrodes within an otherwise active region is a feature of any honest thresholded statistical map, not an anomaly requiring special explanation.

(iii) Figure 5 is mathematically wrong like I mentioned above.

The Reviewer repeats their earlier assertion that Figure 5 is "mathematically wrong" and refers back to their recommendation to use Granger causality. We have addressed this claim comprehensively in our response above.

References

- Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annu Rev Psychol*, 58, 373–403. doi:10.1146/annurev.psych.58.110405.085709
- Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nat Rev Neurosci*, 16(7), 419–429. doi:10.1038/nrn3950
- Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, 100(2), 490–509. doi:10.1016/j.neuron.2018.10.002
- Block, N. (1990). Consciousness and accessibility. *Behavioral Brain Sciences*, 13, 596–598.
- Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292), 1464–1468. doi:10.1126/science.aaf0941
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
- Koban, L., Gianaros, P. J., Kober, H., & Wager, T. D. (2021). The self in context: brain systems linking mental and physical health. *Nat Rev Neurosci*, 22(5), 309–322. doi:10.1038/s41583-021-00446-8
- Lyu, D., Stieger, J. R., Xin, C., Ma, E., Lusk, Z., Aparicio, M. K., . . . Parvizi, J. (2023). Causal evidence for the processing of bodily self in the anterior precuneus. *Neuron*, 111(16), 2502–2512 e2504. doi:10.1016/j.neuron.2023.05.013
- Parvizi, J., Rangarajan, V., Shirer, W. R., Desai, N., & Greicius, M. D. (2013). The will to persevere induced by electrical stimulation of the human cingulate gyrus. *Neuron*, 80(6), 1359–1367. doi:10.1016/j.neuron.2013.10.057
- Parvizi, J., Veit, M. J., Barbosa, D. A. N., Kucyi, A., Perry, C., Parker, J. J., . . . Halpern, C. H. (2022). Complex negative emotions induced by electrical stimulation of the human hypothalamus. *Brain Stimul*, 15(3), 615–623. doi:10.1016/j.brs.2022.04.008
- Paulus, M. P., & Stein, M. B. (2006). An insular view of anxiety. *Biological Psychiatry*, 60(4), 383–387. doi:10.1016/j.biopsych.2006.03.042
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, 8(10), 1391–1400. doi:10.1038/nrn1549
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565–573. doi:10.1016/j.tics.2013.09.007
- Siddiqi, S. H., Kording, K. P., Parvizi, J., & Fox, M. D. (2022). Causal mapping of human brain function. *Nat Rev Neurosci*, 23(6), 361–375. doi:10.1038/s41583-022-00583-8
- Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, 13(8), 334–340. doi:10.1016/j.tics.2009.05.001
- Tan, L. L., & Kuner, R. (2021). Neocortical circuits in pain and pain relief. *Nat Rev Neurosci*, 22(8), 458–471. doi:10.1038/s41583-021-00468-2

Reply to Co-Referees 1 and 2

Summary: I was generally convinced by the authors' response and the revised manuscript. The authors have taken the prior comments seriously and, in my view, have substantially improved the paper accordingly. The revision is stronger in its framing, clearer in several of its methodological explanations, and more appropriately calibrated in its interpretation of the findings. At this stage, I only have a small number of remaining comments, all of which are relatively minor and primarily concern conceptual precision, clarification of cohort structure, statistical wording, and figure labeling.

We are deeply grateful to the Referee for their thorough and generous assessment of our revised manuscript. We have now addressed the remaining minor points in detail and hope that the Referees will find our replies satisfactory.

Minor comments

1. The historical framing of interoception in the Introduction should be corrected

The Introduction currently attributes to Sherrington a definition of interoception as “the sensing of internal bodily states, distinct from exteroception and proprioception.” This is not quite accurate historically. Sherrington's contribution was to distinguish classes of physiological stimulation according to whether they arose from outside the body, within the body, or from the musculoskeletal apparatus; that is, he classified stimuli as exteroceptive, interoceptive, and proprioceptive. The more explicit formulation of interoception as the sensing of internal bodily states is a more contemporary conceptualization rather than Sherrington's original definition.

I would recommend revising this sentence for historical accuracy and adding one or more appropriate contemporary references to support the modern definition being used in the manuscript.

We thank the Referee for this important historical correction. After re-reading the work by Sherrington, we now understand where we made a historical mistake. The Referee is entirely right that our original phrasing misattributed the modern conceptualization of interoception to Sherrington. We have revised the relevant text in the Introduction as follows:

Sherrington, over a century ago, introduced the concept of “intero-ceptive” field (i.e., the sensory field within the body) as opposed to exteroceptive (outside the body) and proprioceptive (the musculoskeletal apparatus)(Sherrington, 1906). The more explicit formulation of “interoception” as the brain's sensing and representation of internal bodily states evolved gradually as a contemporary conceptualization that was refined and expanded in subsequent years (Craig, 1996; Critchley et al, 2004; Seth, 2013).

2. The description of cohort structure is improved, but one point remains unclear

The revised manuscript is clearer than the prior version in detailing the composition of the different data subsets and the degree of overlap between analyses. That said, I am still somewhat confused about the distribution of the Stanford and Italian cohorts in relation to the Somatosensory Evoked Potentials analysis. It would help to make this point fully explicit, ideally in one concise sentence or a simple schematic/table indicating which cohort contributed to which

analysis. This is now a relatively minor clarity issue, but resolving it would make the workflow easier to follow.

We have now added the following statement in the Methods.

We note that our SEP analysis was based entirely on the Italian cohort and does not overlap with any other component of the study. SEP recordings were not part of the protocol at Stanford and were therefore not available for the American participants.

3. The wording of the statistical approach for the Saliency Task should be clarified. Regarding the Saliency Task, the manuscript refers to independent-samples t-tests conducted within each participant. This surprised me, since measurements obtained within the same participant are ordinarily treated as dependent rather than independent. I may be misunderstanding what the authors intended here, but the current wording is potentially confusing. It would therefore be helpful for the authors to clarify exactly what comparison was performed and why this statistical formulation is appropriate. If the analysis was in fact based on within-subject comparisons, the terminology should be revised accordingly.

We thank the Reviewer for this observation and apologize for the confusing terminology. We have added the following clarifying sentence to the Methods to make this explicit and to prevent further confusion.

To clarify, the independent samples t-test was applied to compare HFA power values from individual mountain target trials against HFA power values from individual city baseline trials within each participant. Because mountain and city trials were distinct, non-overlapping events with no natural one-to-one pairing between them, we considered the observations across the two conditions as independent at the trial level, making the independent samples t-test the appropriate choice. A paired t-test would require a meaningful one-to-one correspondence between individual trials across conditions, which did not exist in this experimental design.

4. At least some figure elements may still require checking

Some figures may still need another careful review for labeling clarity. For example, in Panel A of Figure 1, the “responsive” group does not seem to correspond transparently to a visible portion of the bar plots. In particular, the meaning of the grey segment is not immediately clear. The authors should clarify this in the figure and/or legend.

We again thank the Reviewers for this careful observation. We have now modified Figure 1a and have revised the labels and figure legends.

Reply to Reviewer 3

I have reviewed the responses to all of the prior reviews and revisions to the manuscript. I believe this is an important paper that substantially advances the field, and I support publication in Nature. The authors have done a thoughtful and thorough job in responding to all of the comments raised.

The combination of techniques and findings presented here, which constitute over a decade of work collecting data from these rare patients, provide something qualitatively new that builds on previous work, but also adds new insight. The CCEP findings are a good example of novel and important new findings, as they show directional connectivity between the frontal cortex (and other cortical areas) and insula, and within the insula, that tells us something new about how 'top-down' influences and predictive processing works. In addition, the paper provides fascinating, foundational data on the types of feelings, evoked by insular stimulation, which are surprisingly common but also surprisingly limited to particular classes of interoceptive, somatosensory, and pain-related sensations. The combined findings on anterior insular area 4 — that it is devoid of evoked feelings, projects back to influence posterior areas 1 and 2, and receives substantial frontal projections — provides substantive new data that challenges and updates now-classic views of anterior insular function and will lead to new conceptual models. These data will also be shared, which will be a boon for the field. In short, the authors have satisfied the concerns raised.

We are deeply grateful to the Referees for this generous and thoughtful assessment of our work.

References

- Craig, A. D. (1996). An ascending general homeostatic afferent pathway originating in lamina I. [Review] [78 refs]. *Progress in Brain Research*, 107, 225–242.
- Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7(2), 189–195.
doi:10.1038/nn1176
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565–573. doi:10.1016/j.tics.2013.09.007
- Sherrington, C. S. (1906). *The Integrative Action of the Nervous System*. NEW HAVEN: YALE UNIVERSITY PRESS.