How do brain maps affect novices in a simplified scientific investigation task?

Marina Dubova (mdubova@iu.edu)
Cognitive Science Program
Bloomington, IN 47405 USA

Robert L. Goldstone (rgoldsto@iu.edu)
Cognitive Science Program
Bloomington, IN 47405 USA

Abstract

This study explores how scientific conceptualizations, such as partitioning of the brain into distinct regions, shape investigation. One hundred fifty-six undergraduate psychology students (novices) completed a science learning task in which they explored the behavioral functions of a fictional brain segment by conducting simplified neuroimaging and lesioning experiments on it. We investigated how the partitioning of the segment into regions influenced participants' experimental choices and learning outcomes by randomly seeding the brain regions for each participant. The participants exhibited conceptual influences on their experimentation: they preferred to explore the boundaries and prototypical–or "skeletal"–locations of the delineated regions. These conceptual biases significantly shaped learning outcomes; for example, participants were more successful at identifying signals near region boundaries. Additionally, participants demonstrated conceptual expectations that led them to associate a discovered signal with locations within one region rather than locations that straddled region boundaries. This research contributes to our understanding of how the scientific concepts affect scientific investigation.

Keywords: experimentation strategies; scientific concepts; ontologies; neuroimaging; scientific reasoning; brain mapping

Introduction

While science is often regarded as objective, it is increasingly recognized that scientific evidence is not a neutral reflection of reality. Instead, it involves a contextualized interpretation of phenomena, influenced by scientists' social, historical, and cognitive backgrounds (Longino, 2002; Daston & Galison, 2021; Chang, 2022). Scientific theories and concepts–such as the periodic table in chemistry, the taxonomy of species in biology, or the DSM manual of mental disorders in psychiatry–influence how scientists measure, analyze, interpret, and communicate the phenomena they study (Dubova & Goldstone, 2023). For example, scientists tend to interpret and communicate their results in the ways that support their conceptualizations, and develop specialized instruments to measure entities indicated by their concepts (Bloch, 2012; Chang, 2004). This concept-laden nature of evidence poses challenges in reassessing a field's theoretical foundations based on accumulated evidence, such as searching for better classifications of mental disorders (Cuthbert & Insel, 2013; Kotov et al., 2017). Therefore, ensuring steady scientific progress requires empirical understanding of the specific ways in which concepts influence evidence accumulation in a given scientific discipline.

Examining scientists' investigation strategies could involve engaging participants with systems whose internal workings are already understood, such as microprocessors (Jonas & Kording, 2017), radios (Lazebnik, 2002), or simulated systems, such as microworlds with different physical properties (Bramley, Gerstenberg, Tenenbaum, & Gureckis, 2018; Ullman, Stuhlmüller, Goodman, & Tenenbaum, 2018). Exploring the simulated systems provides an opportunity to gain insights onto the general or specific assumptions that might influence the scientific process, by designing systems that either adhere to or deviate from these assumptions. However, a significant drawback of utilizing such simplified, microworld scenarios is their potential lack of resemblance to the actual real-worlds systems scientists study and a limited ability to replicate the exact scientific practices. Therefore, any effects observed in such studies could reflect cognitive strategies or guiding assumptions which may or may not play a role in the actual scientific process.

Here, we present a scientific reasoning study with psychology undergraduate students. We investigate how minimal conceptual assumptions about a fictional brain segment—such as the way it is divided into different regions—affect how participants conduct experiments to learn about this segment. In the experiment, participants were playing the role of a neuroscientist studying the fictional brain segment through imaging and lesioning experiments. Each participant aimed to learn as much as possible about the brain-behavioral associations of the neural segment assigned to them by conducting imaging and lesioning experiments on it. For each participant, the segment map was randomly generated to manipulate how their fictional brain segment was divided into regions (Baribault et al., 2018; Dubova, Moskvichev, & Zollman, 2022).

Importantly, our study is not testing whether the *actual* brain maps are used or are useful in neuroscience. Instead, we are merely probing whether and how partitioning a brain into regions affects novices' exploration of it.

Method

Participants

156 undergraduate psychology students from Indiana University took part in the study for course credit. The participants remotely completed the experiment on their own laptops.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Stimuli

Each participant explored one fictional brain segment. The brain segment was a square-shaped patch that had four different areas indicated with different colors. Participants were told that the fictional brain segment belonged to a fictional subject on which they could conduct experiments. To avoid confusions, we use 'participants' to refer to the people performing our experiment, and 'subjects' to refer to the fictional people that our participants were tasked with studying.

Participants had five types of tasks which they could give their subjects: "Visual", "Auditory", "Olfactory", "Balance", and "Tactile". For each participant, the partitioning of their subject's brain segment into areas ("regions") and the ground truth connecting brain activation to behavioral outcomes ("signals") were randomly created when they started the experiment (Yarkoni, 2022; Baribault et al., 2018; Dubova et al., 2022; Musslick et al., 2023).

Ground Truth The ground truth that participants had to uncover was created using a simple mapping—there was a location in the brain responsible for each behavioral function. For each participant, the x and y coordinates of four locations on the brain segment were randomly selected[1] to serve as signals responsible for the subject's brain activation and behavioral performance. Each type of task was randomly assigned to one signal; one signal corresponded to two tasks.

Brain Segment Map The partitioning of the brain segment into regions was also randomized across participants. For this, four more locations on the brain segment were chosen by randomly generating their x and y coordinates. These four locations served as the seeds for the brain segments' regions. The regions were generated by creating a Voronoi parcellation using the produced seeds. We used Minkowski distance with a power of 6 as a distance formula for the parcellation[2].

Procedure

The experiment consisted of a short demographic survey, instructions, training, and testing phases. At the end, participants had a debriefing revealing the goals and main idea of the experiment. See the demo of the procedure.

Instructions

The instructions introduced participants to the experiment and the different types of trials they would be encountering. We introduced participants to their goal in the following way: "In this experiment, you will be exploring the function of a small segment of the brain. Your objective is to understand the relationship between this brain region and specific behaviors or activities". We introduced participants to the fact that the segment was split into colored regions in the following way: "To help you in your investigation, the brain segment visualizations will be divided into four distinct regions, each filled with a different color. These colors depict the regions characterized by different types of neural transmitters used by the cells, as documented by the scientific literature. That is, each of the four colored regions has its own distinct distribution of neural transmitters. Note that these previously documented regions may or may not be related to the cognitive functions you are trying to investigate." Then, we instructed participants as to the structure and goals of each type of trial, letting them practice one example trial of each type.

[1]With a constraint that each signal has to be at least 25 pixels (Euclidean) away from every other signal

[2]This formula was chosen by trial and error to produce the most natural-looking regions.

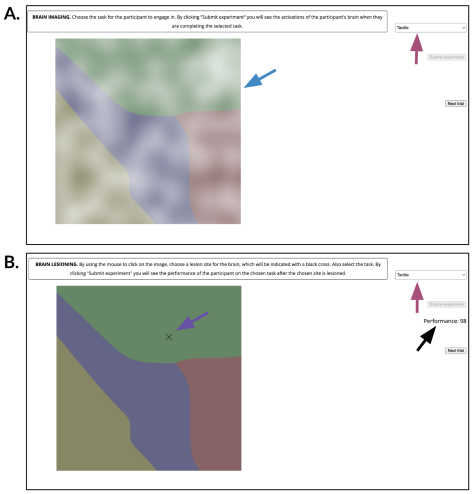

Figure 1: A. Example brain imaging training trial: the magenta arrow indicates a prompt for choosing the task type (chosen task: "tactile"), the blue arrow shows the resulting brain segment activity plot. B. Example brain lesioning training trial: the magenta arrow indicates a prompt for choosing the task type (chosen task: "tactile"), the purple arrow shows the lesion location, the black arrow shows the resulting performance of the subject on the chosen task after the lesion.

Types of trials

Training consisted of imaging and lesioning trials (Figure 1). The testing involved predict-from-image and lesion prediction trials with and without the colored map of the segment.

Neuroimaging During neuroimaging trials, participants selected the properties of the task for their subject to complete and observed the subject's task-induced brain activity plot (Figure 1A). The plot was created by summing a Gaussian distribution centered on the task's associated signal (variance=50) with Perlin noise of amplitude 0.3 at three spatial scales. Perlin noise was used to make the neural activity maps appear more natural and to make the detection of the signals more challenging.

Lesioning On lesioning trials, participants saw the brain segment as a square patch with colored regions. Participants chose the type of task for their subject to engage in. Then, they chose a location on the segment to inhibit by mouse-clicking on it. Participants observed their subject's perfor-

mance from 0 (minimum) to 100 (maximum) – on the chosen task after the chosen segment's location was deactivated (Figure 1B). The closer the placed lesion was to the signal, the worse the performance of the subject on the chosen task. Performance was calculated by measuring the distance between the signal and the lesion.

Predict-from-image On predict-from-image trials, participants saw the image of the segment's activity and had to choose one of five task types that they thought was associated with the provided image.

Lesion prediction On lesion prediction trials, participants saw the type of task that their fictional subject was completing and had to choose the segment's location which they thought would impair their subject's performance the most.

Training

The training phase includes 50 imaging and 50 lesioning trials. The imaging and lesioning trials were interleaved, so that each imaging trial was followed by a lesioning trial. At each point of the training phase, one of four (randomly selected) testing trials was added with a 20% probability. The testing trials incorporated into the training phase showed participants the correct answers after they submitted the guesses.

Testing

The testing phase consisted of 60 randomly shuffled testing trials. Each participant was tested on predict-from-image and lesion prediction trials with and without the colored regions for each type of task. The task type and the brain activation image were randomly selected. Each unique testing trial was repeated three times throughout the testing phase. The testing trials provided no feedback.

Results

Each participant worked with their own unique brain segment; therefore, all our analyses involve customized, relative metrics. All of the analyses are exploratory and do not test a specific theory (Scheel, Tiokhin, Isager, & Lakens, 2021). The analyses reported in different subsections test different hypotheses that are not theoretically exchangeable (and, again, are exploratory), and so they do not entail family-wise inflation of type-1 error (Wilson, 1962; Rubin, 2021).

For all analyses, we split the lesioning data into *exploratory* (far from the probed signal) and *exploitative* (close to the probed signal) trials using a data-driven threshold based on the bimodal distribution of lesion distances from the signals.

Participants preferred to explore regions' boundaries

Some of the most perceptually noticeable aspects of the brain segment images are the boundaries between the regions. The edges between visual regions not only attract attention but are also efficiently processed by specialized neural mechanisms, such as the second-order difference detectors (Marr & Poggio, 1976), which emphasize the transitions between different visual regions. The perceptual importance of the object boundaries is underscored by findings like those of (Davidenko, 2007), who demonstrated that humans can recognize faces from mere silhouettes. Despite their visibility, the boundaries might be considered as the least useful targets for experiments because they are only peripherally, not centrally, related to conceptually-defined regions. Therefore, we wondered whether participants would be influenced by such boundaries in their experimentation–for example, by preferentially testing the boundaries between the regions or by instead avoiding these boundaries when exploring the cognitive functions of regions of the brain segment. To test this, we computed the *perceptual boringness* score for each location of each participant's brain map (Figure 2). *Perceptual boringness* represents how close a location is to the boundaries between two, three, and all four regions. We formalized *perceptual boringness* as a sum of the absolute differences between the Minkowski (r=6) distances of the pixel to its closest, second closest, and least closest region seeds, preferentially weighting the distances to the closer seeds. If a location is equally close to multiple region seeds, then this location is right at the boundary between these seeds. Therefore, a lower boringness score indicates that a location is near a boundary or several boundaries; a higher boringness score of the location indicates that it is far away from the boundaries.

Figure 2: Left: example partitioning of the brain segment into regions investigated by one of the participants. Middle: the heatmap of boringness scores of this segment (darker–higher boringness). Right: the heatmap of skeleton scores of this segment (lighter–higher skeleton score).

For each participant, we computed mean boringness score for the locations of the exploratory (far from the signals) lesions they placed during the training phase. As a baseline, we simulated the same number of random lesion locations and calculated their respective mean boringness score for each participant's map. A paired Wilcoxon test revealed a significant difference between the mean boringness scores of the empirical and simulated data $V = 4607, p = 0.01$), indicating that participants' lesion placement was significantly closer to the boundaries between the regions than the random baseline (Figure 3).

Participants preferred to explore prototypical locations (skeletons)

Brain regions have prototypical locations, such as their centers, which might also attract experimentation. Recent work on human visual perception has identified that when simply

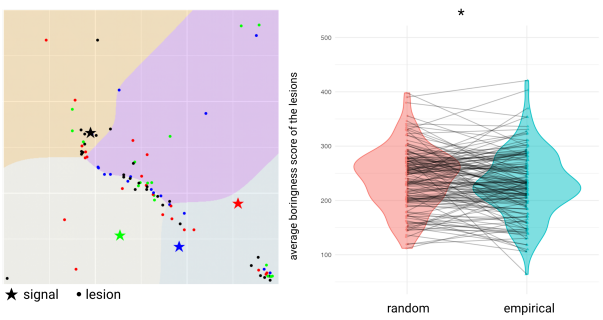

Figure 3: Left: Example participant who was influenced by the boundaries. Filled regions indicate the regions of the subject-specific brain maps; stars indicate the underlying signal locations (not visible to participants), and the dots show chosen lesion locations throughout the whole experiment. Right: Distribution of boringness scores as compared to random baseline.

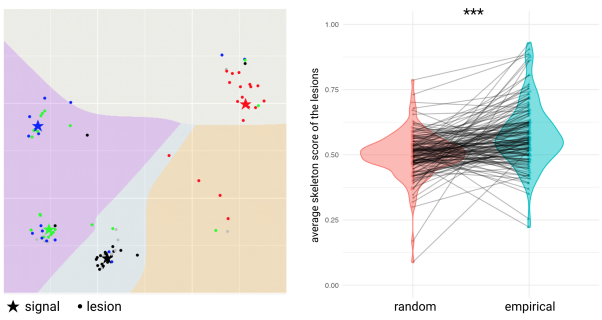

Figure 4: Left: Example participant who preferred to explore the prototypical locations of the regions. Filled regions indicate the regions of the subject-specific brain maps; stars indicate the underlying signal locations (not visible to participants), and the dots show chosen lesion locations throughout the whole experiment. Right: Distribution of skeleton scores of the lesions compared to random baseline.

perceiving shapes humans are attracted to their "skeletons"–which represent medial axes capturing the shape structures invariant to bending, stretching, and other deformations (Firestone & Scholl, 2014). We hypothesized that regions' skeletons might serve as prototypical axes that would attract participants' experimentation, either because they attract participants' attention or because of the assumption that these more prototypical locations will lead to more informative results. We mapped how "skeletal" each point on the region is by computing its normalized Euclidean distance from the nearest boundary (edges included) within its respective region (Figure 2). In this way, the skeletons are locally most distant from the boundaries, while all the other points are counted as less skeletal the closer they are to the boundaries. Then, we computed the average skeleton scores of exploratory lesion experiments chosen by participants. As a baseline, we simulated the same amount of random lesion locations per each participant and calculated their respective skeleton score. Note that the skeleton score is not the the same as perceptual boringness–boringness is high for the pixels in the corners and outer edges of the brain segment image, whereas skeleton score treats every region as an individual unit and considers these outer edges as having low score.

We compared the average skeleton scores of participants' exploratory lesions and random lesions with a paired Wilcoxon test. The test revealed a significant difference $V = 9460, p < 0.001$), suggesting that novices' exploratory lesion placement was closer to the skeletons of the regions than the random baseline (Figure 4).

Participants overexplored smaller regions

Brain regions do not only have locations and boundaries, but also vary in size. Here, we test whether participants tend to disproportionately explore the small (or large) regions, perceiving size as an indication of functional importance. For this, we compared the proportion of lesions participants placed in each region during the training phase against a baseline in which testing proportions mirror each region's size proportion (Fretwell & Calver, 1969; Stephens & Krebs, 1986). According to this baseline, a region occupying 40% of the brain segment should receive 40% of lesions, while one with 10% area should receive 10% of lesions.

Our first analysis involved calculating the residual between the actual proportion of exploratory lesions in a region and the region's size proportion. We then used linear analysis to predict these residuals for each participant based on the size proportion. The direction of the size proportion effect coefficients indicates participant biases: positive coefficients suggest a tendency to undersample smaller regions and oversample larger ones, and vice versa for negative coefficients. The between-participant coefficient distribution was unimodal with the negative median, supported by a Wilcoxon signed rank test indicating that the median coefficient was significantly less than zero ($V = 4711$, $p = 0.01$).

Subsequently, we examined the ratios of size proportions vs. exploratory lesion proportions. Here, coefficients greater than 1 indicate a preference for lesioning larger areas more and smaller areas less, while coefficients less than 1 suggest the opposite trend. Most participants had coefficients under 1, and the overall distribution of coefficients was unimodal. However, a Wilcoxon signed rank test did not find a statistically significant difference to suggest that the median coefficient value was less than 1 ($V = 5352$, $p = 0.13$), indicating that the preference towards lesioning smaller areas more frequently may not be very pronounced.

Once they have found a signal, participants assume it is restricted to one region

Next, we were interested in whether participants, upon detecting a signal, showed a tendency to confine their subsequent lesioning of this signal to the same region (Figure 5). Such tendency would indicate an assumption that signals are

restricted to one region only. For this, we filtered our data to consider only the exploitative trials where the lesions were sufficiently close to the probed signal. For each relevant lesion, we established a baseline – the proportion of hypothetical lesions within the same Euclidean distance from the signal, that would fall within the signal's region. This baseline helps us discern if participants exhibit a bias towards lesioning within or outside the signal's region, beyond the simple likelihood of targeting any nearby region. In scenarios where a participant's lesion was extremely close to the signal (e.g., 99% of all hypothetical lesions within that distance would also be in the same region), the lesion's location within the same region as the signal is statistically unremarkable. Conversely, if the lesion was in a different region than the signal, this would strongly suggest a preference for lesioning outside the signal's region. We excluded lesions from this analysis where the baseline was 1 (all hypothetical lesions within the same distance are in the same region as the signal) or 0 (none are in the same region), as these cases are not diagnostic data for our analysis. We focused on two key variables: 1) *sameRegion* – a boolean indicating whether the lesion is in the same region as the signal (0 for false, 1 for true), and 2) *proportion* – the proportion of pixels, within the same distance from the signal, that are in the same region as the signal (ranging from 0 to 1). For each lesioning trial, we calculated the residual $sameRegion - proportion$. A positive residual suggests a preference for lesioning within the same region as the signal, with a higher residual indicating a stronger preference. Conversely, a negative residual implies a tendency to lesion in a different region than the signal, with a lower residual suggesting a stronger preference for this behavior.

The majority (71%) of participants exhibited a positive median residual, indicating a general preference for lesioning within the same region as the signal (average median residual for this subgroup is 0.24). The between-participant median residual distribution is bimodal, with 29% of participants exhibiting a strong preference to make lesions outside of the originating signal's region (mean median residual for this subgroup is -0.45). This may indicate cases when participants associated a signal with an incorrect region.

Alignment between brain map and ground truth does not predict learning success

One important function of scientific conceptualizations is guiding experimentation towards more informative parts of the design space. Hence, we hypothesized that the participants who were given segment maps that aligned well with their signal distribution might end up learning the segments' functions relatively more successfully. To explore this, we tested whether participants' learning success is influenced by the degree to which brain segment regions match the signal distribution. We evaluated this using two metrics: the *isomorphism overlap* score and the *isomorphism discrete* score. The *isomorphism overlap* score assesses the similarity between two Voronoi diagrams: the original brain segment map that a participant interacted with and a hypothetical map generated using the locations of signals as seeds. This score, ranging from 0 to 1, measures the extent of overlap in region assignments vs. signals, with 1 indicating perfect overlap (the original map exactly matches a signal-centered hypothetical map) and 0 indicating no similarity. The *isomorphism discrete* score counts the number of unique regions containing at least one signal. A maximum score of 4 occurs when each of the four signals is located in a separate region, while a minimum score of 1 indicates that all signals are contained within the same region. We performed Kendall correlation tests to examine the relationship between these isomorphism scores and two outcomes: the average distance of participants' lesions to the signals in lesion prediction testing trials, and accuracy in predict-from-image testing trials.

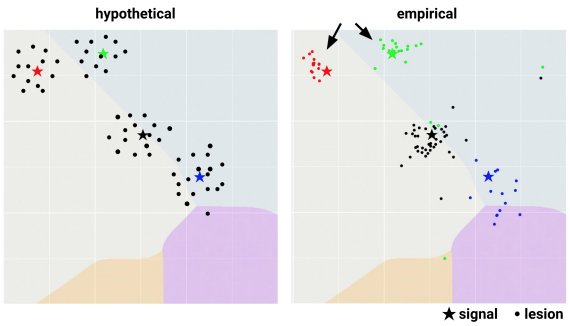

Figure 5: Example participant exhibiting same-region preference when lesioning the signals they found. Filled regions indicate the regions of the subject-specific brain map; stars indicate the underlying signal locations (not visible to the participant), and the dots indicate lesion locations throughout the experiment. Left: hypothetical lesioning data if the participant placed their lesions uniformly with some amount of noise around the signals; right: empirical data with the lesions for the red and green tasks biased away from the boundary.

The *isomorphism overlap* score did not significantly predict learning success, with an estimated correlation of 0.042 ($p = 0.43$) for the distance from signals in lesion prediction trials, and 0.09 ($p = 0.11$) for correctness in predict-from-image trials. Similarly, the *isomorphism discrete* score showed an estimated correlation of 0.03 ($p = 0.62$) for the distance from signals in lesion prediction trials, and 0.04 ($p = 0.52$) for correctness in predict-from-image trials.

More successful learning of signals that were close to boundaries between regions

Earlier, we observed that novices tended to focus their lesioning experiments near the boundaries of brain regions while both experts and novices preferred to lesion the skeletal locations. Based on this, we hypothesized that novices' ability to identify and memorize signals near these boundaries might be enhanced, given that the boundaries could serve as useful landmarks. Hence, we tested whether the proximity of a signal to the nearest region boundary could predict success

on both lesion and predict-from-image testing trials. We employed a mixed effects regression analysis, using success in lesioning and image prediction as the outcomes and random intercepts for each participant.

Novices were more effective at lesioning signals located closer to region boundaries, with a significant positive effect of the distance of the signal from the nearest boundary on distance of the lesion from the signal ($\beta = 0.2774$, $SE = 0.02$, $t(13710) = 12.30$, $p < 0.001$). Similarly, there was a significant increase in participants' success at making predictions based on images for signals closer to these boundaries ($z = -1.982$, $p = 0.0475$), though the effect size was small ($beta = -0.001$, $SE = 0.0006$).

Discussion

We have studied how psychology undergraduate students explore a fictional brain segment when learning about its behavioral functions in a scientific reasoning task. Rather than studying the value of actual brains' partitions into regions or testing ideas about the actual brains' functional-structural mappings (e.g., network vs. localizationist view, etc), in this study we were interested in whether adding region-based information about the brain segment under investigation would influence novices' explorations of it.

Our findings reveal a nuanced relationship between participants' experimentation and the regions of the brain segment. When exploring the segment, novices were drawn to the boundaries between regions, possibly perceiving them as more informative or perhaps because they were simply more perceptually salient. Interestingly, novices' exploration was also attracted by the skeletons—or the medial axes—of the regions. This could be associated with the assumption that "prototypical" locations within regions might be more informative or representative, or, again, to the fact that these skeletons are more perceptually attractive (Firestone & Scholl, 2014; Reed, 1972). Finally, we found marginal evidence that novices overexplored smaller regions and underexplored bigger regions. This could be driven by the perceptual characteristics of the regions of different sizes, or perhaps by the assumption that smaller areas, despite their size, might be as functionally significant as the larger areas (Poldrack, 2006).

However, once having identified a signal, participants preferred to confine their targeted lesions to the region of the inferred signal's location. This shift from boundary attraction to aversion highlights a possible difference between the participants' assumptions about the informative parts of the experimental design space and their ideas about the properties of the signals that have been found (i.e., that they must be contained within one conceptual region). This result is also consistent with the idea that participants might code the signals with respect to the conceptual regions (Sadalla, Burroughs, & Staplin, 1980; Huttenlocher, Hedges, & Duncan, 1991). For instance, a participant might memorize their findings by thinking: "The tactile task is in the red region."

"Good" vs. "bad" concepts

We did not find evidence that a closer match between the conceptual map of the brain segment and its functional structure directly aids participants' learning. This could be attributed to a lack of experimental power since the full randomization of the brain map and ground truth generation prevented us from enforcing a uniform distribution of isomorphism scores in these maps.

While isomorphism in its strictest sense did not prove influential, the boundaries of the regions still played a crucial role in the success of participants' learning. Since novices showed a clear tendency to explore the conceptual boundaries of the brain segment, this attraction proved beneficial when the underlying signals happened to be located near these boundaries–participants were more likely to successfully identify such signals. Therefore, participants were more successful at making predictions for and lesioning the signals which were close to the boundary.

Limitations

Our study presents an initial attempt to empirically study conceptual effects on scientific experimentation, and so it is subject to several limitations. The study's focus is narrowly confined to neuroscience, limiting its applicability to other scientific domains. This specialization, while valuable for understanding concept-ladenness of neuroscientific evidence (Gershman, 2021), restricts the broader generalizability of our findings to different fields of scientific inquiry. Importantly, our experimental paradigm is too simple to faithfully capture the full range of actual investigation practices in neuroscience. Specifically, the "lesioning" and "imaging" experiments available to our participants are simplified and constitute only a small subset of neuroscientist's rich experimental toolkit. Moreover, the region partitioning presented to our participants is not typically present when most neuroscientists make their lesioning choices or interpret imaging results. Thus, our experiment's relevance to actual neuroscientific investigation should be interpreted with caution. Finally, our participant pool did not include expert neuroscientists, also impacting the nature of the insights we can draw from this study. We expect to see many differences between how students and experts approach the task. For example, the boundaries between the regions, which served as exploration magnets for the students, might rather repel the experts. Moreover, the localizationist reasoning exhibited by the novices might not be shared by many expert neuroscientists who are moving beyond the localizationist paradigm (Anderson, 2014; Sporns, 2016; McCaffrey, 2023; Westlin et al., 2023; Pessoa, 2022). We are currently collecting data from expert neuroscientists, hoping to gain a more nuanced understanding of the conceptual biases in neuroscientific research. The potential differences identified in future studies could also point at the discrepancies in laypeople's and experts' interpretations of neuroscientific evidence (e.g., neuroimaging plots).

References

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT Press.

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., Van Ravenzwaaij, D., ... Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, *115*(11), 2607–2612.

Bloch, C. L. (2012). Early concepts in investigative practice-the case of the virus. *Scientific concepts and investigative practice*, 191–218.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., & Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive psychology*, *105*, 9–38.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.

Chang, H. (2022). *Realism for realistic people*. Cambridge University Press.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of rdoc. *BMC medicine*, *11*(1), 1–8.

Daston, L., & Galison, P. (2021). *Objectivity*. Princeton University Press.

Davidenko, N. (2007). Silhouetted face profiles: A new methodology for face perception research. *Journal of Vision*, *7*(4), 6–6.

Dubova, M., & Goldstone, R. L. (2023). Carving joints into nature: reengineering scientific concepts in light of concept-laden evidence. *Trends in Cognitive Sciences*.

Dubova, M., Moskvichev, A., & Zollman, K. (2022). Against theory-motivated experimentation in science. *MetaArXiv. June*, *24*.

Firestone, C., & Scholl, B. J. (2014). "please tap the shape, anywhere you like" shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological science*, *25*(2), 377–386.

Fretwell, S. D., & Calver, J. S. (1969). On territorial behavior and other factors influencing habitat distribution in birds: Ii. sex ratio variation in the dickcissel (spiza americana gmel). *Acta biotheoretica*, *19*(1), 37–44.

Gershman, S. J. (2021). Just looking: The innocent eye in neuroscience. *Neuron*, *109*(14), 2220–2223.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological review*, *98*(3), 352.

Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS computational biology*, *13*(1), e1005268.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., ... others (2017). The hierarchical taxonomy of psychopathology (hitop): A dimensional alternative to traditional nosologies. *Journal of abnormal psychology*, *126*(4), 454.

Lazebnik, Y. (2002). Can a biologist fix a radio?—or, what i learned while studying apoptosis. *Cancer cell*, *2*(3), 179–182.

Longino, H. E. (2002). *The fate of knowledge*. Princeton University Press.

Marr, D., & Poggio, T. (1976). From understanding computation to understanding neural circuitry.

McCaffrey, J. B. (2023). Evolving concepts of functional localization. *Philosophy Compass*, *18*(5), e12914.

Musslick, S., Hewson, J. T., Andrew, B. W., Strittmatter, Y., Williams, C. C., Dang, G. T., ... Holland, J. G. (2023). An evaluation of experimental sampling strategies for autonomous empirical research in cognitive science. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Pessoa, L. (2022). *The entangled brain: How perception, cognition, and emotion are woven together*. MIT Press.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in cognitive sciences*, *10*(2), 59–63.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive psychology*, *3*(3), 382-407.

Rubin, M. (2021). When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese*, *199*(3), 10969–11000.

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of experimental psychology: human learning and memory*, *6*(5), 516.

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, *16*(4), 744–755.

Sporns, O. (2016). *Networks of the brain*. MIT press.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory* (Vol. 1). Princeton university press.

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive psychology*, *104*, 57–82.

Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., ... others (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in cognitive sciences*.

Wilson, W. (1962). A note on the inconsistency inherent in the necessity to perform multiple comparisons. *Psychological Bulletin*, *59*(4), 296.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1.