ORIGINAL ARTICLE

Honeycomb: a template for reproducible psychophysiological tasks for clinic, laboratory, and home use

Nicole R. **Provenza**,[1,2] (iD) Luiz Fernando Fracassi **Gelin**,[3] Wasita **Mahaphanit**,[4,5] Mary C. **McGrath**,[3] Evan M. **Dastin-van Rijn**,[1] (iD) Yunshu **Fan**,[1] (iD) Rashi **Dhar**,[3] Michael J. **Frank**,[4,5] Maria I. **Restrepo**,[3] Wayne K. **Goodman**,[6] David A. **Borton**[1,5,7] (iD)

[1]*Brown University School of Engineering, Providence, RI, USA.* [2]*Charles Stark Draper Laboratory, Cambridge, MA, USA.* [3]*Center for Computation and Visualization, Brown University, Providence, RI, USA.* [4]*Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA.* [5]*Carney Institute for Brain Science, Brown University, Providence, RI, USA.* [6]*Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA.* [7]*Department of Veterans Affairs, Providence VA Medical Center for Neurorestoration and Neurotechnology, Providence, RI, USA.*

Objective: To improve the ability of psychiatry researchers to build, deploy, maintain, reproduce, and share their own psychophysiological tasks. Psychophysiological tasks are a useful tool for studying human behavior driven by mental processes such as cognitive control, reward evaluation, and learning. Neural mechanisms during behavioral tasks are often studied via simultaneous electrophysiological recordings. Popular online platforms such as Amazon Mechanical Turk (MTurk) and Prolific enable deployment of tasks to numerous participants simultaneously. However, there is currently no task-creation framework available for flexibly deploying tasks both online and during simultaneous electrophysiology.
Methods: We developed a task creation template, termed Honeycomb, that standardizes best practices for building jsPsych-based tasks. Honeycomb offers continuous deployment configurations for seamless transition between use in research settings and at home. Further, we have curated a public library, termed BeeHive, of ready-to-use tasks.
Results: We demonstrate the benefits of using Honeycomb tasks with a participant in an ongoing study of deep brain stimulation for obsessive compulsive disorder, who completed repeated tasks both in the clinic and at home.
Conclusion: Honeycomb enables researchers to deploy tasks online, in clinic, and at home in more ecologically valid environments and during concurrent electrophysiology.

Keywords: Neuropsychology; psychophysiology; electrophysiology; task performance and analysis

Introduction

Psychophysiological tasks are a useful tool for studying mental functions, such as cognitive control, learning, and reward evaluation.[1-3] Deficits in executive functioning are a core feature of psychiatric disorders,[4] and analysis of behavioral task performance has revealed significant differences between healthy and psychiatric cohorts.[5-8] Assessing function across specific behavioral constructs, such as those defined by the Research Domain Criteria (RDoC), may provide insight into functional deficits that gold standard clinical assessments may not capture.[9,10] Online tools enable assessment of psychiatric symptoms alongside cognitive performance in large numbers of participants and allow for better delineation of functional deficits that cut across multiple clinical diagnoses. For example, groundbreaking work by Gillan et al.[8] to elucidate symptom dimensions associated with deficits in goal-directed control was made possible by large-scale online studies. Further, analysis of task behavior in large populations and at repeated intervals may help researchers better understand heterogeneity of phenotypes within a single diagnosis and variability in function over time within individuals.

Conducting electrophysiological recordings during psychophysiological tasks may help elucidate the neural mechanisms of functional impairments across psychiatric disorders.[11-15] Electrophysiological recordings, such as scalp electroencephalography (EEG) or intracranial electroencephalography (iEEG), are most commonly conducted in formal research settings such as a laboratory or clinic. However, recent advances in implantable neuromodulation devices present new opportunities for administering tasks during chronic intracranial wireless

recordings in the home environment.[16,17] As noted earlier, these new technologies present particularly exciting opportunities for neural biomarker exploration in severe cases of psychiatric disorders, such as obsessive compulsive disorder (OCD), where the neural underpinnings of symptoms remain unknown.[18,19] Electrophysiological studies require extensive time commitment from researchers and participants alike, and are therefore typically limited to a small number of research participants. Conversely, online task deployment enables the study of population behavior with limited burden on researchers and participants, but this advantage can come at the expense of data quality and participant engagement.

Amazon Mechanical Turk (MTurk) (https://www.mturk.com/) is an internet crowd-sourcing labor marketplace that allows individuals, research groups, and businesses to quickly outsource their jobs to a virtual workforce. Originally intended for human intelligence validation tasks (e.g., labeling images for computer vision algorithm development) and survey participation, MTurk has been heavily used in the past decade for behavioral and psychological studies,[20] revolutionizing psychological science research. Given the platform's initial intention, an open-source software called psiTurk was developed to help researchers interface with MTurk and run online studies without the need to create their own server software.[21] Prolific (Prolific, Oxford, UK), an alternative to MTurk, has gained recent popularity in behavioral research due to its flexible prescreening features. While Prolific does not provide hosting and database management, there are cloud solutions like Firebase (https://firebase.google.com/) and Supabase (https://github.com/supabase/supabase) that can be leveraged as an easy-to-set-up and cost-effective solution.

Deploying the same behavioral task online and in a research setting during concurrent electrophysiology is useful for comparing research participants' behavior to that of a broader population. However, each deployment scenario has unique experimental needs. Online tasks require browser compatibility, whereas conducting concurrent electrophysiological recordings requires time synchronization of task behavior with the external recording system. Desktop-based applications allow event codes or "triggers" to be sent out from the machine running the task during important task events. Event codes are then recorded by the external system and provide a mechanism for precise time alignment between electrophysiological recordings and saved task data. Event code transmission is not possible during online task deployment; browsers are sandboxed environments without access to the system at large, including ports, for security and safety considerations. Today, enabling task deployment in different configurations and environments would likely require two different versions of the task maintained in separate software packages, which may impact task presentation and task performance. To our knowledge, there are currently no standardized methods to deploy psychophysiological tasks from a single code-base consistently to multiple targets online or in lab, and with or without concurrent electrophysiology.

Here, we have combined, into an open-source task-template repository, well-accepted practices and technologies from the cognitive science and web development communities to build psychophysiological tasks that are ready for deployment to different settings (desktop or online) and support electrophysiological recordings, without significant changes to the code base. Deployment specifications are abstracted as parameters that are easy to configure, and application building is automated via GitHub actions, providing continuous delivery of easy-to-download executables, easing setup burden across research sites. The same code-base is used to maintain and deploy the identical task on Mechanical Turk, Prolific, and in research settings during concurrent electrophysiological recordings. We describe our best practice strategy for guaranteeing accuracy in the alignment of task events and electrophysiology with ± 12 ms latency precision, and contribute plans for a low cost, open-source event trigger device that is compatible with Honeycomb and adaptable to most electrophysiology systems. Further, we demonstrate the utility of Honeycomb in psychiatric research by presenting task performance metrics over time from one participant with deep brain stimulation (DBS) for OCD that performed the Multi-Source Interference Task (MSIT) at repeated intervals in the clinic and at home. While the use of Honeycomb in the study of OCD is an example to showcase the utility of the approach, this work applies to a wide array of use cases beyond the example mentioned here (e.g., Honeycomb enables EEG, fMRI, and MEG studies to compare behavioral data against a complementary database from online studies). Lastly, we have curated a library of ready-to-use tasks made publicly available online on The Behavioral Task Hub, termed Beehive, to foster reproducible research across groups. The ability to write one codebase and use it flexibly across settings (with guaranteed consistency in instructions, timing, and other aspects) is an important advance that is highly relevant for psychiatry researchers and beyond.

Method

Honeycomb is a task template repository that standardizes the configuration of jsPsych-based psychophysiological tasks to maintain readability while supporting the flexibility required to enable varied deployment scenarios. Honeycomb provides functionality for web-based and desktop application-based deployment, where the latter requires communication to peripherals for the synchronization of electrophysiological recordings.[22] Honeycomb addresses the key challenge of using a single code base to deliver a task in multiple environments. Tasks built using Honeycomb can easily be deployed as web applications and as cross-platform desktop applications.

The architecture of Honeycomb is summarized in Figure 1 and explained in greater detailed in the online-only supplementary material. Relevant definitions, detailed instructions on how to get started on Honeycomb, and the link to a central and open registry of behavioral tasks (Beehive) are also available as supplementary material.
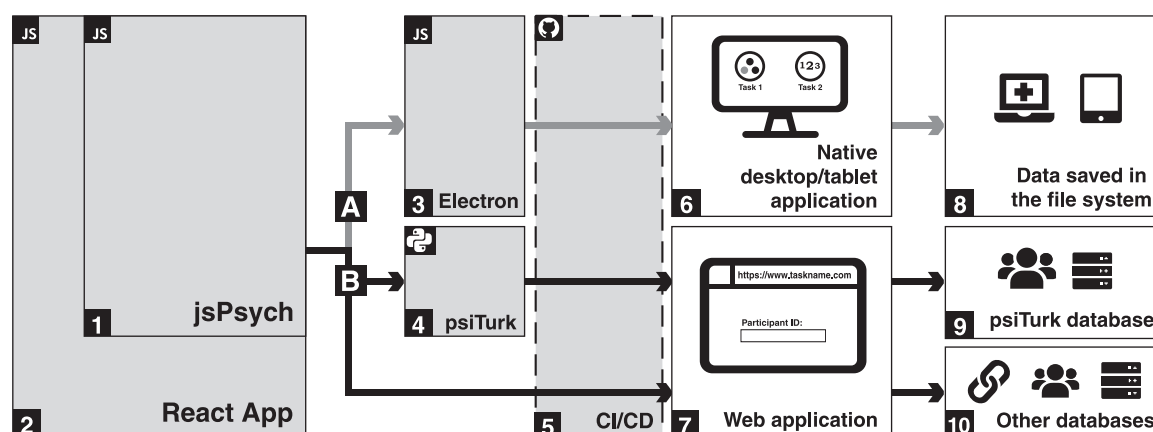
Figure 1 Honeycomb architecture and workflow. Grey boxes show the pieces that are provided with Honeycomb. White boxes represent what the end-user will interact with. Icons on top left represent the language each piece is written on (⬛ javascript, 🐍 python, 🐙 GitHub Actions workflow). Dashed line indicates that continuous integration/continuous delivery (CI/CD) happens on GitHub's servers. Path A (gray arrows) summarizes the flow to generate native desktop applications, while path B (black arrows) show options for web-based deployments. 1. jsPsych is the core of Honeycomb. 2. React is a wrapper around jsPsych allowing for modularity and for setting up connections with backends. The arrows show that the same core code base can be used to create desktop applications via (3) Electron.js, or web applications either via (4) psiTurk, or directly as a React application. 5. The provided workflow files will trigger builds and deployments on GitHub servers via GitHub Actions. These actions can build desktop applications for Windows, Mac, and Linux and create an installer that the users/researchers can download to a desktop or tablet for use in the clinic or at home. In addition, GitHub Actions workflows can build and deploy web applications for access via a web browser. 8. Desktop applications are used in the clinic and at home. In this case the data are saved to the device's file system. 9. psiTurk can be used to manage crowdsourcing via MTurk and the data are saved in the psiTurk database. 10. Researchers can choose to send a link directly to the participants or use other crowd-sourcing services, such as Prolific to distribute the link. In this case the data can be saved to any database (not provided by Honeycomb), or using services like Firebase to host the application and store the data.

To achieve deployment flexibility, Honeycomb is architected in a way that the task core, which is coded with jsPsych and React (Figures 1.1 and 1.2), can be built either as a desktop or a web application (Figure 1, paths A and B, respectively). A task built with Honeycomb can be easily deployed as a static (serverless) web application to services like GitHub Pages, Firebase Hosting, Heroku, or others. This type of deployment can be used with services like Prolific. Furthermore, deploying a psiTurk task requires additional steps that are facilitated by included scripts (Figure 1.4). To build cross-platform desktop applications, Honeycomb includes all the necessary files, helper functions, and scripts that create the installers and executables for Linux, MacOS, and Windows (Figure 1.3). While a researcher can choose to manually build the applications, Honeycomb leverages GitHub Actions to preconfigure and automate the different deployment workflows (Figure 1.5). Finally, to ensure that data are properly managed under the different deployments, Honeycomb supports different formats and mechanisms for saving data (Figures 1.8, 1.9, and 1.10).

Case study

One participant (P1) underwent DBS surgery for treatment of severe, intractable OCD. The participant performed the MSIT at repeated intervals in the clinic and at home. To compare task performance between different locations, we deployed the same Honeycomb-based MSIT (Figure 2A and 2B) both when participant P1 visited the clinic for DBS

programming and when the participant was at home. Response times during the MSIT have been found to improve in response to effective DBS treatment for OCD.[23] Clinical assessments, including the Yale-Brown Obsessive Compulsive Scale (YBOCS), YBOCS II, Hamilton Depression Rating Scale (HDRS), and Young's Mania Rating Scale (YMRS) were administered at each clinical visit.

Ethics statement

The participant gave informed consent and data presented were collected in accordance with recommendations of the federal human subjects regulations and under protocol H-40255 approved by the Baylor College of Medicine Institutional Review Board.

Multi-source interference task

The MSIT was designed to reliably elicit cognitive conflict in human participants.[24,25] MSIT consists of eight 48-trial blocks. Each trial begins with a 2-second fixation period followed by an image of three integers ranging from 0 to 3. The participant was given a keypad and asked to identify the unique number ("target") ignoring its position. Congruent and incongruent trials are illustrated in Figure 2A and 3B. The trial was congruent if the distractor was flanked by invalid targets (0). The trial was incongruent if the distractor was flanked by valid targets, and if the identity of the target did not match its keypad position. Task code is available at https://github.com/brown-ccv/task-msit.
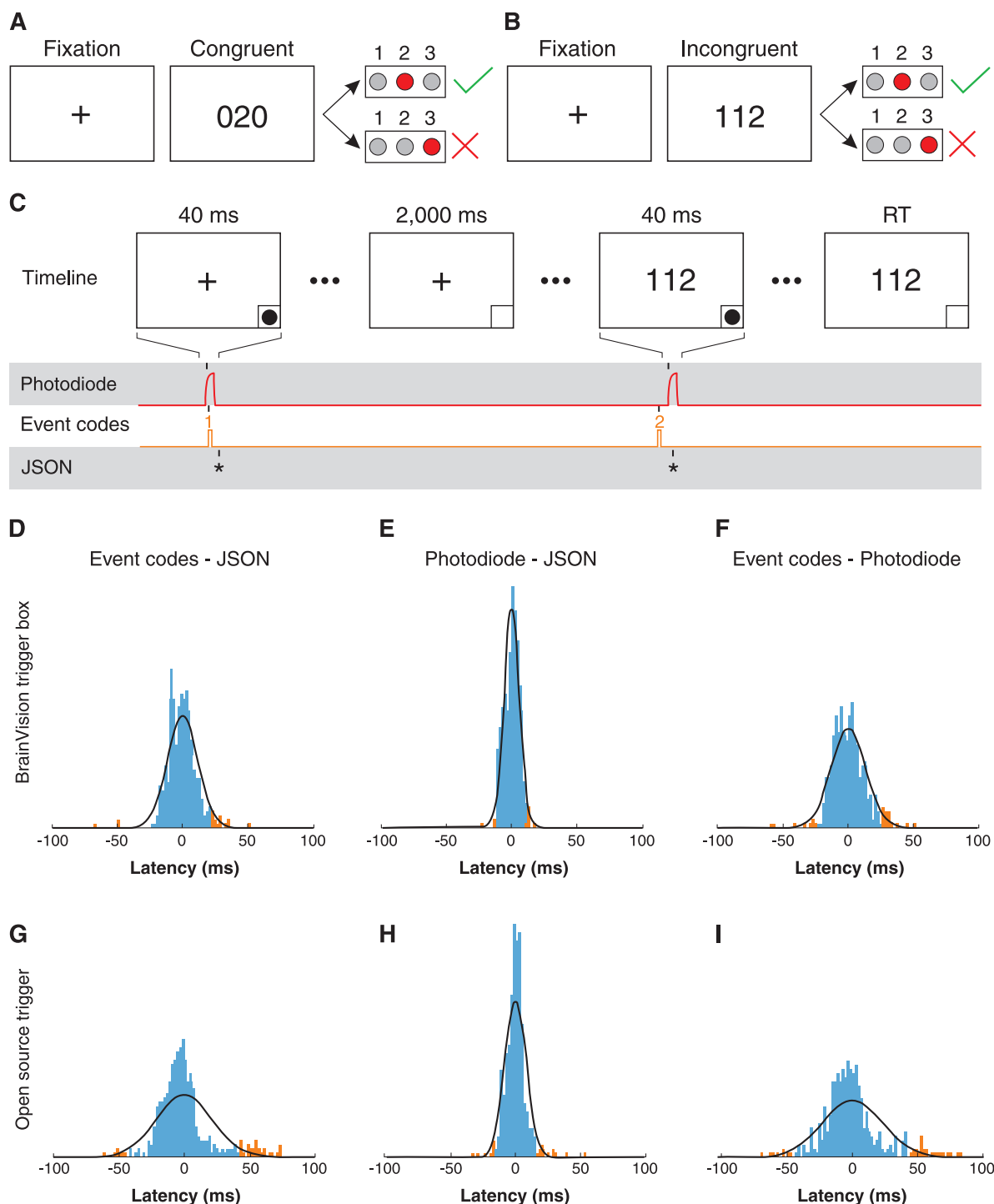
Figure 2 Task overview and event latency testing. A, B) For the Multi Source Interference Task (MSIT), participants are asked to report the one number that is different from the other two using a button box corresponding to the numbers 1, 2, and 3. Trials consist of a fixation period followed by the stimulus. In congruent trials (A), the distractor numbers are zeros and in incongruent trials (B) distractors are other valid numbers. C) Start timing for the fixation and stimulus events were saved in three streams: photodiode pulses, numeric event codes, and timestamps in JavaScript Object Notation (JSON). D-I) Timing latencies between the three streams (D/G: event codes – JSON, E/H: photodiode – JSON, F/I: event codes – photodiode) using BrainVision (D/E/F) and Open Ephys (G/H/I) were estimated by computing the difference in timing for each event. Normal distributions (black) were fit to the latency data and points within the 95% range are indicated in blue and those outside the range are indicated in orange. RT = response time.

Electroencephalography recordings

Two systems were used to record EEG during MSIT: a 64-channel ActiCap BrainVision system with a BrainVision Trigger Box and photodiode (Brain Vision, Morrisville, USA) sampled at 5 kHz, and an Open Ephys acquisition board (Open Ephys, Cambridge, USA) with an open-source event trigger device and photodiode sampled at 30 kHz.

Synchronization of task behavior to external recording system

Previous studies have shown that stimulus and response timing varies across online and application-based experiment generators. Application-based deployment has been found to have 1 ms precision in saving behavioral timing data (e.g., stimulus presentation and response times), led by packages such as PsychToolBox and PsychoPy, while most online packages, including jsPsych, achieved sub 10 ms precision in saving behavioral timing data.[26] While application-based deployment enables higher timing precision, some groups have concluded that the sub 10 ms precision provided by online deployment is also suitable for measuring response times in behavioral research.[27] During concurrent electrophysiological recordings, an additional form of latency should be considered; the latency between the experiment generator (saved task behavior) and the event triggers detected on the external recording system. We thus propose a best practice method for synchronization of EEG and saved behavioral timing data and estimate the timing error for the method.

To guarantee robust synchronization of EEG with saved task behavior, we utilized both a photodiode placed on the screen of the task monitor and event codes sent from the COM port of the task computer to either the BrainVision Trigger Box or the open-source event trigger device. The photodiode captures changes in brightness on the screen at very small latencies and is considered to be a ground truth estimate for when new stimuli are presented on a screen.[28] For example, during the MSIT, the lower right-hand corner of the screen was programmed to show a white dot for 40 ms when new stimuli were presented. The change in brightness is captured in an analog recording channel that provides an accurate time estimation of when the event happened but provides no information about what happened. The event codes can be programmed to carry information about what happened on the screen. For example, during MSIT, fixation crosses send a "1" code, and stimulus presentation sends a "2" code (Figure 2C). Due to latencies of COM port communication however, timing of event codes is less accurate than that of the photodiode.

Time synchronization error between task behavior and external recording system

We quantified the latency between the photodiode recording, event codes, and jsPsych JSON saving to estimate the accuracy at which we can align task behavior to electrophysiological data. A single MSIT run (eight 48-trial

blocks) from participant P1 was used to evaluate each acquisition system: the BrainVision/Trigger Box system and the Open Ephys/open-source event trigger system. Timing information for events was saved in three streams: elapsed time in JSON, photodiode pulses, and numeric event codes. Photodiode crossings were computed by finding the base of peaks in the difference time series identified using the MATLAB (MathWorks, Natick, USA) function findpeaks. In order to align the three streams, the time for the first photodiode threshold crossing was subtracted from the timing for all other photodiode and event code markers, and the corresponding event time in JSON was subtracted from all the JSON times. Differences in timing between the three streams were then computed for each event. The mean of each distribution of differences was then subtracted to account for any offset in the initial alignment point, and normal distributions were fit to the data using the MATLAB function fitdist. To quantify time synchronization error, we computed the latencies corresponding to the 2.5 and 97.5% percentiles of the normal distribution, and reported the absolute deviation from the mean. When using BrainVision, 95% of the latencies were within 23 ms between the JSON and event codes (Figure 2D), within 12 ms between the JSON and photodiode (Figure 2E), and within 25 ms between the photodiode and event codes. When using Open Ephys with the open-source event trigger, 95% of the latencies were within 42 ms between the JSON and event codes (Figure 2D), within 16 ms between the JSON and photodiode (Figure 2E), and within 45 ms between the photodiode and event codes.

Utilizing the photodiode led to the least amount of error in time synchronization between EEG and task behavior ($\sim$ 12 ms using the BrainVision Trigger Box and $\sim$ 16 ms using our open-source event trigger with Open Ephys). These time delays are dependent on the latency between when the stimulus is updated on the screen and the aforementioned latency ($\sim$ 10 ms) between the jsPsych call and the saved timestamp in the output JSON (inherent to the experiment generator). The display of task stimuli on a presentation screen is dependent on-screen refresh rate. For example, 60 Hz screen refresh rate leads to a maximum of 16 ms of potential error between when the jsPsych call is made, and when the task stimulus is presented. We recommend including both the photodiode sensor and event trigger during a task, as it is beneficial both for redundancy and improved synchronization accuracy. The open-source event trigger device has a 20 ms increase in latency when compared to the BrainVision Trigger Box, however, in both cases, the photodiode will alleviate any latencies occurring due to COM port communication.

Synchronization of task behavior to electrophysiological data collected on the same device

Collecting synchronized task data and electrophysiology in the home environment presents different needs than a typical laboratory or clinic scenario. External electrophysiological recording systems are typically not available in the home environment, and therefore the event codes

and photodiode cannot be used to aid in synchronization. Wireless, implantable neuromodulation devices are ideal for enabling participants to trigger their own recordings at home. Electrophysiological data were recorded using the Summit RC + S (Medtronic, Minneapolis, USA) via wireless data streaming from implanted electrodes to the tablet running the task. Each Summit RC + S data packet contains timing information for when it was received and an estimate for the latency between data creation and receipt. Together, these variables can be used to estimate the time of day when each packet was sampled. Task-related timestamps are saved to JSON on the same clock, allowing for direct alignment between the electrophysiological data and the task JSON file. The accuracy of this alignment will depend on the reliability of the estimate for the time between data generation and receipt, and latency inherent to the experiment generator. Without timing correction provided by the photodiode, we rely on JavaScript task timing accurate to the sub 10 ms level, as reported by Bridges et al.[26]

Results

The participant met responder status ($> 35\%$ reduction in YBOCS score) at ~ 180 days after DBS (Figure 3A). The participant performed the task accurately, and as expected we found a significant decrease in accuracy (Figure 3B and 3C) and significant increase in response time (Figure 3D and 3E) for incongruent trials compared to congruent trials. We found that, overall, the participant's MSIT performance in the clinic and at home followed a similar trend. Accuracy and response time in the clinic and at home were not significantly different (*t* test, $p > 0.05$) after accounting for the effect of drift over time. We also found that the characteristic conflict effects in the incongruent trials (i.e., lower accuracy and longer response time) were not only observed in the clinic (triangles in Figures 3C and 3E), but also at home (dots in Figures 3C and 3E). Collecting behavioral data at home did not compromise the quality of the data collected during the MSIT. Moreover, the flexibility of deploying a task at home has enabled ongoing data collection during the COVID-19 pandemic. We were able to keep collecting data despite restrictions on in-person interaction and travel.

Discussion

Currently, using a single code base to build online and desktop-based applications is time-consuming and difficult in popular task creation frameworks such as psychoPy, jsPsych, and PsychToolBox.[29] Maintaining separate code bases for online and desktop-based deployment is time-consuming and error prone. We developed software infrastructure for maintaining and deploying behavioral tasks both online and in research settings during concurrent electrophysiological recordings. We have organized all components that are common to task development and deployment into a template repository that follows best practices in web-deployment and psychophysiological task creation. This template allows for creation of new tasks that are easy to configure and

deployment ready. Further, we have shown that we can accurately align Honeycomb task behavior with external electrophysiology systems with a ± 12 ms latency, and provide plans for a low-cost, open-source event trigger device that is compatible with Honeycomb and is adaptable to most electrophysiology recording platforms. Our motivation was to reduce startup costs for psychology and psychiatry researchers to build, deploy, maintain, and share their own tasks. Simultaneous online and desktop-based task deployment could enable characterization of population behavior and neural biomarker discovery for functional deficits underlying psychiatric disorders.

Administering behavioral tasks both in large populations online and in small participant cohorts during electrophysiological recordings is important for furthering computational psychiatry research.[14,15] Assessing task behavior in a large population of participants allows for characterization of the full spectrum of behavior that can be expected from a behavioral task. Researchers can examine how subclinical measures of behavioral constructs implicated in psychiatric disorders, such as impulsivity or delusions, may predict task behavior in large samples.[30,31] For electrophysiological studies with relatively small cohort sizes, a study participant's behavior can then be put in the context of a broader population. Flexible deployment enables researchers to quantify how subclinical measures and task behavior varies over time within single participants with limited burden on researchers and participants alike. Lastly, online deployment enables the collection of large samples for the sake of task parameter optimizations to achieve desired behavioral effects.

To demonstrate the utility of cross-platform deployment in psychiatry, we presented MSIT performance over time from one participant with DBS for OCD. This work is part of an ongoing study aiming to identify neural biomarkers of OCD symptoms and adverse side effects of DBS in order to enable an adaptive DBS (aDBS) system for OCD. MSIT engages brain circuits related to cognitive control, and deficits in cognitive control have been found in OCD.[24,32-34] By monitoring MSIT performance both in the clinic before regularly scheduled DBS programming sessions and at home, we are able to quantify subtle changes in MSIT performance that may be reflected in or provide supplemental information to routine clinical assessments such as the YBOCS, YBOCS-II, HDRS, and YMRS. Honeycomb enabled us to avoid inconsistencies across the at-home and in clinic version of the task, such as subtle differences in stimulus presentation that could impact behavioral performance and the conclusions drawn from analysis. Additionally, without Honeycomb, it would have been time consuming to set up and maintain the different versions of the task, especially as minor tweaks (e.g., wording of instructions) were made over time. Analyzing how MSIT performance metrics change over time with clinical measures, both in the clinic and at home, may allow us to draw conclusions about how task performance may provide insight to functional deficits that contribute to symptoms, which may be a more productive avenue toward finding neural biomarkers relevant to psychiatry.[10,18] Further, in the future, we plan
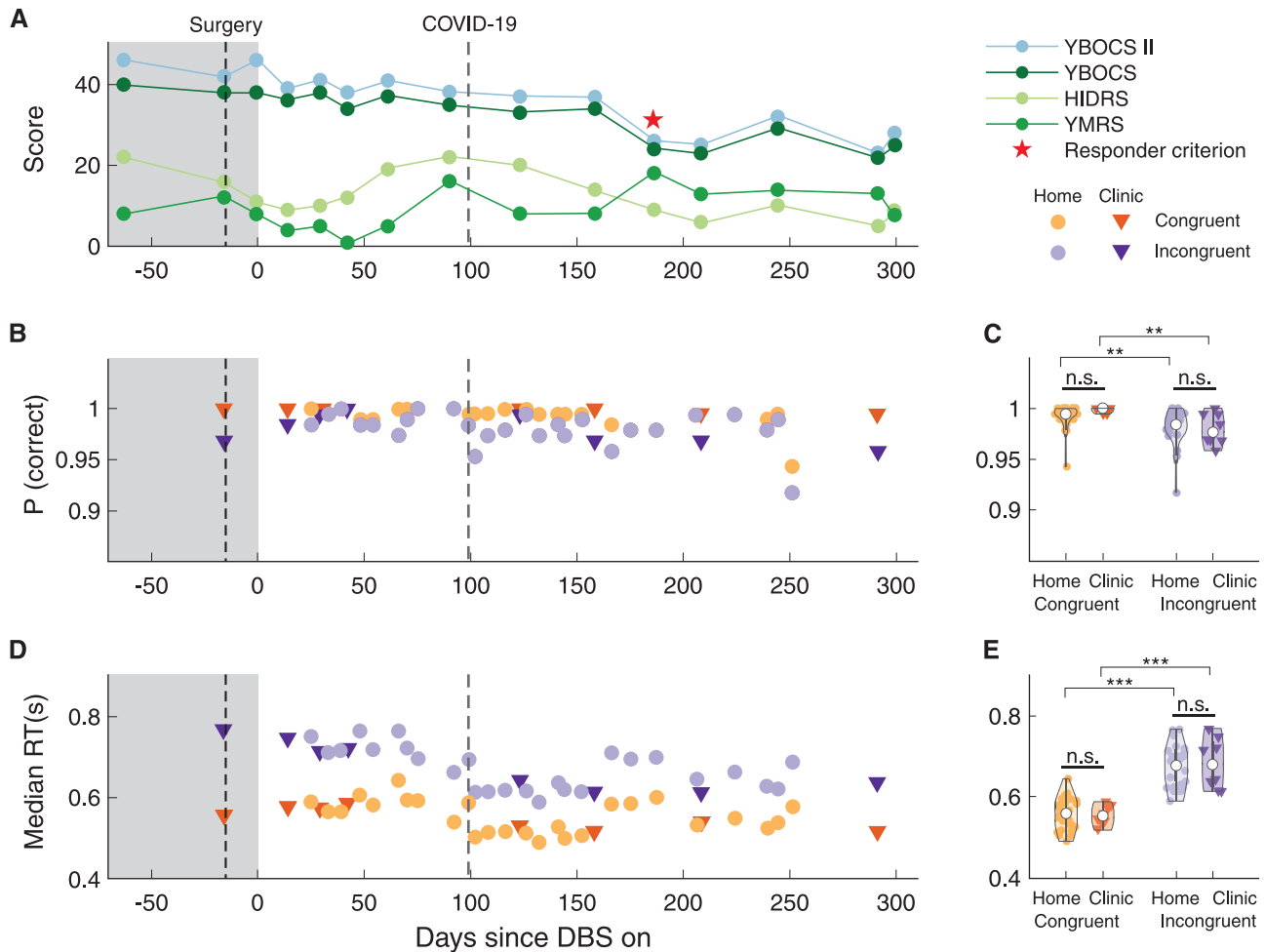
Figure 3 Task summary metrics collected in the clinic and at home are consistent over time. A) Clinical assessments measured during each clinical visit for participant P1. The shaded area corresponds to pre-deep brain stimulation (DBS). Black and grey dotted lines indicate the time of surgical implantation and the start of coronavirus disease (COVID-19) pandemic escalation, respectively. The red star marks the criterion of becoming a responder to DBS treatment, measured as a 35% decrease in YBOCS II score from pre-DBS baseline. (B, D) Participant P1's MSIT task behavior measured in accuracy (P(correct)) in (B) and median response time (median RT) in (D) as a function of days since DBS was turned on. Orange and purple correspond to the congruent and incongruent task conditions, respectively. The triangles represent data collected in the clinic, and the circles represent data collected at home. (C, E) Violin plot showing the distribution of accuracy and median RT in congruent (orange) and incongruent (purple) trials for data collected at home (circles; lighter shade) and in the clinic (triangles; darker shade). White dots indicate the median of the distributions. Note that the at home distributions are not significantly different from the in-clinic distributions, after accounting for the influence of time (multiple linear regression with accuracy (or median RT) as the dependent variable, days from DBS on and the recording location (home or clinic) as the regressors. Congruent trials and incongruent trials were analyzed separately. In each case, the beta coefficient of the recording location is not significant (t test, p > 0.05). Also note that accuracies were lower and median RTs were longer in incongruent trials than in congruent trials for both at-home data and in-clinic data (multiple linear regression with accuracy (or median RT) as the dependent variable, days from DBS on and the trial type (congruent or incongruent) as the regressors. At-home and in-clinic data were analyzed separately. In each case, the beta coefficient of the trial type variable is significant (t test, p < 0.05). HDRS = Hamilton Depression Rating Scale; n.s. = non-significant; YBOCS II = the Second Edition of the Yale-Brown Obsessive-Compulsive Scale; YBOCS = Yale-Brown Obsessive-Compulsive Scale; YMRS = Young Mania Rating Scale.

to use Honeycomb to compare MSIT performance in OCD participants to that of a normative population in a large-scale online study.

In order to choose a task platform and language that would work for the complex needs of our experimental set up, customization was essential. Many groups aim to create psychophysiological tasks with minimal programming (more GUI based), however customizing

these tasks to synchronize with electrophysiology or configure for flexible deployment is difficult or even impossible. While Honeycomb requires programming skills, which some may see as a disadvantage, tasks created with Honeycomb support various deployment configurations and can be extended beyond the functionality provided by our template. Likewise, we acknowledge that many laboratories and researchers have legacy code in

other coding languages that do not allow for online deployment. To take advantage of Honeycomb, legacy code and new tasks would have to be converted to JavaScript. While this does require a time investment up front, there are several jsPsych tutorials to help programmers that are new to JavaScript get started.

In the future, we would like to expand the capabilities of Honeycomb to be compatible with a greater variety of external electrophysiological recording systems. Currently, Honeycomb event code messages are compatible with the BrainVision Trigger Box, and our open-source event trigger device. While the open-source plans for the event trigger device provide most of the functionality for integration with external electrophysiology systems, custom-built adapters are necessary to connect the output of the event trigger device to the analog or digital input of the electrophysiology rig. We would also like to provide out-of-the-box builds for mobile platforms. Finally, we look to continue enhancing modularity and maintainability by introducing a command-line interface program that interactively asks questions to the user and generates a starter task according to the answers. This approach reduces the size of the source code and the builds by removing unused features and deployments. Moreover, upcoming changes and improvements to jsPsych will need to be incorporated to future versions of the Honeycomb.

Lastly, through flexible online deployment of behavioral tasks, we hope to diversify the sample of the population used to represent human behavior in psychology and psychiatry research. Online deployment allows researchers to access populations that would otherwise be impossible to reach. Additionally, we aim to facilitate a way for researchers to continue data collection during the coronavirus disease 2019 (COVID-19) pandemic that can be easily transitioned back to research settings with concurrent electrophysiology when the pandemic is over.

Data and material availability

Code used to produce this manuscript is available on GitHub at https://github.com/brown-ccv/honeycomb.

Acknowledgements

Disclosure

Summit RC + S devices were donated by Medtronic. WKG and DAB have received device donations from Medtronic as part of the NIH BRAIN public-private-partnership program (PPP). WKG has received honoraria from Biohaven Pharmaceuticals and Neurocrine Biosciences. The other authors report no conflicts of interest.

References

1 Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. Psychol Rev. 2001;108:624-52.
2 Gläscher J, Daw N, Dayan P, O'Doherty JP. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron. 2010;66:585-95.
3 Schultz W, Tremblay L, Hollerman JR. Reward processing in primate orbitofrontal cortex and basal ganglia. Cereb Cortex. 2000;10:272-84.
4 Grisanzio KA, Goldstein-Piekarski AN, Wang MY, Ahmed AP, Samara Z, Williams LM. Transdiagnostic symptom clusters and associations with brain, behavior, and daily function in mood, anxiety, and trauma disorders. JAMA Psychiatry. 2018;75:201-9.
5 Fineberg NA, Potenza MN, Chamberlain SR, Berlin HA, Menzies L, Bechara A, et al. Probing compulsive and impulsive behaviors, from animal models to endophenotypes: a narrative review. Neuropsychopharmacology. 2010;35:591-604.
6 Trivedi MH, Greer TL. Cognitive dysfunction in unipolar depression: implications for treatment. J Affect Disord. 2014;152-4:19-27.
7 Robbins TW, Gillan CM, Smith DG, de Wit S, Ersche KD. Neuro-cognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. Trends Cogn Sci. 2012;16:81-91.
8 Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. Elife. 2016;5:e11305.
9 Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry. 2010;167: 748-51.
10 Widge AS, Ellard KK, Paulk AC, Basu I, Yousefi A, Zorowitz S, et al. Treating refractory mental illness with closed-loop brain stimulation: progress towards a patient-specific transdiagnostic approach. Exp Neurol. 2017;287:461-72.
11 Johnson EL, Kam JW, Tzovara A, Knight RT. Insights into human cognition from intracranial EEG: a review of audition, memory, internal cognition, and causality. J Neural Eng. 2020;17:051001.
12 Pujara M, Koenigs M. Mechanisms of reward circuit dysfunction in psychiatric illness: prefrontal-striatal interactions. Neuroscientist. 2014;20:82.95-82.95.
13 Whitton AE, Treadway MT, Pizzagalli DA. Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. Curr Opin Psychiatry. 2015;28:7-12.
14 Browning M, Carter CS, Chatham C, Den Ouden H, Gillan CM, Baker JT, et al. Realizing the clinical potential of computational psychiatry: report from the banbury center meeting, February 2019. Biol Psychiatry. 2020;88:e5-10.
15 Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016;19: 404-13.
16 Skarpaas TL, Jarosiewicz B, Morrell MJ. Brain-responsive neuro-stimulation for epilepsy (RNS® System). Epilepsy Res. 2019;153: 68-70.
17 Stanslaski S, Herron J, Chouinard T, Bourget D, Isaacson B, Kremen V, et al. A chronically implantable neural coprocessor for investigating the treatment of neurological disorders. IEEE Trans Biomed Circuits Syst. 2018;12:1230-45.
18 Provenza NR, Matteson ER, Allawala AB, Barrios-Anderson A, Sheth SA, Viswanathan A, et al. The case for adaptive neuromodulation to treat severe intractable mental disorders. Front Neurosci. 2019;13:152.
19 Olsen ST, Basu I, Bilge MT, Kanabar A, Boggess MJ, Rockhill AP, et al. Case report of dual-site neurostimulation and chronic recording of cortico-striatal circuitry in a patient with treatment refractory obsessive compulsive disorder. Front Hum Neurosci. 2020;14: 569973.
20 Buhrmester M, Kwang T, Gosling SD. Amazon's mechanical Turk: a new source of inexpensive, yet high-quality data? Perspect Psychol Sci. 2011;6:3-5.

21 Gureckis TM, Martin J, McDonnell J, Rich AS, Markant D, Coenen A, et al. psiTurk: an open-source framework for conducting replicable behavioral experiments online. Behav Res Methods. 2016;48:829-42.

22 de Leeuw JR. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. Behav Res Methods. 2015;47:1-12.

23 Widge AS, Zorowitz S, Basu I, Paulk AC, Cash SS, Eskandar EN, et al. Deep brain stimulation of the internal capsule enhances human cognitive control and prefrontal cortex function. Nat Commun. 2019;10:1536.

24 Bush G, Shin LM. The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. Nat Protoc. 2006;1:308-13.

25 Bush G, Shin LM, Holmes J, Rosen BR, Vogt BA. The Multi-Source Interference Task: validation study with fMRI in individual subjects. Mol Psychiatry. 2003;8:60-70.

26 Bridges D, Pitiot A, MacAskill MR, Peirce JW. The timing megastudy: comparing a range of experiment generators, both lab-based and online. PeerJ. 2020;8:e9414.

27 de Leeuw JR, Motz BA. Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. Behav Res Methods. 2016;48:1-12.

28 Schultz BG, Biau E, Kotz SA. An open-source toolbox for measuring dynamic video framerates and synchronizing video stimuli with neural and behavioral responses. J Neurosci Methods. 2020;343:108830.

29 Kleiner M, Brainard DH, Pelli D, Ingling A, Murray R, Broussard C. What's new in psychtoolbox-3? Percept. 2007;36:1-16.

30 Shapiro DN, Chandler J, Mueller PA. Using mechanical Turk to study clinical populations. Clin Psychol Sci. 2012;1:213-20.

31 Bronstein MV, Everaert J, Castro A, Joormann J, Cannon TD. Pathways to paranoia: analytic thinking and belief flexibility. Behav Res Ther. 2019;113:18-24.

32 Shin NY, Lee TY, Kim E, Kwon JS. Cognitive functioning in obsessive-compulsive disorder: a meta-analysis. Psychol Med. 2014;44:1121-30.

33 Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, Enander J, et al. Disorders of compulsivity: a common bias towards learning habits. Mol Psychiatry. 2015;20:345-52.

34 Vaghi MM, Vértes PE, Kitzbichler MG, Apergis-Schoute AM, van der Flier FE, Fineberg NA, et al. Specific frontostriatal circuits for impaired cognitive flexibility and goal-directed planning in obsessive-compulsive disorder: evidence from resting-state functional connectivity. Biol Psychiatry. 2017;81:708-17.