Neuroimaging Article Reexecution and Reproduction Assesment System

Horea-Ioan Ioanas[1] Austin Macdonald[1] Yaroslav O. Halchenko[1]
[1]Center for Open Neuroscience, Department of Psychological and Brain Sciences, Dartmouth College

Abstract — The value of research articles is increasingly contingent on the results of complex data analyses which substantiate their claims. Compared to data production, data analysis more readily lends itself to a higher standard of both full transparency and repeated operator-independent execution. This higher standard can be approached via fully reexecutable research outputs, which contain the entire instruction set for end-to-end generation of an entire article solely from the earliest feasible provenance point, in a programatically executable format. In this study, we make use of a peer-reviewed neuroimaging article which provides complete but fragile reexecution instructions, as a starting point to formulate a new reexecution system which is both robust and portable. We render this system modular as a core design aspect, so that reexecutable article code, data, and environment specifications could potentially be substituted or adapted. In conjunction with this system, which forms the demonstrative product of this study, we detail the core challenges with full article reexecution and specify a number of best practices which permitted us to mitigate them. We further show how the capabilities of our system can subsequently be used to provide reproducibility assessments, both via simple statistical metrics and by visually highlighting divergent elements for human inspection. We argue that fully reexecutable articles are thus a feasible best practice, which can greatly enhance the understanding of data analysis variability and the trust in results. Lastly, we comment at length on the outlook for reexecutable research outputs and encourage re-use and derivation of the system produced herein.

Background

Reexecutable Research

Independent verification of published results is a crucial step for establishing and maintaining trust in shared scientific understanding [5]. The basic feasibility of *de novo* research output generation from the earliest recorded data provenance is known as reexecutability, and has remained largely unexplored as distinct phenomenon in the broader sphere of research reproducibility. While the scope of *reexecution* is narrower than that of *reproduction*, it constitutes a more well-defined and therefore tractable issue in improving the quality and sustainability of research. In all cases, reexecutability increases the feasibility of reproduction assessments. Further, in the case of complex analysis processes with vast parameter spaces, reexecutability is a prerequisite for detailed reproducibility assessments. Lastly, reexecution constitutes a capability in and of itself, with ample utility in education, training, and resource reuse for novel research purposes (colloquially, "hacking") — which may accrue even in the absence of accurate result reproduction.

Free and Open Source Software [27] has significantly permeated the world of research, and it is presently not uncommon for researchers to publish part of the analysis instructions used in generating published results under free and open licenses. However, such analysis instructions are commonly disconnected from the research output document, which is manually constructed from static inputs. Notably, without fully reexecutable instructions, data analysis outputs and the positive claims which they support are not verifiably linked to the methods which generate them.

Reexecutability is an emergent topic in research, with a few extant efforts attempting to provide solutions and tackle associated challenges. Such efforts stem both from journals and independent researchers interested in the capabilities which reexecutable research processes offer to the ongoing development of their work. Among these, an effort by the eLife journal [24] provides dynamic article figures based on the top-most data processing output and executable code conforming to journal standards. NeuroLibre [23] provides a Jupyter Notebook based online platform for publishing executable books along with a selection of reexecutability assets, namely code, data, and a reexecution runtime. Independent researcher efforts offer more comprehensive and flexible solutions, yet provide reference implementations which are either applied to comparatively simple analysis processes [7] or tackle complex processes, but assume environment management capabilities which may not be widespread [14].

In order to optimally leverage extant efforts pertaining to full article reexecution and in order to test

reexecutability in the face of high task complexity, we have selected a novel neuroimaging study, identified as OPFVTA based on author naming conventions [16]. The 2022 article is accompanied by a programmatic workflow via which it can be fully regenerated — based solely on raw data, data analysis instructions, and the natural-language manuscript text — and which is initiated via a simple executable script in the ubiquitous GNU Bash [26] command language. The reexecution process in this effort relies on an emerging infrastructure approach, RepSeP [14], also in use by other articles, thus providing a larger scope for conclusions that can be drawn from its study.

Data Analysis

One of the hallmarks of scientific data analysis is its intricacy — resulting from the manifold confounds which need to be accounted for, as well as from the breadth of questions which researchers may want to address. Data analysis can be subdivided into *data preprocessing* and *data evaluation*. The former consists of data cleaning, reformatting, standardization, and sundry processes which aim to make data suitable for evaluation. Data evaluation consists of various types of statistical modeling, commonly applied in sequence at various hierarchical steps.

The OPFVTA article, which this study uses as an example, primarily studies effective connectivity, which is resolved via stimulus-evoked neuroimaging analysis. The stimulus-evoked paradigm is widespread across the field of neuroimaging, and thus the data analysis workflow (both in terms of *data processing* and *data evaluation*) provides significant analogy to numerous other studies. The data evaluation step for this sort of study is subdivided into "level one" (i.e. within-subject) analysis, and "level two" (i.e. across-subject) analysis, with the results of the latter being further reusable for higher-level analyses [8]. In the simplest terms, these steps represent iterative applications of General Linear Modeling (GLM), at increasingly higher orders of abstraction.

Computationally, in the case of the OPFVTA article as well as the general case, the various data analysis workflow steps are sharply distinguished by their time cost. By far the most expensive element is a substage of data preprocessing known as registration. This commonly relies on iterative gradient descent and can additionally require high-density sampling depending on the feature density of the data. The second most costly step is the first-level GLM, the cost of which emerges from to the high number of voxels modeled individually for each subject and session.

The impact of these time costs on reexecution is that rapid-feedback development and debugging can be stifled if the reexecution is monolithic. While ascertaining the effect of changes in registration instructions on the final result unavoidably necessitate the reexecution of registration and all subsequent steps — editing natural-language commentary in the article text, or adapting figure styles, should not. To this end the reference article employs a hierarchical Bash-script structure, consisting of two steps. The first step, consisting in data preprocessing and all data evaluation steps which operate in voxel space, is handled by one dedicated sub-script. The second step handles document-specific element generation, i.e. inline statistics, figure, and TeX-based article generation. The nomenclature to distinguish these two phases introduced by the authors is "low-iteration" and "high-iteration", respectively [14].

Analysis dependency tracking — i.e. monitoring whether files required for the next hierarchical step have changed, and thus whether that step needs to be reexecuted — is handled for the high-iteration analysis script via the RepSeP infrastructure, but not for the low-iteration script.

Software Dependency Management

Beyond the hierarchically chained data dependencies, which can be considered internal to the study workflow, any data analysis workflow has additional dependencies in the form of software. This refers to the computational tools leveraged by the workflow — which, given the diversity of research applications, may encompass numerous pieces of software. Additionally, individual software dependencies commonly come with their own software dependencies, which may in turn have further dependencies, and so on. The resulting network of prerequisites is known as a "dependency graph", and its resolution is commonly handled by a package manager.

The OPFVTA article in its original form relies on Portage [2], the package manager of the Gentoo Linux distribution. This package manager offers integration across programming languages, source-based package installation, and wide-ranging support for neuroscience software [15]. As such, the dependencies of the target article itself are summarized in a standardized format, which is called an ebuild — as if it were any other piece of software. This format is analogous to the format used to specify dependencies at all further hierarchical levels in the dependency tree. This affords a homogeneous environment for dependency resolution, as specified by the Package Manager Standard [4]. Additionally, the reference article contextualizes its raw data resource as a dependency, integrating data provision in the same network as software provision.

While the top-level ebuild (i.e. the direct software dependency requirements of the workflow) is included in the article repository and distributed alongside it, the ebuilds which specify dependencies further down the tree are all distributed via separate repositories. These repositories are version controlled, meaning that their state at any time point is documented, and they can thus be restored to represent the environment as it would have been generated at any point in the past.

Software Dependencies

The aforementioned infrastructure is relied upon to provide a full set of widely adopted neuroimaging tools, including but not limited to ANTs [3], nipype [9], FSL [22], AFNI [6], and nilearn [1]. Nipype in particular provides workflow management tools, rendering the individual sub-steps of the data analysis process open to introspection and isolated reexecution. Additionally, the OPFVTA study employs a higher-level workflow package, SAMRI [19, 21], which provides workflows optimized for the preprocessing and evaluation of animal neuroimaging data.

Containers

Operating system virtualization is a process whereby an ephemeral "guest" environment is started in and may be reused between persistent "host" systems. Virtual machines (VMs), as these "guest" environments are called, can thus provide users with environments tailored to a workflow, while eschewing the need to otherwise (e.g. manually or via a package manager) provide the tools it requires. Once running, VMs are self-contained and isolated from the host, also eliminating the risk of unwanted persistent changes being made to the host environment. Perhaps the most important benefit of virtual isolation is significantly improved security, allowing system administrators to safely grant users relatively unrestricted access to large-scale computational capabilities. However, VMs can also help mitigate issues arising from package updates by locking a specific dependency resolution state which is known to work as required by a workflow, and distributing that instead of a top-level dependency specification which might resolve differently across time.

Modern advances in container technology allow the provision of the core benefits of system virtualization, but lighten the associated overhead by making limited use of the host system, specifically the hypervisor. Container technology is widespread in industry applications, and many container images are made available via public image repositories. While container technology has gained significant popularity specifically via the Docker toolset, it refers to an overarching effort by numerous organizations, now best represented via a Linux Foundation project, the "Open Container Initiative" (OCI). The OCI governing body has produced an open specification for containers, which can be used by various container runtimes and toolsets. Generally, OCI-compliant container images can be executed analogously with Docker, Podman, or other OCI compliant tools.

While OCI images are nearly ubiquitous in the software industry, Singularity (recently renamed to Apptainer) is a toolset that was developed specifically for high-performance computing (HPC) and tailored to research environments. A significant adaptation of Singularity to HPC environments is its capability to run without root privileges. However, recent advances in container technology have provided similar capabilities. Further, Singularity permits the conversion of OCI images into Singularity images, and recent versions of Apptainer have also added support for natively running OCI containers — thus making reuse of images between the two technologies increasingly convenient.

Container technology thus represents a solution to providing stable reusable environments for complex processes, such as the automatic generation of research articles. In particular, containers provide a convenient way of making advanced package management solutions — as seen in the original OPFVTA article — available to users which may lack them on their host systems.

Results

Repository Structure

In order to improve the reexecution reliability of the OPVFTA article we have constructed a parent repository which leverages Git and DataLad to link all reexecution requirements. This framework uses Git submodules for resource referencing, and DataLad [10] in order to permit Git integration with data resources.

These submodules include the original article, the raw data it operates on, and a reference mouse brain templates package. Additionally, the top-level repository directly tracks the code required to coordinate the OPFVTA article reexecution and subsequent generation of *this* article. The code unique to the reexecution framework consists of container image generation and container execution instructions, as well as a Make system for process coordination (fig. 1). This repository structure enhances the original reference article by directly linking the data at the repository level, as opposed to relying on its installation via a package manager. Notably, however, the article source code itself is not duplicated or further edited here, but handled as a Git submodule, with all proposed improvements being recorded in the original upstream repository. The layout constructed for this study thus provides robust provenance tracking and constitutes an implementation of the YODA principles (a recursive acronym for "YODAs Organigram on Data Analysis" [12]).

The Make system is structured into a top-level Makefile, which can be used for container image regeneration and upload, article reexecution in a containerized environment, and meta-article production. There are independent entry points for both *this* and the original article — making both articles reexecutable (fig. 2). Versioning of the original article reexecution is done via file names (as seen in the `outputs/` subdirectories of fig. 1) in order to preserve shell accessibility to what are equivalent resources. Versioning of the meta-article is handled via Git, so that the

most recent version of the work is unambiguously exposed.

The meta-article targets redirect to a Makefile in the `article/` subdirectory, which contains this document's human-readable text in TEX format, alongside scripts for generating dynamical elements based on the reexecution results seen in the `outputs/` directory. The original article reexecution is provided by two alternative targets, using either the Open Container Initiative standard, or Singularity. Both original article reexecution targets wrap the `produce_analysis.sh` script, which is a thin compatibility layer accessing the Make system of the original article. This alternative is introduced in order to assess feasibility as well as potential variability across virtualization infrastructures.

Resource Refinement

As a notable step in our article reproduction effort, we have updated resources previously only available as tarballs (i.e. compressed `tar` archives), to DataLad. This refinement affords both the possibility to cherry-pick only required data files from the data archive (as opposed to requiring a full archive download), as well as more fine-grained version tracking capabilities. In particular, our work encompassed a re-write of the Mouse Brain Templates package [18] Make system. In its new release [20], developed as part of this study, Mouse Brain Templates now publishes tarballs, as well as DataLad-accessible unarchived individual template files.

Best Practice Guidelines

As part of this work we have contributed substantial changes to the original OPFVTA repository, based on which we formulate a number of best practice guidelines, highly relevant in the production of reexecutable research outputs.

Errors should be fatal more often than not.

By default, programs written in the majority of languages (including e.g. Python and C) will exit immediately when running into an unexpected operation. The POSIX shell and other similar or derived shells, such as Bash and Zsh, behave differently. Their default is to continue with execution of the next scripted command, and only exit with a non-zero code when the list of commands is exhausted or the `exit` command is called explicitly. As a result, an execution of the script could continue for hours before it fails, and the original error message might be lost in the flood of output, making it hard or impossible to localize and address the original problem. This behavior can be mitigated by prepending `set -e` to the respective shell script, which changes the default behavior so that execution is stopped as soon as a command exits with an error code. Additionally, shell scripts treat undefined variables as a variable having an empty value, with empty values causing no errors. This can lead to numerous ill-defined behaviors, including a command such as `rm -rf "$PREFIX/"` removing all files on the system if `PREFIX` is not defined. This can be addressed by prepending `set -u` so that an error is raised and execution is stopped as soon as an undefined variable is referenced. To summarize, we recommend including `set -eu` at the top of every shell script to guarantee it exits as soon as any command fails or an undefined variable is encountered. This is in line with the "Fail Early" principle advocated in the ReproNim Reproducible Basics Module [11].

Avoid assuming or hard-coding absolute paths to resources.

Ensuring layout compatibility in different article reexecution environments is contingent on processes being able to find required code or data. Absolute paths, which are hard-coded into scripts, are likely to not exist anywhere but the original execution environment, rendering the scripts non-portable. This problem is best avoided by adhering to YODA principle [12] of being able to reference all needed resources (data, scripts, container images, etc.) *under* the study directory. Use of relative paths within the study scripts consequently improve their portability. Paths to external resources (scratch directories or reusable resources such as atlases) should additionally be parameterized so that they can be controlled via command line options or environment variables.

Avoid assuming a directory context for execution.

As previously recommended, resources may be linked via relative paths, which are resolved based on their hierarchical location with the respect to the execution base path. However, scripts could be executed from various locations and not necessarily from the location of the script, thus rendering relative paths fragile. A good way of making script execution more robust is ensuring that they set base execution directories to their respective parent directories. This can be accomplished in POSIX shell scripts by prepending `cd "$(dirname "$0")"`.

Workflow granularity greatly benefits efficiency.

The high time cost of executing a full analysis workflow given contemporary research complexity and technical capabilities makes debugging errors very time-consuming. Ideally, it should not be necessary to reexecute the entire workflow for every potentially resolved error. It is thus beneficial to segment the workflow into self-contained steps, which can be executed and inspected independently. Workflows should as a minimum separate such large steps as preprocessing, individual levels of analysis (e.g. per-subject vs. whole-population), and article generation. One way to integrate such steps is to formulate a workflow which automatically checks for the presence of results from prior stages, and, if present, proceeds to the next stage without triggering prior processes. This property is known as itempotence and is again

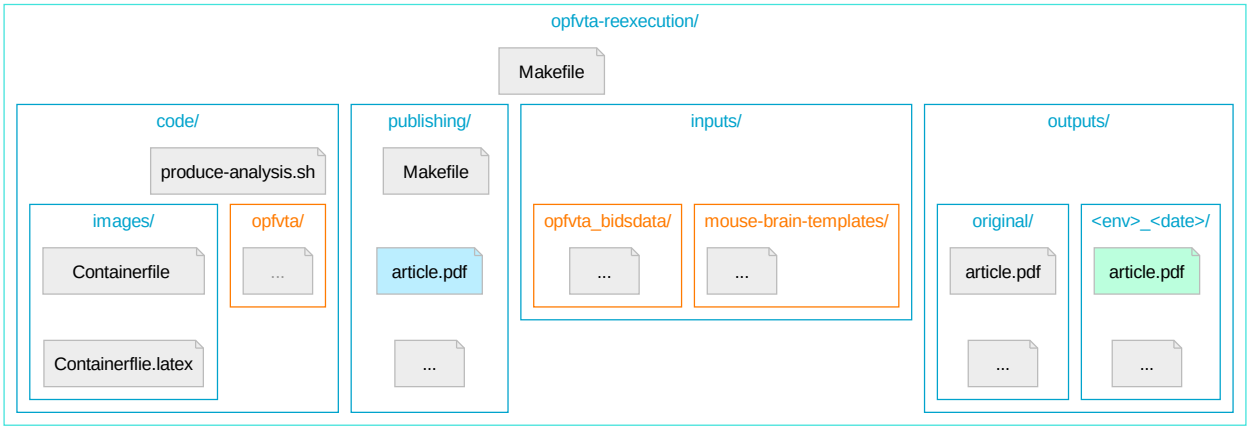


Figure 1: The directory topology of the new reexecution system nests all resources and includes a Make system for process coordination. Depicted is the directory tree topology of the repository coordinating OPFVTA reexecution. Nested directories are represented by nested boxes, and Git submodules are highlighted in orange. The article reexecution PDF results are highlighted in light green, and the PDF of the resulting meta-article (i.e. this article) is highlighted in light blue.

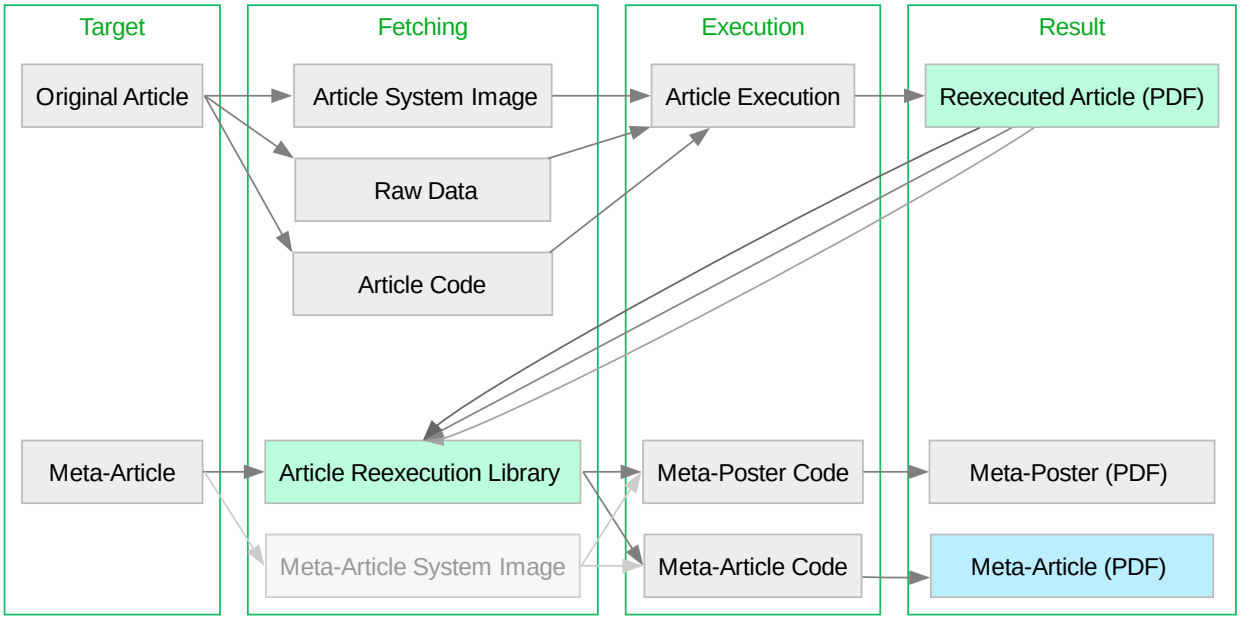


Figure 2: The reexecution system encompasses both the Original Article and Meta-Article as independent Make targets. Depicted is the reexecution system workflow, with two reexecution entry points, the "Original Article and the "Meta-Article" (i.e. *this* article, which also performs the reproduction assessment). Notably, for the generation of the meta-article, the Original Article can be executed, or not — the meta-article will dynamically include all reexecution results which are published, as well as all which are locally produced. The article reexecution PDF results are highlighted in light green, and the PDF of the resulting meta-article (i.e. this article) is highlighted in light blue. Optional nodes (such as fetching a container image for meta-article reexecution) are faded gray.

advocated by the YODA principles, and implemented in this article via both the Make system, as well as internally by the original article's usage of NiPype.

Container image size should be kept small.
Due to a lack of persistency, addressing issues in container images requires an often time-consuming rebuilding process. One way to mitigate this is to make containers as small as possible. In particular, when using containers, it is advisable to *not* provide *data* via a package manager or via manual download inside the build script. Instead, data provisioning should be handled outside of the container image and resources should be bind-mounted after download to a persistent location on the host machine.

Resources should be bundled into a superdataset.
As external resources might change, it is beneficial to use data version control system, such as git-annex and DataLad. The git submodule mechanism permits bundling multiple repositories with clear provenance and versioning information, thus following the modularity principle promoted by YODA. Moreover, git-annex supports multiple data sources and data integrity verification, thus increasing the reliability of a resource in view of providers potentially removing its availability.

Containers should fit the scope of the underlying workflow steps.
In order to constrain the workload of rebuilding a container image, it is advisable to not create a bundled container image for sufficiently self-contained substeps of the workflow. For example, as seen in this study, the article reexecution container image should be distinct from container images required for producing a summary meta-article. Conversely, if sub-steps share toolkit requirements, containers can be re-used between different steps by leveraging different *entry points* to the same target.

Do not write debug-relevant data inside the container.
Debug-relevant data, such as intermediary data processing steps and debugging logs should not be deleted by the workflow or written to an ephemeral location inside the container, but should instead be written to persistent storage. When using some container technologies, such as Docker, files written to hard-coded paths will disappear once the container is removed. As numerous workflow files beyond the main data output may be relevant for debugging, they should not be lost. In order to achieve this, intermediary and debugging outputs should be written to paths which are bind-mounted to persistent directories on the parent system, from which they can be freely inspected.

Scratch directories should be parameterized.
Complex workflows commonly generate large amounts of scratch data — intermediary data processing steps, whose main utility is being read by subsequent steps or consulted for debugging. If these data are written to the same hard-coded path on the host system, multiple reexecutions will lead to race conditions, compromising one or multiple instances of the process. This can be avoided by parameterizing the path and/or setting a default value based on a unique string (e.g. generated from the timestamp). When using containers, this should be done at the container instantiation level, as the relevant path for such potential conflicts is the path on the parent system, and not the path inside the container.

Dependency versions inside container environments should be frozen as soon as feasible.
The need for full image rebuilding means that assuring consistent functionality in view of frequent updates is more difficult for containers than interactively managed environments. This is compounded by the frequent and often API-breaking releases of many scientific software packages. While dependency version freezing is not without cost in terms of assuring continued real-life functionality for an article, it can aid stable re-execution if this is done as soon as all required processing capabilities are provided. How this is accomplished differs greatly based on the package manager used inside the container. Gentoo's Portage package manager allows freezing versions both explicitly, or — as done in this study — by checking out a specific commit of the dependency tree, in view of which the package manager will resolve the same versions. Other distributions (such as Debian and Neurodebian), or language-specific package managers (such as Python's pip), provide analogous functionality, via e.g. `nd_freeze` or `pip freeze`, respectively.

Reproduction Quality

As a top-level view of reexecution results we have produced a simple infrastructure to analyze reproduction quality. This provides both quality control for successful reexecution as well as a showcase of how automatic article reexecutability can be leveraged to evaluate *reproducibility* at a glance.

For this purpose we compare the difference between the Historical Manuscript Record — a product of the original executable article generation — and multiple results generated via the new reexecution system. Reproduction differences between the article versions are extracted by evaluating rasterized page-wise PDF differences (fig. 3).

This overview shows a consistent minimum baseline of differing pixels between reexecutions, around 10^{-4} (i.e. 0.01 %), best seen in pages 6 to 10. When examined closely (fig. 4a), this difference corresponds to the modified date of the Historical Manuscript Record (2022-07-25) and the new reexecution system results (2023-..). While otherwise inconsequential, this difference provides a good litmus test for whether the article was indeed reexecuted or simply preserved, and should be expected throughout all comparisons. Throughout other pages we see difference percentages

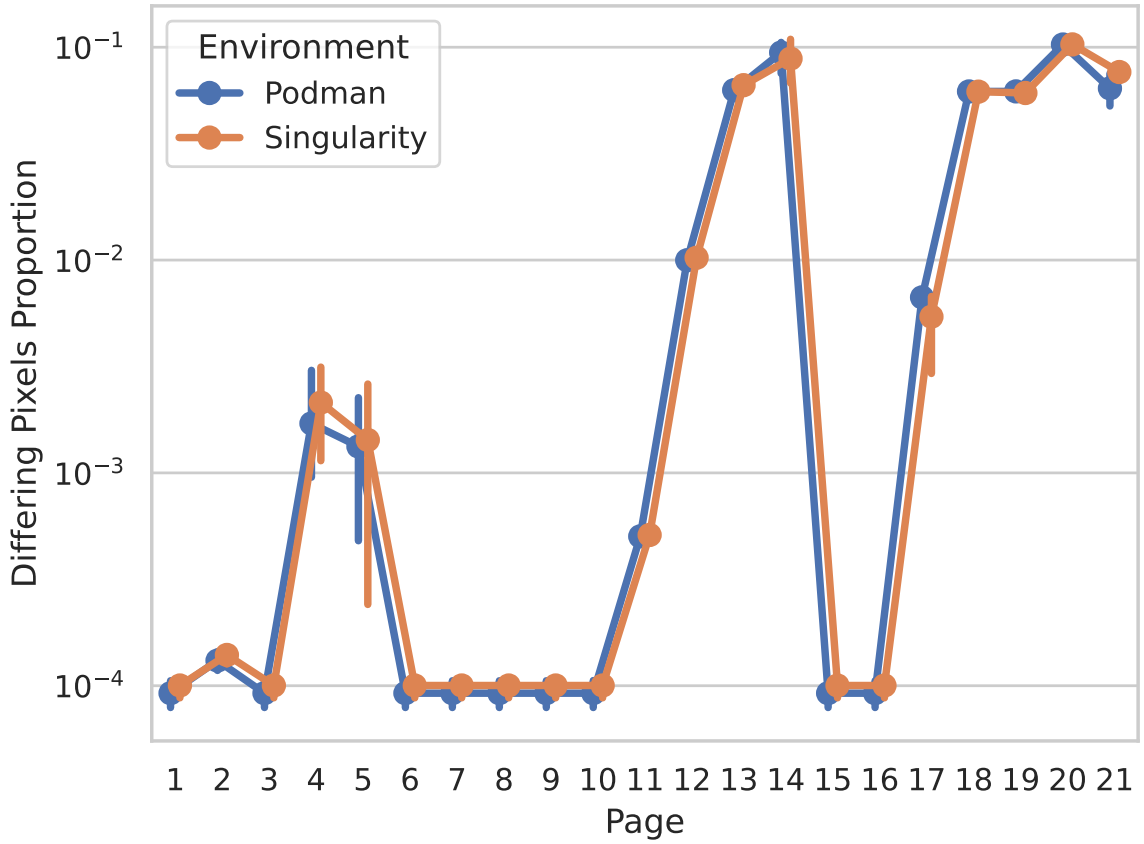
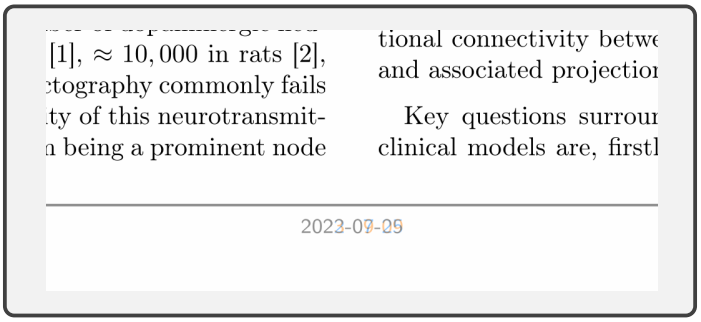


Figure 3: Page-wise visual differences between the Historical Manuscript Record and new reexecution system outputs help identify overall reproduction fidelity, and identify pages with noteworthy differences. Depicted are rasterized document differences, weighted 1 for changes in any pixel color channel, and rounded to four decimal points. Error bars represent the 95th percentile confidence interval.

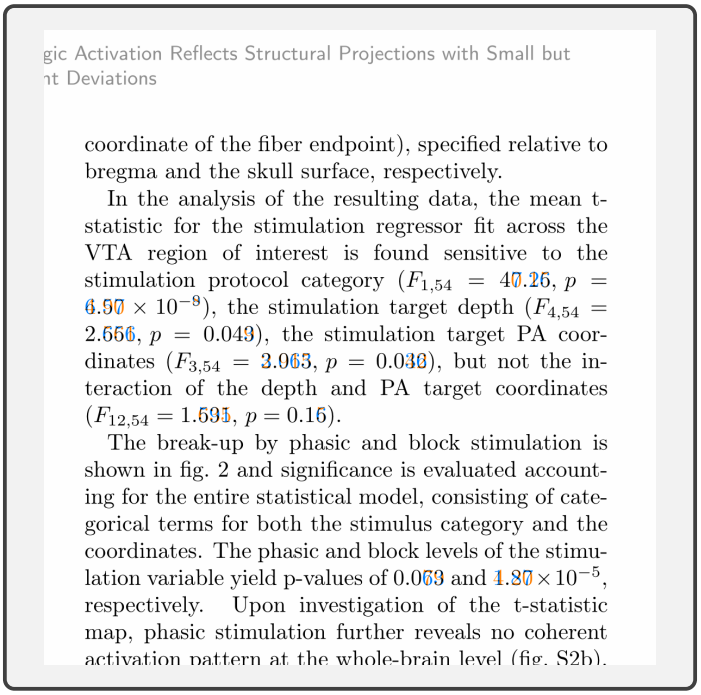
which are broadly consistent across reexecutions and environments, but vary from page to page over almost 2 degrees of magnitude. Upon inspection, more variable but comparatively lower-percentage differences (pages 4 and 5, detail depicted in fig. 4b) are revealed as text differences. This is caused by the target article being fully reexecuted, including the reexecution of inline statistic summaries (e.g. p and F-values). Higher-percentage differences (detail depicted in fig. 4c) correspond to dynamically generated data figures, in which the high variability of non-deterministic preprocessing results in changes to the majority of figure pixels.

Notably, inspecting these differences reveals a strong coherence at the qualitative evaluation level in spite of high quantitative variability. This coherence manifests in the statements from the original article remaining valid with regard to statistical summaries which emerge from *de novo* data processing (as seen in 4b, 4c). This is particularly true for p-values, the magnitude of which can vary substantially at the lower tail of the distribution without impacting qualitative statements.

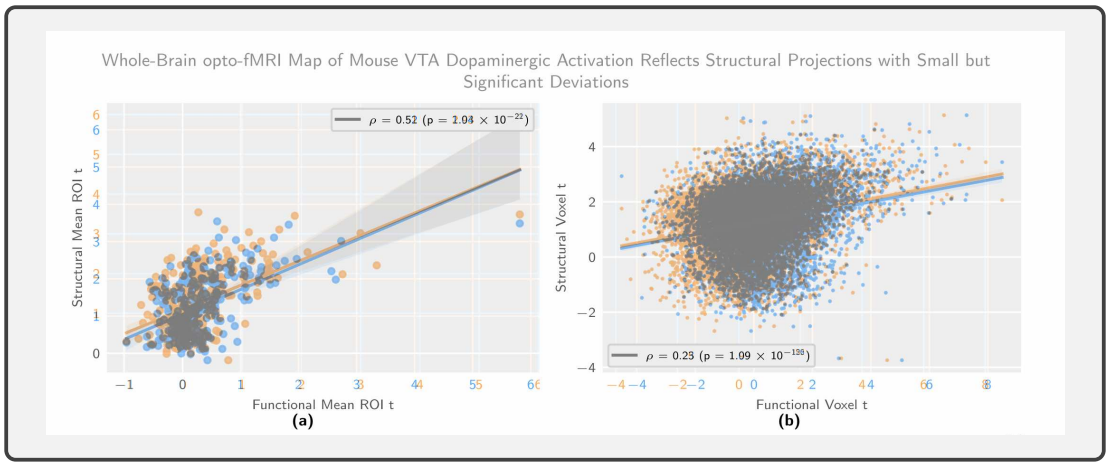
Further, we find that text differences are well localized, as a function of the original article implementing fixed decimal rounding and magnitude notation for statistical outputs (fig. 4). Thus, changes in inline statistic values do not impact text length and do not generally propagate to subsequent lines via word shifts, where they would be recorded as false positives.



(a) The date change is correctly identified throughout the document, as seen in this example from page 1 of the article.



(b) Statistical summary values change, but maintain qualitative evaluation brackets with respect to e.g. p-value thresholds, as seen in this example from page 4 of the article.



(c) In regression analysis, data points are highly variable, the slope and significance remain constant, as seen in this example from page 14 of the article.

Figure 4: The article differences showcase expected quantitative and metadata variability, while maintaining overall validity of qualitative statements. The figures are extracted from a full article `diff`, with tinted highlighting (blue for the Historical Manuscript Record, and orange for the new reexecution system result).

Methods

Data Acquisition

No new animal data was recorded. The data forming the substrate for the reproduction analysis was produced by extracting the output `article.pdf` files from iterative reexecutions of the original article code.

Computing Environments

Article reexecution was performed on a Debian 6.1.8-1 (2023-01-29) system using the `x86_64` architecture, inside containers handled by Podman version `4.3.1` and Singularity version `3.10.3`. Git version `2.39.2` and DataLad version `0.19.2` were used for data and code orchestration. The top-level make targets were executed via Bash version `5.2.15`.

Data Sources

The raw data for the article was sourced in BIDS form from Zenodo, an open data repository, via the identifier specified by the original publication [17]. Mouse brain templates were sourced via a Git repository, "Mouse Brain Templates", which was updated as part of this study to allow individual file fetching [20].

Discussion

In this article and its accompanying source code [13] we present an automated workflow for full, end-to-end article reexecution. We generate the full research communication output (including inline statistics, figures, and brain maps) from solely the raw data and automatically executable code. This work substantiates the feasibility of article reexecution as a process, based on a real-life peer-reviewed study example. To this end, we also detail important and transferable principles, and document common pitfalls in creating a reexecution workflow. Lastly, we leverage the capabilities of this reexecution system in order to provide a simple reproducibility assessment, showcasing the relevance of reexecutable research outputs for investigating reproducibility.

Reexecutability

We argue that reexecutability is a core aspect of reliable research output creation. Reexecutability implies that instructions are formulated in such a way that they can be automatically deployed without human operator bias. In contrast to arbitrary reporting standards, the property of reexecutability implicitly guarantees that required instructions are fully recorded.

We demonstrate the feasibility of full research output reexecution by integrating cutting-edge technological capabilities, and publish all resources for open access, inspection, re-use, and adaptation. The article reexecution system which we produced isolates data and original resources, and does not make assumptions about the internal structure of a reexecutable article, and is of course, not domain-specific. Our system initiates article execution via a Bash entry point, meaning it itself is programmatically accessible for integration into higher-order reexecutable research. We demonstrate the feasibility of this by integrating the original article reexecution with the reexecution of the meta-article. Dependency resolution for the original article is provided via an ebuild-style specification present in the original article code. This means that its dependencies are resolved seamlessly with all lower-level dependencies, and could be resolved seamlessly with higher-order dependencies making use of the reexecutable article as a piece of software.

We sharply distinguish between reexecutability and reproducibility. The former refers to the capability of producing an analogue research output from the same data through automatic execution of data analysis. The latter refers to the coherence between an analogue research output (whether automatically reexecuted or manually recreated) and an original research finding. We further distinguish those two terms from replicability, which describes an identical reproduction of a finding.

Reproducibility

We supplement the reexecution workflow description of this article with a brief demonstration of how it can be used to provide a reproducibility assesment. For this purpose we use a difference computation tool (in computational contexts known simply as "diff") which summarizes and visually displays mismatches between a historical manuscript record and multiple reexecutions over various environments. Such a process makes mismatches visible at-a-glance throughout the article figures and text, rendering them easy to locate and interpret via human inspection.

Based on these results we lay out a few key findings for further reproducibility assessments. In particular, we notice that figures which directly map output data are highly — and to a consistent extent — variable across multiple reexecution attempts. However, in as far as such figures are accompanied by statistical evaluations, we find these to be qualitatively consistent. This indicates that reproduction quality is not only reliant on whether or not data processing is deterministic, but also on which aspects of the top-level data the authors seek to highlight. While the above observations describe the current article specifically, we suspect that they may reflect a phenomenon of broader relevance.

In neuroimaging workflows, the most notorious source for non-deterministic data analysis behavior is the registration. This process commonly operates via a random starting point — specified by a seed value — and iterates according to a gradient descent algorithm. While the toolkit used by the article reexecuted here allows the specification of a particular seed, this was not done for the Historical Manuscript

Record, nor is it a feature commonly used by operators. In light of our results, the question emerges whether or not seed specification should be introduced as a best practice. While a fixed seed would aid in numerical reproducibility, it is possible that a specific seed — whether by coincidence or *ex post facto* selection — may result in anomalous conclusions. It may then be that a stronger finding is one which is statistically robust with respect to preprocessing variability, even if this comes at the cost of compromising numerical replicability. Conversely, it could be argued that reproduction analysis can be better targeted and more concise, if seed values were fixed to universally accepted numbers (analogous to the usage of nothing-up-my-sleeve numbers in cryptography).

Challenges

For this meta-article we have selected an original neuroimaging article which already published all of the instructions needed to reproduce it in its entirety from raw data and automatically executable instructions. Even in light of this uncommon advantage, setting up a portable reexecution system has proven to be an ample effort.

Difficulties arose primarily due to the instability of the software stack. It is common (and increasingly so as researchers become involved in software development) for scientific software to be subjected to frequent interface changes and loss of support for older dependency versions. In this article we propose version-frozen container technology as a mitigation method for such fragility. However, this is not without draw-backs, as it can make introspection more challenging. In view of this, we defined interactive container entry points (`make` targets), whereby the user may enter the container dedicated to automatic reexecution to inspect and test changes in the environment. Even so, on account of these containers being dedicated to automatic execution, features such as an advanced text processor are missing, and the inclusion of such features may not be ultimately desired.

A more easily surmountable challenge was data management. Whereas the original article strove to integrate all provision of computational requirements with the package manager, the usage of containers made the cost of this all-encompassing solution prohibitive. As such, Git submodules and DataLad were used, providing enhanced functionality for e.g. data version specification, but at the cost of spreading requirements provision over different technologies.

Lastly, an unavoidable challenge consisted in execution time-cost. While not prohibitive, the time cost not only slows iterative development work, but presages a potential decrease in the feasibility of reexecution given the trend towards larger and larger data. This means that process complexity and experimental data size may need to be evaluated in light of the diminished accessibility to such processes as reexecution.

Outlook

We propose a few key considerations for the further development of article reexecution — though we note that practical reuse of this system might identify promising enhancements better than theoretical analysis.

In particular, we find that reexecutable article debugging in a container environment can be a significant challenge, and one which will only be more severe if such an environment is already implemented in the development phase of an article. In order to provide seamless integration of both flexible development and portable reexecution, we envision a workflow system which, for each analysis step, permits either usage of locally present executables, or entry points to a container. These two approaches may also be integrated by bind-mount overloading of container components with their local counterparts. We implement a version of this concept for the meta-article generation, where the `make article` target which generates this article will use the local environment, and the `make container-article` target executes the same code via an entry point to a TEX container.

The reproduction quality assessment methods provided in this study serve as a starting point for assessing full article reexecution. We argue that for the reproducibility assessment of a specific article, there is currently no substitute for the human-readable article as the foremost output element, as it most accurately documents all variable elements in the context of the statements they underpin. However, it should be noted that crude pixel-diff comparison, as showcased here, cannot provide automatic evaluation of differences (i.e. determining whether or not statistical thresholds have been crossed) — so machine-readable outputs are necessary for numerical comparisons. There are ongoing efforts, such as NIDM [25], to establish a framework and language for describing numerical results in neuroimaging. This requires custom tooling to export result descriptors in a language aiming to approximate — but distinct from — human readable commentary, and was not yet implemented in our analysis workflow. There are also *supplementary* outputs which may provide additional capabilities, not in lieu of, but in addition to the article text. The foremost among these — specifically pertaining to neuroimaging — are statistical brain maps. Such supplementary data would not only let studies generate reusable outputs, but would also aid the inspection of the original article. Our workflow produces and records all of the top-level data (statistical maps, data tables, etc.) from which the article extracts elements relevant to its statements. We have uploaded the main statistical map of reexecution results to NeuroVault, and are working to provide a corresponding template for our mouse brain data. In-

tegration between the present reexecutable article system and statistical map libraries is thus a promising endeavor for further development.

Lastly, we highlight the relevance of reexecutable articles for reuse and adaptation. Their key strength is that they can easily be derived based on a reliable starting point with respect to successful process execution. This pertains not only to reuse of reexecutable article code for novel or derived studies, but also reuse for the inspection of specific parameter or data modifications. In view of this we recommend a practical approach to the work described herein [13], whereby the parent reexecution system repository can be considered immediately and freely available for inspection, personal exploration, and re-use by the reader.

Acknowledgement

This work was supposed by NIH grants 1R24MH117295 (DANDI: Distributed Archives for Neurophysiology Data Integration) and 2P41EB019936-06A1 (ReproNim: A Center for Reproducible Neuroimaging Computation).

References

[1] Alexandre Abraham et al. “Machine learning for neuroimaging with scikit-learn”. In: **Frontiers in Neuroinformatics** 8 (Feb. 2014). DOI: 10.3389/fninf.2014.00014. URL: https://doi.org/10.3389/fninf.2014.00014.

[2] Guilherme Amadio and Benda Xu. “Portage: Bringing Hackers’ Wisdom to Science”. In: (2016). DOI: 10.48550/ARXIV.1610.02742. URL: https://arxiv.org/abs/1610.02742.

[3] Brian B. Avants et al. “A reproducible evaluation of ANTs similarity metric performance in brain image registration”. In: **NeuroImage** 54.3 (Feb. 2011), pp. 2033–2044. DOI: 10.1016/j.neuroimage.2010.09.025. URL: https://doi.org/10.1016/j.neuroimage.2010.09.025.

[4] Stephen P. Bennett et al. “Package Manager Specification”. English. In: (Apr. 2017). URL: https://projects.gentoo.org/pms/6/pms.html.

[5] Open Science Collaboration. “Estimating the reproducibility of psychological science”. In: **Science** 349.6251 (Aug. 2015). DOI: https://doi.org/10.1126/science.aac4716.

[6] Robert W Cox. “AFNI: software for analysis and visualization of functional magnetic resonance neuroimages”. In: **Computers and Biomedical research** 29.3 (June 1996), pp. 162–173. DOI: 10.1006/cbmr.1996.0014. URL: https://www.sciencedirect.com/science/article/pii/S0010480996900142.

[7] Asim H. Dar, Adina S. Wagner and Michael Hanke. “REMoDNaV: Robust Eye-Movement Classification for Dynamic Stimulation”. In: (Apr. 2019). DOI: 10.1101/619254. URL: https://doi.org/10.1101/619254.

[8] K.J. Friston et al. “Characterizing Evoked Hemodynamics with fMRI”. In: **NeuroImage** 2.2 (June 1995), pp. 157–165. DOI: 10.1006/nimg.1995.1018. URL: https://doi.org/10.1006/nimg.1995.1018.

[9] Krzysztof Gorgolewski et al. “Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python”. In: **Front. Neuroinform.** 5 (2011). ISSN: 1662-5196. DOI: 10.3389/fninf.2011.00013. URL: http://dx.doi.org/10.3389/fninf.2011.00013.

[10] Yaroslav Halchenko et al. **DataLad: distributed system for joint management of code, data, and their relationship**. Vol. 6. 63. The Open Journal, July 2021, p. 3262. DOI: 10.21105/joss.03262. URL: https://doi.org/10.21105/joss.03262.

[11] Yaroslav O. Halchenko et al. **ReproNim Reproducible Basics Module**. 2021. URL: https://www.repronim.org/module-reproducible-basics/.

[12] Michael Hanke et al. **YODA: YODA’s organigram on data analysis**. Poster presented at the annual meeting of the Organization for Human Brain Mapping, Singapore. 2018. eprint: https://github.com/myyoda/poster/blob/master/ohbm2018.pdf. URL: https://github.com/myyoda/poster/blob/master/ohbm2018.pdf.

[13] Horea-Ioan Ioanas, Austin Macdonald and Yaroslav O. Halchenko. **Neuroimaging Article Reexecution and Reproduction Assesment System**. Apache License, verison 2.0. 2024. URL: https://github.com/con/opfvta-reexecution/releases/tag/1.20240119.0.

[14] Horea-Ioan Ioanas and Markus Rudin. “Reproducible Self-Publishing for Python-Based Research”. In: EuroSciPy, Aug. 2018. DOI: 10.6084/m9.figshare.7247339.v1. URL: https://figshare.com/articles/Reproducible_Self-Publishing_for_Python-Based_Research/7247339.

[15] Horea-Ioan Ioanas, Bechara Saab and Markus Rudin. “Gentoo Linux for Neuroscience — a replicable, flexible, scalable, rolling-release environment that provides direct access to development software”. In: **Research Ideas and Outcomes** 3 (Feb. 2017), e12095. DOI: 10.3897/rio.3.e12095. URL: https://doi.org/10.3897/rio.3.e12095.

[16] Horea-Ioan Ioanas, Bechara John Saab and Markus Rudin. “A Whole-Brain Map and Assay Parameter Analysis of Mouse VTA Dopaminer-

gic Activation". In: **bioRxiv** (Apr. 2020). DOI: 10.1101/2020.04.03.023648. eprint: https://www.biorxiv.org/content/early/2020/04/04/2020.04.03.023648.full.pdf. URL: https://www.biorxiv.org/content/early/2020/04/04/2020.04.03.023648.

[17] Horea-Ioan Ioanas, Bechara John Saab and Markus Rudin. **BIDS Data for "A Whole-Brain Map and Assay Parameter Analysis of Mouse VTA Dopaminergic Activation"**. June 2019. DOI: 10.5281/zenodo.3236930. URL: https://doi.org/10.5281/zenodo.3236930.

[18] Horea-Ioan Ioanas, Tina Segessemann and Markus Rudin. **"Mouse Brain Atlases" Generator Workflows**. Version 0.5. Jan. 2019. DOI: 10.5281/zenodo.2545838. URL: https://doi.org/10.5281/zenodo.2545838.

[19] Horea-Ioan Ioanas et al. "An optimized registration workflow and standard geometric space for small animal brain imaging". In: 241 (Nov. 2021), p. 118386. DOI: 10.1016/j.neuroimage.2021.118386. URL: https://doi.org/10.1016/j.neuroimage.2021.118386.

[20] Horea-Ioan Ioanas et al. **Mouse Brain Templates — Generator Workflows and Data Links**. Version 1.0. Jan. 2019. DOI: 10.5281/zenodo.8214739. URL: https://doi.org/10.5281/zenodo.8214739.

[21] Horea-Ioan Ioanas et al. **SAMRI — Small Animal Magnetic Resonance Imaging**. Jan. 2019. DOI: 10.5281/zenodo.3234918. URL: https://doi.org/10.5281/zenodo.3234918.

[22] Mark Jenkinson et al. "FSL". eng. In: **Neuroimage** 62.2 (Aug. 2012), pp. 782–790. DOI: 10.1016/j.neuroimage.2011.09.015. URL: http://dx.doi.org/10.1016/j.neuroimage.2011.09.015.

[23] Agah Karakuzu et al. "NeuroLibre: A preprint server for full-fledged reproducible neuroscience". In: (2022).

[24] Giuliano Maciocci, Michael Aufreiter and Nokome Bentley. "Introducing eLife's first computationally reproducible article". In: **editorial** (Feb. 2019). URL: https://elifesciences.org/labs/ad58f08d/introducing-elife-s-first-computationally-reproducible-article.

[25] Camille Maumet et al. "Sharing brain mapping statistical results with the neuroimaging data model". In: **Scientific Data** 3.1 (Dec. 2016). ISSN: 2052-4463. DOI: 10.1038/sdata.2016.102. URL: http://dx.doi.org/10.1038/sdata.2016.102.

[26] Chet Ramey. "What's GNU: Bash-The GNU Shell". In: **Linux Journal** 1994.4es (1994), 13–es.

[27] Richard Stallman. **Free software, free society: Selected essays of Richard M. Stallman**. Free Software Foundation, 2002. ISBN: 1882114981.

Supplementary

Below, we include the full article difference view, from which showcase excerpts in the article body are sourced.

979

Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

Horea-Ioan Ioanas[1] Bechara John Saab[2] Markus Rudin[1]
[1]Institute for Biomedical Engineering, ETH and University of Zurich
[2]Preclinical Laboratory for Translational Research into Affective Disorders, DPPP, Psychiatric Hospital, University of Zurich

Abstract — Ascending dopaminergic projections from neurons located in the Ventral Tegmental Area (VTA) are key to the etiology, dysfunction, and control of motivation, learning, and addiction. Due to evolutionary conservation of this nucleus and the extensive use of mice as disease models, establishing an assay for VTA dopaminergic signalling in the mouse brain is crucial for the translational investigation of motivational control as well as of neuronal function phenotypes for diseases and interventions. In this article we use optogenetic stimulation directed at VTA dopaminergic neurons in combination with functional Magnetic Resonance Imaging (fMRI), a method widely used in human deep brain imaging. We present a comprehensive assay producing the first whole-brain opto-fMRI map of dopaminergic activation in the mouse, and show that VTA dopaminergic system function is consistent with its structural VTA projections, diverging only in a few key aspects. While the activation map predominantly highlights target areas according to their relative projection densities (e.g. strong activation of the nucleus accumbens and low activation of the hippocampus), it also includes areas for which a structural connection is not well established (such as the dorsomedial striatum). We further detail the variability of the assay with regard to multiple experimental parameters, including stimulation protocol and implant position, and provide evidence-based recommendations for assay reuse, publishing both reference results and a reference analysis workflow implementation.

Background

The dopaminergic system consists of a strongly localized, and widely projecting set of neurons with cell bodies clustered in the midbrain into two lateralized nucleus pairs, the Substantia Nigra pars compacta (SNc) and the Ventral Tegmental Area (VTA, fig. 1a). On account of the small number of dopaminergic neurons ($\approx 300,000$ in humans [1], $\approx 10,000$ in rats [2], and $\approx 4,000$ in mice [3]), tractography commonly fails to resolve the degree centrality of this neurotransmitter system, precluding it from being a prominent node in such graph representations of the brain. However, it is precisely the small number of widely branching and similar neurons, which makes the dopaminergic system a credible candidate for truly node-like function in coordinating brain activity. As is expected given such salient features, the system is widely implicated in neuropsychiatric phenomena (including addiction [4, 5], attentional control [6], motivation [7], creativity [8], personality [9], neurodegeneration [10], and schizophrenia [11]), and is a common target for pharmacological interventions. Lastly, due to high evolutionary conservation [12], the dopaminergic system is also an excellent candidate for translational study.

Imaging a neurotransmitter system comprised of a small number of cells based only on spontaneous activity is highly unreliable due to an intrinsically low signal to noise ratio (SNR). This limitation can, however, be overcome by introducing exogenous stimulation. While the colocalization of widely projecting dopaminergic cell bodies into nuclei renders temporally precise and population-wide targeting feasible, dopaminergic nuclei also contain notable subpopulations of non-dopaminergic cells, which may confound an intended dopaminergic read-out [13]. In order to specifically target dopaminergic cells, they need to be sensitized to an otherwise inert stimulus in a transcription-dependent manner. This can be achieved via optogenetics, which is based on light-stimulation of cells expressing light-sensitive proteins such as channelrhodopsin [14]. Cell-type selectivity can be achieved by Cre-conditional channelrhodopsin vector delivery [15] to transgenic animals expressing Cre-recombinase under a dopaminergic promoter. Following protein expression, stimuli can be delivered via an implanted optic fiber. The combination of this stimulation method with fMRI is commonly referred to as opto-fMRI and can provide information on functional connectivity between a primary activation site and associated projection areas [16, 17].

Key questions surrounding VTA function in preclinical models are, firstly, method feasibility in ani-

mal models more accessible to transgenic techniques, such as the mouse; and secondly, a mapping of the efferent spectrum for dopaminergic VTA output. In particular, in the study of the Rat VTA, it has both been suggested that the efferent dopaminergic spectrum encompasses but extends beyond well-documented structural projections [18] — or alternatively, that VTA dopaminergic efferences are comparatively sparse and that based on translational insight the dopaminergic paradigm of motivation-related VTA function could be questioned [19].

The current study of whole-brain VTA dopaminergic function in mice aims to produce three novel research outputs. Firstly, a proof-of-principle documenting the feasibility of midbrain dopaminergic opto-fMRI in the mouse should be demonstrated, using a protocol which affords qualitative comparability with extant rat data, such as block stimulation and right VTA targeting. Pursuing open questions in the field, results should be quantitatively benchmarked with respect to histologically documented structural projections in the mouse. Secondly, the procedure needs to be optimized by systematic variation of experimental parameters (such as targeting and stimulation protocol variations) in order to ascertain reliability and reproducibility, as is required for a general-purpose dopaminergic system assay. Lastly, a reference neurophenotype of stimulus-evoked dopaminergic function (represented as a brain-wide voxelwise map) should be published in standard space to facilitate co-registered data integration, operative targeting, and comparative evaluation of pathology or treatment induced effects.

These goals presuppose not only the production of experimental data, but also the development of a transparent, reliable, and publicly accessible analysis workflow, which leverages pre-existing standards for mouse brain data processing [20] and extends them to the statistical analysis.

Methods

Animal Preparation

VTA dopaminergic neurons were specifically targeted via optogenetic stimulation. As shown in fig. 1d, this entails a triple selection process. Firstly, cells are selected based on gene expression (via a transgenic mouse strain), secondly the location is selected based on the injection site, and thirdly, activation is based on the overlap of the aforementioned selection steps with the irradiation volume covered by the optic fiber.

A C57BL/6-based mouse strain was chosen, which expresses Cre recombinase under the dopamine transporter (DAT) promoter [21]. Transgenic construct presence was assessed via polymerase chain reaction (PCR) for the Cre construct, using the forward primer ACCAGCCAGCTATCAACTCG and the reverse primer TTGCCCCTGTTTCACTATCC. A total of 25 transgenic animals and 7 wild type control animals are included in the study. The animal sample consisted of 18 males and 14 females, with a group average age of 302 days (standard deviation 143 days) at the study onset. The sample size was determined based on the range found sufficient to uncover opto-fMRI results in the mouse serotonergic system [17].

The right VTA (fig. 3e, green contour) of the animals was injected with a recombinant Adeno-Associated Virus (rAAV) solution. The vector delivered a plasmid containing a floxed channelrhodopsin and YFP construct: pAAV-EF1a-double floxed-hChR2(H134R)-EYFP-WPRE-HGHpA, gifted to a public repository by Karl Deisseroth (Addgene plasmid #20298). Viral vectors and plasmids were produced by the Viral Vector Facility (VVF) of the Neuroscience Center Zurich (Zentrum für Neurowissenschaften Zürich, ZNZ). The solution was prepared at a titer of 5.7×10^{12} vg/ml and volumes from 0.8 to 1.6 µl were injected into the right VTA. Injection coordinates ranged in the posteroanterior (PA) direction from −3.5 to −3.05 mm (relative to bregma), in depth from 4.0 to 4.4 mm (relative to the skull), and were located 0.5 mm right of the midline. Construct expression was ascertained post mortem by fluorescent microscopy of formaldehyde fixed 200 µm brain slices.

For optical stimulation, animals were fitted with an optic fiber implant ($l = 4.7$ mm $d = 400$ µm NA $= 0.22$) targeting the right VTA, at least two weeks before imaging. Implant target coordinates ranged in the PA direction from −3.5 to −3.05 mm (relative to bregma), in depth from 4.0 to 4.6 mm (relative to the skull), and were located 0.5 to 0.55 mm right of the midline.

Stimulation was delivered via an Omicron LuxX 488-60 laser (488 nm), tuned to a power of 30 mW at contact with the fiber implant, according to the protocols listed in tables S1 to S7. Stimulation protocols were delivered to the laser and recorded to disk via the COSplayer device [22]. Animal physiology, preparation, and measurement metadata were tracked with the LabbookDB database framework [23].

MR Acquisition

Over the course of preparation and measurement, animals were provided a constant flow of air with an additional 20 % O_2 gas (yielding a total O_2 concentration of ≈36 %). For animal preparation, anesthesia was induced with 3 % isoflurane, and maintained at 2 to 3 % during preparation — contingent on animal reflexes. Animals were fixed to a heated MRI-compatible cradle via ear bars and a face mask equipped with a bite hook. A subcutaneous (s.c.; right dorsal) and intravenous (i.v.; tail vein) infusion line were applied. After animal fixation, a bolus of medetomidine hydrochloride (Domitor, Pfizer Pharmaceuticals, UK) was delivered s.c. to a total dose of 100 ng/(g BW) and the inhalation anesthetic was reduced to 1.5 %

isoflurane. After a 5 min interval, the inhalation anesthetic was set to 0.5 % and medetomidine was continuously delivered at 200 ng/(g BW h) for the duration of the experiment. This anesthetic protocol is closely based on extensive research into animal preparation for fMRI [24].

All data were acquired with a Bruker Biospec system (7 T, 16 cm bore), and an in-house built transmit/receive surface coil, engineered to permit optic fiber implant protrusion.

Anatomical scans were acquired via a TurboRARE sequence, with a RARE factor of 8, an echo time (TE) of 30 ms, an inter-echo spacing of 10 ms, and a repetition time (TR) of 2.95 s. Thirty adjacent (no slice gap) coronal slices were recorded with a nominal in-plane resolution of $\Delta x(\nu) = \Delta y(\phi) = 75\,\mu m$ (sampled as 180 voxels sagittally and 120 voxels horizontally), and a slice thickness of $\Delta z(t) = 450\,\mu m$.

Functional scans were acquired with a gradient-echo EPI sequence, a flip angle of 60°, and TR/TE = 1000 ms/5.9 ms. Thirty adjacent (no slice gap) coronal slices were recorded with a nominal in-plane resolution of $\Delta x(\nu) = \Delta y(\phi) = 225\,\mu m$ (sampled as 60 voxels sagittally and 29 voxels horizontally), and a slice thickness of $\Delta z(t) = 450\,\mu m$. Functional scans were acquired over a period of 25 min, totalling 1500 repetitions. Changes in cerebral blood volume (CBV) are measured as a proxy of neuronal activity following the administration of an intravascular iron oxide nanoparticle based contrast agent (Endorem, Laboratoire Guebet SA, France) [25]. The contrast agent (30.24 µg/(g BW)) is delivered as an i.v. bolus 10 min prior to the fMRI data acquisition, to achieve a pseudo steady-state blood concentration. This contrast is chosen to enable short echo-time imaging thereby minimizing artefacts caused by gradients in magnetic susceptibility.

The total duration of the scan session, including induction, preparation, and scanning (including the 10 min delay after contrast agent administration, taking place between the structural and functional scan) was approximately 80 min.

MR acquisition was performed blindly with respect to the implant parameter variation, the measurement order was not systematically separated between the conditions. All animal experiments and handling were performed in accordance with the relevant requirements of the Cantonal Veterinary Office of Zurich, under licence ZH263/14 and extension ZH128/18.

Preprocessing

Data conversion from the proprietary ParaVision format was performed via the Bruker-to-BIDS repositing pipeline [26] of the SAMRI package (version `0.4` [27]). Following conversion, data was dummy-scan corrected, registered, and subject to controlled smoothing via the SAMRI Generic registration workflow [20]. As part of this processing, the first 10 volumes were discarded (automatically accounting for volumes excluded by the scanner software). Registration was performed using the standard SAMRI mouse-brain-optimized parameter set for ANTs [28] (version `2.3.1`). Data was transformed to a stereotactically oriented standard space (the DSURQEC template space, as distributed in the Mouse Brain Atlases Package [29], version `0.5.3`), which is based on a high-resolution T_2-weighted atlas [30]. Controlled spatial smoothing was applied in the coronal plane up to 250 µm via the AFNI package [31] (version `19.1.05`).

The registered time course data was frequency filtered depending on the analysis workflow. For stimulus-evoked activity, the data was low-pass filtered at a period threshold of 225 s, and for seed-based functional connectivity, the data was band-pass filtered within a period range of 2 to 225 s.

Statistics and Data

Volumetric data was modelled using functions from the FSL software package [32] (version `5.0.11`). First-level regression was applied to the temporally resolved volumetric data via FSL's `glm` function, whereas the second-level analysis was applied to the first-level contrast and variance estimates via FSL's `flameo`.

Stimulus-evoked first-level regression was performed using a convolution of the stimulus sequence with an opto-fMRI impulse response function, estimated by a beta fit of previously reported mouse opto-fMRI responses [17]. Seed-based functional connectivity analysis was performed by regressing the time course of the voxel most sensitive to the stimulus-evoked activity (per scan) in the VTA region of interest.

Brain parcellation for region-based evaluation was performed using a non-overlapping multi-center labelling [30, 33, 34, 35], as distributed in version `0.5.3` of the Mouse Brain Atlases data package [29]. The mapping operations were performed by a SAMRI function, using the nibabel [36] and nilearn [37] libraries (versions `2.3.1` and `0.5.0`, respectively). Classification of implant coordinates into "best" and "rejected" categories was performed via 1D k-means clustering, implemented in the scikit-learn library [38] (version `0.20.3`). Distribution density visualizations were created using the Scott bandwidth density estimator [39], as implemented in the seaborn software package (`0.9.0`).

Higher-level statistical modelling was performed with the Statsmodels software package [40] (version `0.9.9`), and the SciPy software package [41] (version `1.1.0`). Model parameters were estimated using the ordinary least squares method, and a type 3 analysis of variance (ANOVA) and a heteroscedasticity consistent covariance matrix [42] were employed to control estimate variability for unbalanced categories. All t-tests producing explicitly noted p-values are two-tailed.

The VTA structural projection data used to compare and contrast the activation maps produced in this study was sourced from the Allen Brain Institute (ABI) mouse brain connectome dataset [43]. As the target promoter of this study (DAT) is not included in the ABI connectome study, all available promoters were used (Sty17, Erbb4, Slc6a3, Th, Cck, Pdzk1ip1, Chrna2, Hdc, Slc18a2, Calb2, and Rasgrf2). Datasets with left-handed VTA injection sides were flipped to provide right-hand VTA projection estimates. The data was converted and registered to the DSURQEC template space by the ABI Connectivity Data Generator package [44]. For the second-level statistical comparison between functional activation and structural projection, individual activation (betas) and projection maps were normalized to a common scale by subtracting the average and dividing by the standard deviation.

Software management relevant for the exact reproduction of the aforementioned environment was performed via neuroscience package install instructions for the Gentoo Linux distribution [45].

All data analysis was performed on the entire dataset, without any data being removed, and in the absence of individual category investigation.

Reproducibility and Open Data

The resulting t-statistic maps (i.e. the top-level data visualized in this document), which document the opto-fMRI dopaminergic map in the mouse model, are distributed along the source-code of all analyses [46]. The BIDS [47] data archive which serves as the raw data recourse for this document is openly distributed [48], as is the full instruction set for recreating this document from the aforementioned raw data [46]. The source code for this document and all data analysis shown herein is structured according to the RepSeP specifications [49].

Results

Opto-fMRI experiments were carried out in C57BL/6 mice expressing Cre recombinase under the dopamine transporter promoter [21], with Cre-conditional viral vector induced expression of channelrhodopsin (ChR2) and yellow fluorescent protein (YFP) in the dopaminergic midbrain. Light stimuli were delivered via an optic fiber pointing above the right VTA. Different stimulation protocols were applied to the animals, consisting of variations within two main categories: block stimulation (with light stimuli delivered in continuous blocks of at least 8 s — tables S1 to S5) and phasic stimulation (with light stimuli delivered in short bursts of up to 1 s in lenght — tables S6 and S7). Additionally, the dataset details the effects of variation in the posteroainerior (PA) coordinates and the implant depth (equivalent to the dorsoventral coordinate of the fiber endpoint), specified relative to bregma and the skull surface, respectively.

In the analysis of the resulting data, the mean t-statistic for the stimulation regressor fit across the VTA region of interest is found sensitive to the stimulation protocol category ($F_{1,54} = 47.26$, $p = 6.57 \times 10^{-9}$), the stimulation target depth ($F_{4,54} = 2.656$, $p = 0.043$), the stimulation target PA coordinates ($F_{3,54} = 3.063$, $p = 0.036$), but not the interaction of the depth and PA target coordinates ($F_{12,54} = 1.591$, $p = 0.16$).

The break-up by phasic and block stimulation is shown in fig. 2 and significance is evaluated accounting for the entire statistical model, consisting of categorical terms for both the stimulus category and the coordinates. The phasic and block levels of the stimulation variable yield p-values of 0.063 and 4.87×10^{-5}, respectively. Upon investigation of the t-statistic map, phasic stimulation further reveals no coherent activation pattern at the whole-brain level (fig. S2b).

The main and interaction effects of the implant coordinate variables are better described categorically than linearly (figs. S1 and 2b). Consequently, the most suitable implant coordinate group for the assay can best be determined on the basis of categorical classification of implant coordinates. We classify the implant coordinates into a "best" and a "rejected" group by k-means clustering the aggregate VTA t-statistic scores into two clusters, and find spatial coherence for the "best" coordinate group (categorization highlighted in fig. 2b).

For block stimulation, the best implant category group (fig. 3a) and the rejected implant category group (fig. 3c) show not only a difference in overall stimulus-evoked signal intensity, but also a difference in efferent distribution, with the rejected implant category efferent spectrum more strongly weighted towards caudal brain areas. This distinction specifically arises for implant categorization based on block scan VTA t-statistic means, and is not as salient if implants are categorized based on a posteroanterior implant coordinate delimiter (fig. S3).

The activation pattern elicited by block stimulation in the best implant category group shows strong coherent clusters of activation. The top activation areas are predominantly located in the right hemisphere, with highly significant laterality ($p = 8.27 \times 10^{-7}$) seen in the comparison of left and right hemisphere atlas parcellation region averages. Activation is seen in regions surrounding the stimulation site, such as the ventral tegmental decussation and the interpeduncular nucleus. The largest activation cluster encompasses well-known dopaminergic VTA projection areas in the subcortical rostroventral regions of the brain (nucleus accumbens, striatum, and the basal forebrain), with weaker activation observed in smaller structures in the vicinity of these regions, such as the fasciculus retroflexus, anterior commissure and the claustrum.

983

This activation pattern is is largely consistent with structural projection data, as published by the Allen Brain Institute [43] with a few notable distinctions (fig. 4). At the parcellation level, we see a moderately strong positive correlation between functional activation and structural projection (fig. 4a), which is weaker at the voxel level (fig. 4b). In the midbrain, the coronal slice map shows areas of increased functional activation with respect to structural projection density in the contralateral VTA and the ipsilateral substantia nigra. Coherent clusters of increased activation are also observed in projection areas, most prominently in the ipsilateral and contralateral dorsomedial striatum (fig. 4c). Parcellation-based distributions (figs. 4d and 4e) show this increased activation map encompassing additional areas in the contralateral hemisphere, in particular the contralateral nucleus accumbens, with activity extending into the claustrum. Areas for which structural projections clearly outweigh the functional response are few and dispersed. These small clusters yield only weak negative contrast distributions and are located predominantly in the cerebellum (fig. 4d).

We differentiate VTA transmission from VTA excitability by mapping functional connectivity using a seed region in the right VTA, which yielded the projection pattern shown in fig. 3e. These clusters are more sparse compared to those identified by stimulus-evoked analysis, yet follow a similar distribution. While areas displaying the highest functional connectivity are located in the right hemisphere, the whole brain parcellation-resolved response displays no significant laterality ($p = 0.11$). Strong activation can be seen in the parcellation regions surrounding the seed, such as the ventral tegmental decussation and the closely located interpeduncular nucleus. In the midbrain, seed-based functional connectivity highlights both the ipsilateral and the contralateral VTA with great specificity, unlike sitmulus-evoked analysis (figs. 3a and 3e). Rostrovental dopaminergic projection areas remain prominently featured, including the nucleus accumbens and the striatum (fig. 3f).

Stimulation in wild type control animals (which is corrected for in the aforementioned stimulus-evoked analyses) does not exhibit a pattern of activity consistent with dopaminergic projections. Sparse grains containing regression scores of $t \geq 3$ can be observed, with the largest cluster in the lateral geniculate nucleus area of the thalamus, suggesting visual activity (fig. S5b). Atlas parcellation score distributions (fig. S5c) do not strongly deviate from zero, with the highest scoring areas being in the vicinity of the fiber, possibly indicating VTA heating artefacts. Comparable region t-statistic distributions are also found in areas of the cerebellum. Overall the whole brain parcellation-resolved response shows no significant laterality ($p = 0.08$).

Histological analysis of the targeting site reveals that the optic fiber implant displaces the YFP labelled neurons of the VTA (fig. 5). This dislocation was observed irrespective of the targeting area or the speed of implant insertion (10 to 50 µm/s). Yet, labelled filaments and soma remain in the imediate vecinity of the fiber tip, as seen in higher magnification images (fig. 5c).

Discussion

Whole-Brain Dopaminergic Map

In this article we present the first whole-brain opto-fMRI map of VTA dopaminergic activity in the mouse. Published as voxelwise reusable data and discussed in terms of regions of interest in the article text, this constitutes an essential resource for preclinical investigation of the dopaminergic system. The areas identified as functional VTA dopaminergic targets are largely consistent with histological and electrophysiologic literature (as summarized in fig. 1a). This highlights the suitability of opto-fMRI for interrogating the mouse dopaminergic system, which opens the way for longitudinal recording with whole-brain coverage.

The predominant VTA projection area identified both in literature and in our study is the nucleus accumbens. This area is involved in numerous neuropsychological phenomena, and its activation further supports the method's suitability to resolve meaningful brain function and increase the predictability of novel interventions using the mouse model organism. Particularly, potential limitations of dopaminergic VTA imaging as shown in recent literature [19], appear to not constrain the protocol detailed in this study.

Throughout brain regions with high signal amplitudes on either metric, we observe a high degree of correspondence between functional activation and structural projection density. Yet, we also document a number of notable differences between opto-fMRI derived projection areas and the structural substrate of the dopaminergic system. Overall, the contrast between function and structure shows stronger signal and wider coverage for the functional activation pattern, particularly in projection areas. Notably, the functional map extends into the contralateral ventral striatum, and both the contralateral and ipsilateral dorsal striatum. Activation of the contralateral ventral striatum might be attributed to an extension of the functional map to the contralateral VTA. This interpretation is supported by the contralateral projection areas showing lower overall significance scores than the ipsilateral areas (figs. 3b and 3f). The explanation of projection area extension into the dorsal striatum on account of secondary activation of the ipsilateral substantia nigra is however less reliable, since the most relevant cluster of increased functional activation — the dorsomedial striatum — can be observed bilaterally, though potential nigral activation is only seen ipsilaterally (fig. 4c). Together with other recent

literature [18, 50], it is also possible that VTA activation on its own elicits dorsomedial striatial activity. Not least of all, the local deformation of the VTA upon fiber implantation may additionally confound parcellation in the vicinity of the fiber tip (fig. 5).

Negative contrasts clusters between functional activation and structural projection are overall very sparse (fig. 4d). Yet, the amygdala, hippocampus, and the medial prefrontal cortex — known targets for VTA dopaminergic projections — do not reveal strong activation in opto-fMRI. Comparison with published structural projection data indicates that this is due to low fiber bundle density, as these areas also do not show high amounts of structural projections.

In the pursuit of differentiating primary activation from subsequent signal transmission (and resolving a dopaminergic graph relay model, as depicted in fig. 1b) we present an analysis workflow based on VTA seed-based connectivity. Our results indicate that this analysis is capable of identifying projection areas, but is significantly less powerful than stimulus-evoked analysis (fig. 3a). VTA seed-based analysis highlights only a small number of activation clusters and fails to show significant projection laterality. This is an interesting outcome, as — given the superior performance of stimulus-evoked analysis — it describes two possible features of dopaminergic neurotransmission in the VTA. The first is that the relay of primary VTA stimulation has higher fidelity than the fMRI measurement of VTA activity itself (i.e. VTA activity is relayed accurately, but outweighed by measurement noise). The second is that there is a significant threshold to dopaminergic neurotransmission, by which fMRI-measurable baseline activity is predominantly not propagated (i.e. VTA activity is measured accurately, but is relayed in a strongly filtered fashion). The seed-based analysis workflow, however successfully disambiguates VTA activation from adjacent midbrain activation including for the contralateral VTA, which is outside of the seed region of interest. This indicates that VTA susceptibility to optogenetic stimulation may have a unique signature compared to surrounding midbrain tissue in which activation is also elicited in opto-fMRI.

Assay Parameters

This article presents an evidence-based outline for assay reuse and refinement. In particular, we detail the effects of stimulus protocol categories and optogenetic targeting coordinates on the performance of the method.

The break-down of target coordinates for optical stimulation (fig. 2) indicates that more rostral and deeper implant coordinates elicit stronger VTA signal responses to block stimulation trials. Based on our data we suggest targeting the optic implant at a posteroanterior distance of −3.05 mm from bregma, a left-right distance of 0.5 to 0.55 mm from the midline, and a depth of 4.5 mm from the skull surface. Additional coordinate exploration might be advisable, though further progression towards bregma may lead to direct stimulation of specific efferent fibers rather than the VTA.

The absence of VTA activation as well as coherent activity patterns elicited by phasic stimulation (figs. 2a and S2b) highlights that phasic stimulation is unable to elicit activation measurable by the assay in its current form. The overall low susceptibility to phasic stimulation is most likely due to the intrinsically lower statistical power of such stimulation protocols in fMRI.

Regarding the distribution of activation across projection areas, we note a strong and unexpected divergence between the most sensitive ("best") and least sensitive ("rejected") implant coordinate category responses to block stimulation (figs. 3a and 3c). In addition to a difference in VTA and efferent signal intensity (expected as per the selection criterion), we also notice a different pattern of target areas. Interestingly, the activity pattern elicited in the "rejected" group is more strongly weighted towards the hindbrain, and the efferent pattern includes the periaqueductal gray, a prominent brainstem nucleus involved in emotional regulation [51]. This effect might be related to the activation of descending dopaminergic projections, though further investigation is needed to clarify this point and, in general, to better understand the cross-connectivity between deep brain nuclei.

The activation patterns in wild type control animals are very sparse (fig. S5), and — whether or not they are controlled for in the form of a second-level contrast — do not meaningfully impact the dopaminergic block stimulation contrast (figs. 3a and S4). Based on the activation distribution, however, it may be inferred that trace heating artefacts (midbrain activation) and visual stimulation (lateral geniculate nucleus thalamic activation) are present. On account of this, for further experiments, we suggest using eye occlusion and dark or dark-painted ferrule sleeves (to avoid visual stimulation), as well as laser power lower than the 30 mW (239 mW/mm^2) used in this study (to further reduce heating artefacts).

Stimulus-evoked analysis displayed significant laterality; nevertheless, large clusters displaying significant activation were also observed on the contralateral side. Fluorescence microscopy (fig. 4c) revealed that expression of the viral construct injected at the site of the right VTA extends over a large area, including part of the contralateral VTA. Inspection of the functional map at the midbrain stimulation site corroborates that activity in fact spreads to the contralateral VTA (fig. 3a). This explains the occurrence of contralateral fMRI responses, which are most likely weaker due to a lower photon fluence at the site of the left VTA. Together, these data suggest that the solution volume and virus amount injected for the assay could be significantly reduced, to less than the 0.8 µl

985

(5.7×10^{12} vg/ml) used as the minimal volume in this study.

The most salient qualitative feature of fig. 5 is, however, the displacement of labelled neurons from the area in the proximity of the optic fiber implant tip. This feature was consistent across animals and implantation sites, and is a relevant concern as it affects the accuracy of targeting small structures. In particular, such a feature could exacerbate limitations arising from heating artefacts, since the maximum SNR attainable at a particular level of photon fluence may be capped to an unnecessarily low level. This effect might be mitigated by using thinner optic fiber implants (e.g. ⌀200 µm, as opposed to the ⌀400 µm fibers used in this study).

Conclusion

In this article we demonstrate the suitability of opto-fMRI for investigating a neurotransmitter system which exhibits node-like function in coordinating brain activity. We present the first whole-brain map of VTA dopaminergic signalling in the mouse in a standard space aligned with stereotactic coordinates [46]. We determine that the mapping is consistent with known structural projections, and note the instances where differences are observed. Further, we explore network structure aware analysis via functional connectivity (fig. 3e), finding that the assay provides superior identification of the VTA, but limited support for signal relay imaging. In-depth investigation of experimental variation, based on open source and reusable workflows, supports the current findings by identifying detailed evidence-based instructions for assay reuse. Our study provides a reference dopaminergic stimulus-evoked functional neurophenotype map and a novel and thoroughly documented workflow for the preclinical imaging of dopaminergic function, both of which are crucial to elucidating the etiology of numerous disorders and improving psychopharmacological interventions in health and disease.

Funding

This work was funded by the Swiss National Science Foundation grant number 310030-179257, which was awarded to MR.

Author Contributions

HII performed the methods development, experiments, data analysis, and drafted the article. BJS consulted on methods development, provided materials, and reviewed the article. MR supervised the project, provided materials, consulted on MRI methods, and reviewed and edited the article.

Conflict of Interest

There are no financial or other relations that could lead to a conflict-of-interest.

References

[1] Matthew W Rice, Rosalinda C Roberts, Miguel Melendez-Ferro, and Emma Perez-Costas. Mapping dopaminergic deficiencies in the substantia nigra/ventral tegmental area in schizophrenia. *Brain Structure and Function*, 221(1):185–201, January 2016.

[2] Dwight C German and Kebreten F Manaye. Midbrain dopaminergic neurons (nuclei a8, a9, and a10): three-dimensional reconstruction in the rat. *Journal of Comparative Neurology*, 331(3):297–309, May 1993.

[3] LC Triarhou, J Norton, and Bernardino Ghetti. Mesencephalic dopamine cell deficit involves areas a8, a9 and a10 in weaver mutant mice. *Experimental brain research*, 70(2):256–265, April 1988.

[4] G. Di Chiara and A. Imperato. Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proceedings of the National Academy of Sciences*, 85(14):5274–5278, July 1988.

[5] Gaetano Di Chiara. Drug addiction as dopamine-dependent associative learning disorder. *European Journal of Pharmacology*, 375(1-3):13–30, June 1999.

[6] A Nieoullon. Dopamine and the regulation of cognition and attention. *Progress in Neurobiology*, 67(1):53–83, May 2002.

[7] John D. Salamone. The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behavioural Brain Research*, 61(2):117–133, April 1994.

[8] Soghra Akbari Chermahini and Bernhard Hommel. The (b)link between creativity and dopamine: Spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition*, 115(3):458–465, June 2010.

[9] Richard A. Depue and Paul F. Collins. Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences*, 22(3):491–517, June 1999.

[10] E. Masliah. Dopaminergic loss and inclusion body formation in -synuclein mice: Implications for neurodegenerative disorders. *Science*, 287(5456):1265–1269, February 2000.

[11] O. D. Howes and S. Kapur. The dopamine hypothesis of schizophrenia: Version III–the final common pathway. *Schizophrenia Bulletin*, 35(3):549–562, March 2009.

[12] Kei Yamamoto and Philippe Vernier. The evolution of dopamine systems in chordates. *Frontiers in Neuroanatomy*, 5, March 2011.

[13] Seth R. Taylor, Sylvia Badurek, Ralph J. Dileone, Raad Nashmi, Liliana Minichiello, and Marina R. Picciotto. GABAergic and glutamatergic efferents of the mouse ventral tegmental area. *Journal of Comparative Neurology*, 522(14):3308–3334, July 2014.

[14] Edward S Boyden, Feng Zhang, Ernst Bamberg, Georg Nagel, and Karl Deisseroth. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience*, 8(9):1263–1268, August 2005.

[15] P. C. Orban, D. Chui, and J. D. Marth. Tissue- and site-specific DNA recombination in transgenic mice. *Proceedings of the National Academy of Sciences*, 89(15):6861–6865, August 1992.

[16] M. Desai, I. Kahn, U. Knoblich, J. Bernstein, H. Atallah, A. Yang, et al. Mapping brain networks in awake mice using combined optical neural control and fMRI. *Journal of Neurophysiology*, 105(3):1393–1405, March 2011.

[17] Joanes Grandjean, Alberto Corcoba, Martin C. Kahn, A. Louise Upton, Evan S. Deneris, Erich Seifritz, et al. A brain-wide functional map of the serotonergic responses to acute stress and fluoxetine. *Nature Communications*, 10(1), January 2019.

[18] S Lohani, A J Poplawsky, S-G Kim, and B Moghaddam. Unexpected global impact of VTA dopamine neuron activation as measured by opto-fMRI. *Molecular Psychiatry*, 22(4):585–594, July 2016.

[19] Marta Brocka, Cornelia Helbing, Daniel Vincenz, Thomas Scherf, Dirk Montag, Jürgen Goldschmidt, et al. Contributions of dopaminergic and non-dopaminergic neurons to VTA-stimulation induced neurovascular responses in brain reward circuits. *NeuroImage*, 177:88–97, August 2018.

[20] Horea-Ioan Ioanas, Markus Marks, Valerio Zerbi, Mehmet Fatih Yanik, and Markus Rudin. An optimized registration workflow and standard geometric space for small animal brain imaging. 241:118386, November 2021.

[21] Xiaoxi Zhuang, Justine Masson, Jay A. Gingrich, Stephen Rayport, and René Hen. Targeted gene expression in dopamine and serotonin neurons of the mouse brain. *Journal of Neuroscience Methods*, 143(1):27–32, April 2005.

[22] Florian Aymanns, Markus Rudin, and Horea-Ioan Ioanas. COSplay: Contrast optimized stimulation player. *Journal of Open Source Software*, 4(39):1171, July 2019.

[23] Horea-Ioan Ioanas, Bechara Saab, and Markus Rudin. LabbookDB: A wet-work-tracking database application framework. *Proceedings of the 16th Python in Science Conference*, July 2017.

[24] Joanes Grandjean, Aileen Schroeter, Imene Batata, and Markus Rudin. Optimization of anesthesia protocol for resting-state fmri in mice based on differential effects of anesthetics on functional connectivity patterns. *NeuroImage*, 102 Pt 2:838–847, November 2014.

[25] John J.A. Marota, C. Ayata, Michael A. Moskowitz, Robert M. Weisskoff, Bruce R. Rosen, and Joseph B. Mandeville. Investigation of the early response to rat forepaw stimulation. *Magnetic Resonance in Medicine*, 41(2):247–252, February 1999.

[26] Horea-Ioan Ioanas, Markus Marks, Clément M Garin, Marc Dhenain, Mehmet Fatih Yanik, and Markus Rudin. An automated open-source workflow for standards-compliant integration of small animal magnetic resonance imaging data. *Frontiers in neuroinformatics*, 14:5, February 2020.

[27] Horea-Ioan Ioanas, Markus Marks, Tina Segessemann, Dominik Schmidt, Florian Aymanns, and Markus Rudin. SAMRI — Small Animal Magnetic Resonance Imaging, January 2019.

[28] Brian B. Avants, Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033–2044, February 2011.

[29] Horea-Ioan Ioanas, Tina Segessemann, and Markus Rudin. "Mouse Brain Atlases" generator workflows, January 2019.

[30] A.E. Dorr, J.P. Lerch, S. Spring, N. Kabani, and R.M. Henkelman. High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult c57bl/6j mice. *NeuroImage*, 42(1):60–69, August 2008.

[31] Robert W Cox. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, June 1996.

987

[32] Mark Jenkinson, Christian F. Beckmann, Timothy E J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *Neuroimage*, 62(2):782–790, August 2012.

[33] Patrick E. Steadman, Jacob Ellegood, Kamila U. Szulc, Daniel H. Turnbull, Alexandra L. Joyner, R. Mark Henkelman, et al. Genetic effects on cerebellar structure across mouse models of autism using a magnetic resonance imaging atlas. *Autism Research*, 7(1):124–137, October 2013.

[34] Jeremy F.P. Ullmann, Charles Watson, Andrew L. Janke, Nyoman D. Kurniawan, and David C. Reutens. A segmentation protocol and MRI atlas of the c57bl/6j mouse neocortex. *NeuroImage*, 78:196–203, September 2013.

[35] Kay Richards, Charles Watson, Rachel F. Buckley, Nyoman D. Kurniawan, Zhengyi Yang, Marianne D. Keller, et al. Segmentation of the mouse hippocampal formation in magnetic resonance images. *NeuroImage*, 58(3):732–740, October 2011.

[36] Matthew Brett, Michael Hanke, Chris Markiewicz, Marc-Alexandre Côté, Paul McCarthy, Chris Cheng, et al. nipy/nibabel: 2.3.1, October 2018.

[37] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, et al. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8, February 2014.

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, October 2011.

[39] David W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, December 1979.

[40] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with Python. In *9th Python in Science Conference*. June 2010.

[41] Travis E. Oliphant. Python for scientific computing. *Computing in Science & Engineering*, 9(3):10–20, June 2007.

[42] J Scott Long and Laurie H Ervin. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224, August 2000.

[43] Seung Wook Oh, Julie A Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, et al. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207, April 2014.

[44] Tina Segessemann, Markus Rudin, and Horea-Ioan Ioanas. Abi connectivity data package generator, October 2019.

[45] Horea-Ioan Ioanas, Bechara Saab, and Markus Rudin. Gentoo linux for neuroscience - a replicable, flexible, scalable, rolling-release environment that provides direct access to development software. *Research Ideas and Outcomes*, 3:e12095, February 2017.

[46] Horea-Ioan Ioanas, Bechara John Saab, and Markus Rudin. Source Code for "A Whole-Brain Map and Assay Parameter Analysis of Mouse VTA Dopaminergic Activation", June 2019.

[47] Krzysztof J Gorgolewski, Tibor Auer, Vince D Calhoun, R Cameron Craddock, Samir Das, Eugene P Duff, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3:160044, June 2016.

[48] Horea-Ioan Ioanas, Bechara John Saab, and Markus Rudin. BIDS Data for "A Whole-Brain Map and Assay Parameter Analysis of Mouse VTA Dopaminergic Activation", June 2019.

[49] Horea-Ioan Ioanas and Markus Rudin. Reproducible self-publishing for Python-based research. EuroSciPy, August 2018.

[50] Weixing X. Pan, Tianyi Mao, and Joshua T. Dudman. Inputs to the dorsal striatum of the mouse reflect the parallel circuit architecture of the forebrain. *Frontiers in Neuroanatomy*, 4, December 2010.

[51] E. E. Benarroch. Periaqueductal gray: An interface for behavioral control. *Neurology*, 78(3):210–217, January 2012.

[52] Ana Aransay, Claudia Rodríguez-López, María García-Amado, Francisco Clascá, and Lucía Prensa. Long-range projection neurons of the mouse ventral tegmental area: a single-cell axon tracing analysis. *Frontiers in Neuroanatomy*, 9, May 2015.

[53] Howard L. Fields, Gregory O. Hjelmstad, Elyssa B. Margolis, and Saleem M. Nicola. Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annual Review of Neuroscience*, 30(1):289–316, July 2007.

[54] Satoshi Ikemoto. Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens–olfactory tubercle complex. *Brain Research Reviews*, 56(1):27–78, November 2007.

[55] Gonzalo E. Torres, Raul R. Gainetdinov, and Marc G. Caron. Plasma membrane monoamine transporters: structure, regulation and function. *Nature Reviews Neuroscience*, 4(1):13–25, January 2003.

989

Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

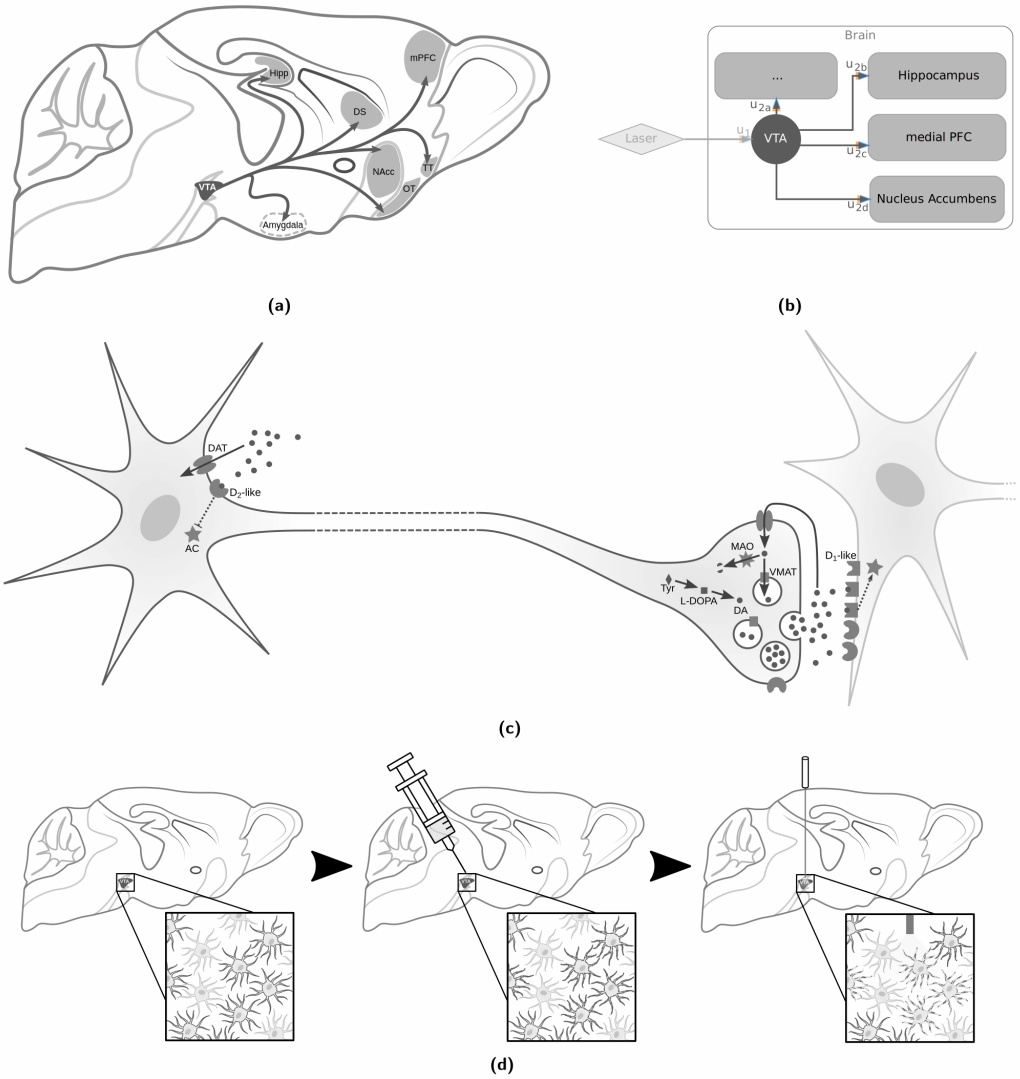


Figure 1: The cell biological compartmentalization of dopaminergic neurotransmission (and susceptibility to psychopharmacology) can partly be mapped onto neuroanatomical features by a simple network model, using optogenetics. Depicted are schematic overviews of the VTA dopaminergic system at various spatial resolutions. **(a)** Schematic map of VTA dopaminergic projections [52, 53, 54, 50]. Dotted structures are off-slice, and projection arrows do not reflect actual fiber bundle paths. **(b)** Simplified network model of 1-step signal relay following optogenetic stimulation of the VTA. The u_1 weighting corresponds to VTA somatic excitability and u_{2a}, u_{2b}, u_{2c} and u_{2d} correspond to transmission at the dopaminergic synapses in the respective projection areas. **(c)** Schematic overview of VTA dopaminergic neurons, with the soma located in the VTA and synapses in one or multiple other projection area voxels. Excitability at the soma are contingent on D_2 autoinhibition, while transmission at the synapse is contingent on dopamine metabolism, turnover, and postsynaptic D_1 expression. [55]. **(d)** Schematic of optogenetic cell selection and activation. Orange denotes dopaminergic cells, gray enlarged elements on the cell periphery indicate channelrhodopsin expression, and cyan segments on the cell periphery denote depolarization events. **Abbreviations:** AC (adenylyl cyclase), DA (dopamine), DAT (dopamine transporter), DS (Dorsal Striatum), Hipp (Hippocampus), MAO (monoamine oxydase), mPFC (medial Prefrontal Cortex), NAcc (Nucleus Accumbens), OT (Olfactory Tuberculum), TT (Tenia Tecta), Tyr (tyrosine), VMAT (vesicular monoamine transporter).

990

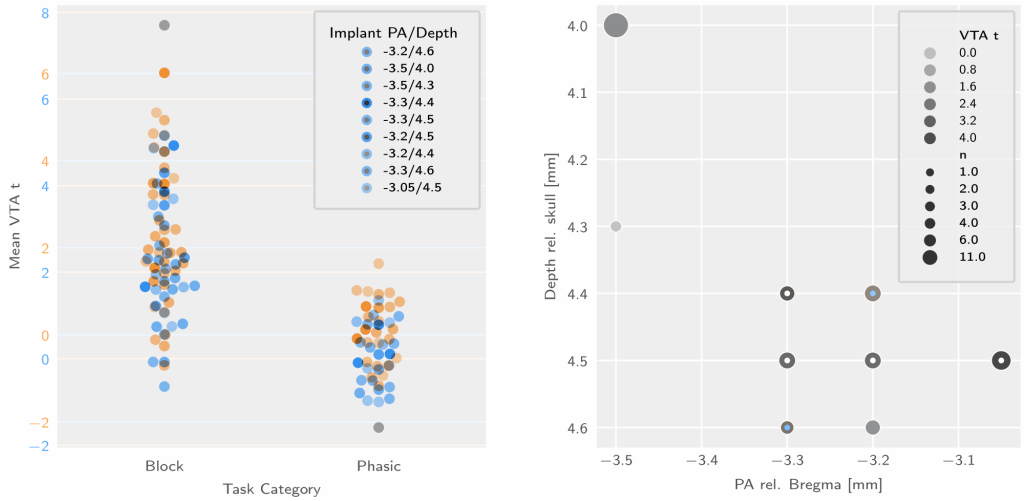


Figure 2: VTA activation is sensitive to the stimulation protocol category and the implant coordinates, with different trends in block and phasic stimulation trials. Depicted are multifactorial (protocol and implant coordinates) comparisons of signal intensity in the VTA region of interest. **(a)** Tavivisk group comparison for animals targeted at all explored combinations of implant coordinates. **(b)** Implant coordinate comparison for block stimulation trials (inner dots indicate best category group). **Abbreviations:** n (sample size), PA (posteroanterior), rel. (relative to).

991

Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

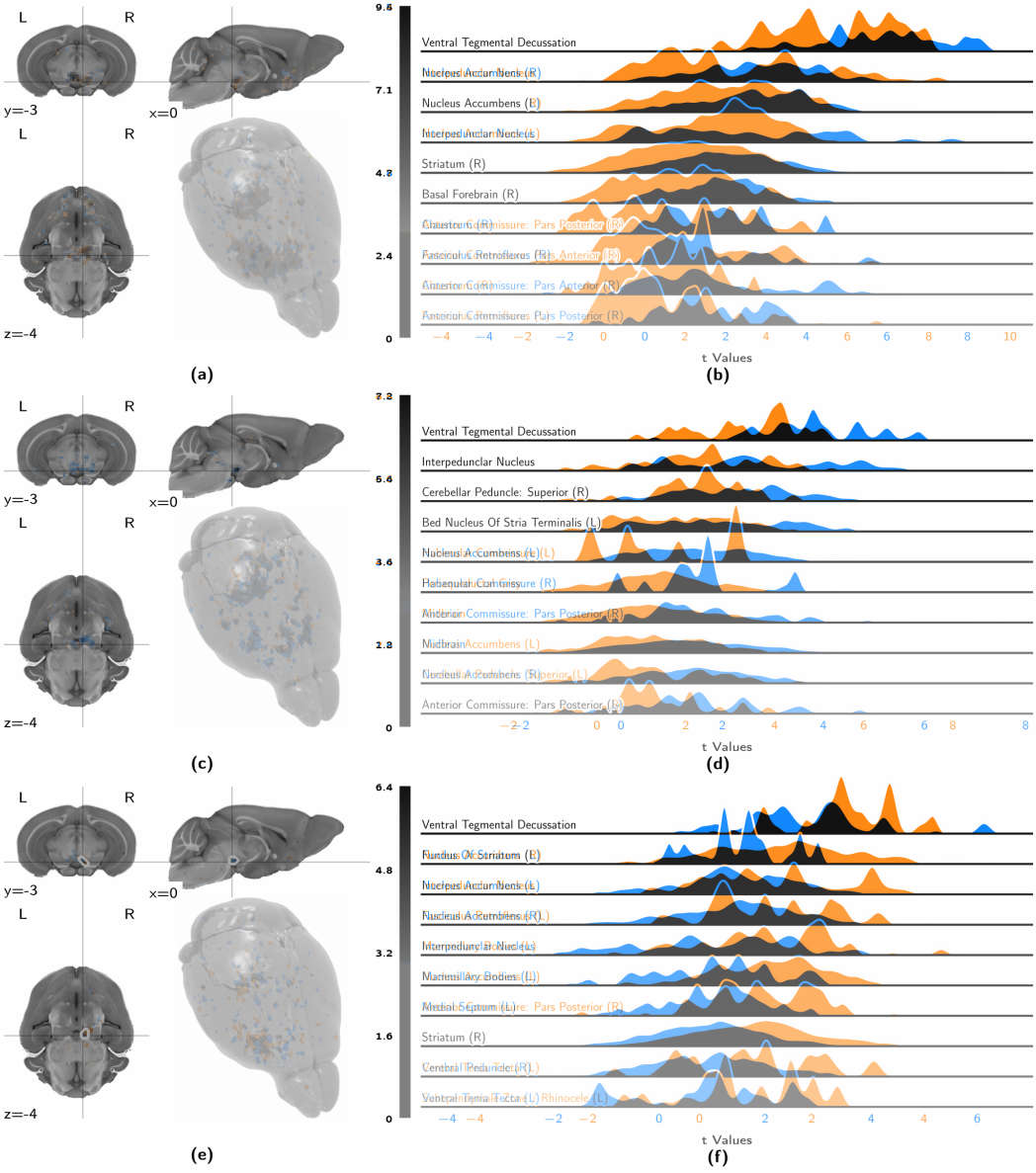


Figure 3: Block stimulation elicits strong ventral striatal activity in the best implant group, more rostrally weighted activity in the rejected implant group, and generates similar but weaker contrasts for VTA seed-based analysis. The figures show volumetric population t-statistic maps **(a, e, c)** thresholded at $t \geq 3$ and centered on the VTA target, as well as a break-down of activation along atlas parcellation regions **(b, d, f)**. **(a)** Second-level t-statistic map for block-stimulus-evoked activity in best implant group animals (corrected for the wild type control response). **(b)** Distribution densities of statistic values from block-stimulus-evoked activity analysis in best implant group animals (corrected for the wildtype control response). Depicted are the 10 most strongly activated areas. **(c)** Second-level t-statistic map for block-stimulus-evoked activity in rejected implant group animals (corrected for the wild type control response). **(d)** Distribution densities of statistic values from block-stimulus-evoked activity analysis in rejected implant group animals (corrected for the wild type control response). Depicted are the 10 most strongly activated areas. **(e)** Second-level t-statistic map for VTA seed-based functional connectivity during block stimulation in best implant group animals (VTA region in green). **(f)** Distribution densities of statistic values from seed-based functional connectivity analysis of best implant group animal block stimulation scans. Depicted are the 10 most strongly activated areas.

992

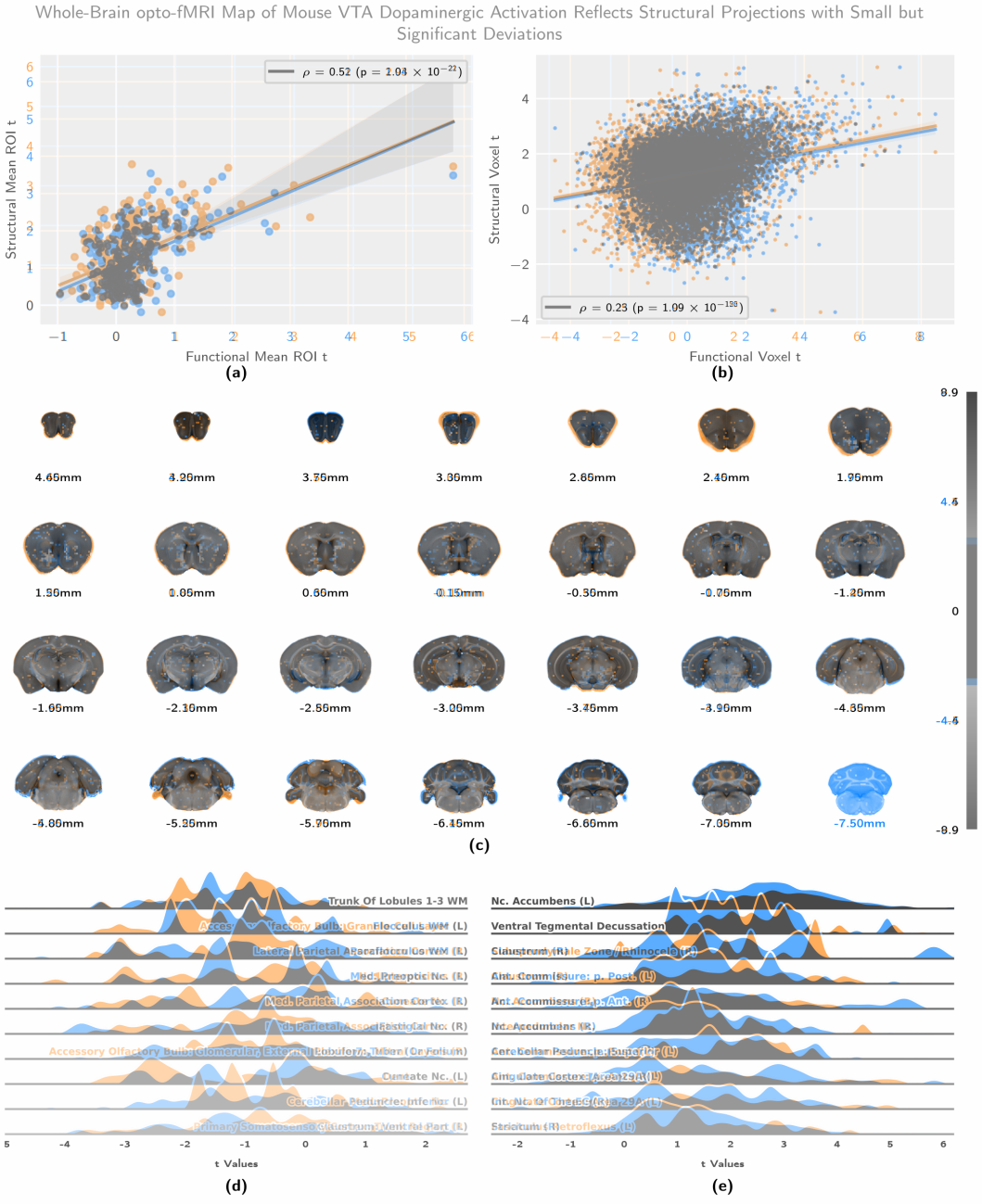
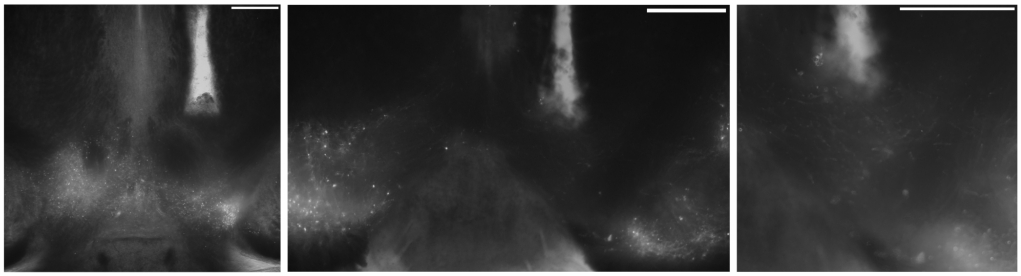


Figure 4: Comparing VTA functional activation to structural projection data reveals good correspondence, with deviations involving the dorsomedial striatum and the contralateral ventral striatum. Depicted are correlation analyses **(a, b)** of the population-level functional and structural statistic scores, alongside statistic distributions **(c, d, e)** for the contrast, taking into account variability across subjects. **(a)** Region-wise regression plot between functional and structural projection maps. Tinted area indicates the 99 % confidence interval of the regression estimate. **(b)** Voxel-wise regression plot between functional and structural projection maps. Tinted area indicates the 99 % confidence interval of the regression estimate. **(c)** Coronal slices, showing the population-level contrast t-statistic between VTA functional activation and VTA structural projections. **(d)** Distribution densities of t-statistics, showing the regions where VTA structural projection exceeds functional activation most strongly. **(e)** Distribution densities of t-statistics, showing the regions where VTA functional activation exceeds structural projection most strongly. **Abbreviations:** Ant. (Anterior), EC (Endopiriform Claustrum), Int. (Intermediate), Med. (Medial), Nc. (Nucleus), p. (Pars), Post. (Posterior), WM (White Matter).

993



(a) **(b)** **(c)**

Figure 5: Fiber implantation causes strong local cell displacement in the VTA. Depicted are YFP (coexpressed with Channelrhodopsin) fluorescent microscopy images of the VTA, overlaid on corresponding transmission microscopy images of the same focal plane. All slices are seen in neurological orientation (the right of the image corresponds to the right of the animal). A higher magnification of **(b)** is depicted in **(c)**. White bars indicate a scale of 500 µm, and slices are shown in neurological orientation. **(a)** 3.05 mm caudal of Bregma. **(b)** 3.5 mm caudal of Bregma. **(c)** 3.5 mm caudal of Bregma.

Supplementary Materials

Onset [s]	Duration [s]	Frequency [Hz]	Pulse Width [s]
182.0	20.0	20.0	0.005
332.0	20.0	20.0	0.005
482.0	20.0	20.0	0.005
632.0	20.0	20.0	0.005
782.0	20.0	20.0	0.005
932.0	20.0	20.0	0.005
1082.0	20.0	20.0	0.005
1232.0	20.0	20.0	0.005

Table S1: Block stimulation protocol, coded "CogB".

Onset [s]	Duration [s]	Frequency [Hz]	Pulse Width [s]
180.0	20.0	20.0	0.005
310.0	20.0	20.0	0.005
480.0	20.0	20.0	0.005
630.0	20.0	20.0	0.005
780.0	20.0	20.0	0.005
930.0	20.0	20.0	0.005
1080.0	20.0	20.0	0.005
1230.0	20.0	20.0	0.005

Table S2: Block stimulation protocol, coded "CogBr".

Onset [s]	Duration [s]	Frequency [Hz]	Pulse Width [s]
192.0	30.0	20.0	0.005
342.0	30.0	20.0	0.005
492.0	30.0	20.0	0.005
642.0	30.0	20.0	0.005
792.0	30.0	20.0	0.005
942.0	30.0	20.0	0.005
1092.0	30.0	20.0	0.005
1242.0	30.0	20.0	0.005

Table S3: Block stimulation protocol, coded "CogBl".

Onset [s]	Duration [s]	Frequency [Hz]	Pulse Width [s]
180.0	8.0	20.0	0.005
330.0	10.0	20.0	0.005
480.0	12.0	20.0	0.005
630.0	14.0	20.0	0.005
780.0	16.0	20.0	0.005
930.0	28.0	20.0	0.005
1080.0	20.0	20.0	0.005
1230.0	22.0	20.0	0.005

Table S4: Block stimulation protocol, coded "CogBm".

Onset [s]	Duration [s]	Frequency [Hz]	Pulse Width [s]
150.0000	20.0	15.0	0.005
280.0000	20.0	25.0	0.005
410.0000	20.0	15.0	0.010
540.0000	20.0	25.0	0.010
670.0000	20.0	15.0	0.005
799.9999	20.0	25.0	0.005
930.0000	20.0	15.0	0.010
1060.0000	20.0	25.0	0.010
1190.0000	20.0	15.0	0.005
1320.0000	20.0	25.0	0.005
1450.0000	20.0	15.0	0.010
1580.0000	20.0	25.0	0.010

Table S5: Block stimulation protocol, coded "CogMwf".

Onset [s]	Duration [s]	Frequency [Hz]	Pulse Width [s]
190.0	0.8	25.0	0.005
192.0	0.8	25.0	0.005
194.0	0.8	25.0	0.005
196.0	0.8	25.0	0.005
290.0	0.8	25.0	0.005
292.0	0.8	25.0	0.005
294.0	0.8	25.0	0.005
296.0	0.8	25.0	0.005
390.0	0.8	25.0	0.005
392.0	0.8	25.0	0.005
394.0	0.8	25.0	0.005
396.0	0.8	25.0	0.005
490.0	0.8	25.0	0.005
492.0	0.8	25.0	0.005
494.0	0.8	25.0	0.005
496.0	0.8	25.0	0.005
590.0	0.8	25.0	0.005
592.0	0.8	25.0	0.005
594.0	0.8	25.0	0.005
596.0	0.8	25.0	0.005

Table S6: Phasic stimulation protocol, coded "CogP".

Onset [s]	Duration [s]	Frequency [Hz]	Pulse Width [s]
50.0	1.0	20.0	0.005
90.0	1.0	20.0	0.005
130.0	1.0	20.0	0.005
170.0	1.0	20.0	0.005
210.0	1.0	20.0	0.005
250.0	1.0	20.0	0.005
290.0	1.0	20.0	0.005
330.0	1.0	20.0	0.005
370.0	1.0	20.0	0.005
410.0	1.0	20.0	0.005
450.0	1.0	20.0	0.005
490.0	1.0	20.0	0.005
530.0	1.0	20.0	0.005
570.0	1.0	20.0	0.005
610.0	1.0	20.0	0.005

Table S7: Phasic stimulation protocol, coded "JPogP".

995

In a linear modelling of the implant coordinate variables, the VTA mean t statistic is found sensitive only to the stimulation protocol category ($F_{1,59} = 57.31$, $p = 2.22 \times 10^{-9}$), but not to the stimulation target depth ($F_{1,59} = 0.48$, $p = 0.49$), the stimulation target posterior-anterior (PA) coordinates ($F_{1,59} = 0.59$, $p = 0.37$), and the interaction of the depth and PA target coordinates ($F_{1,59} = 0.48$, $p = 0.49$).

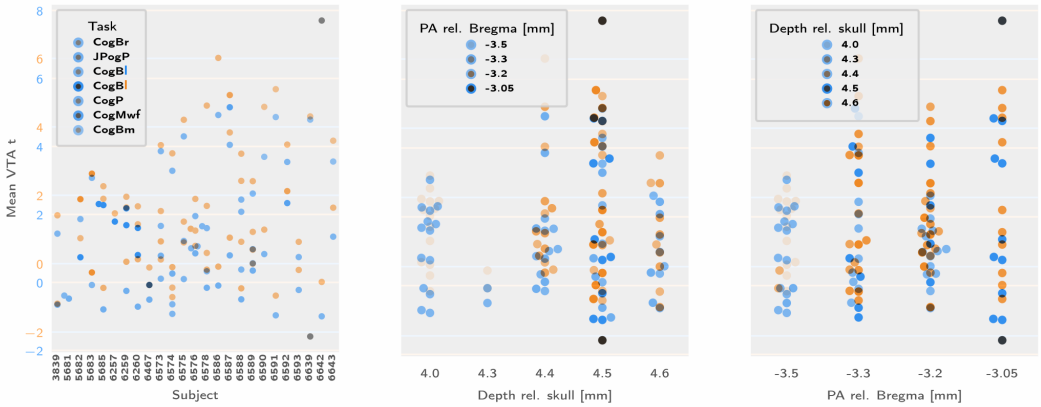
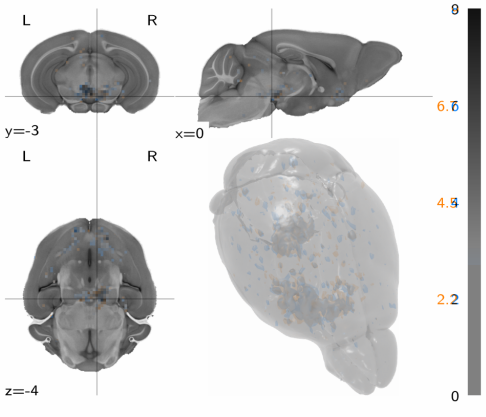
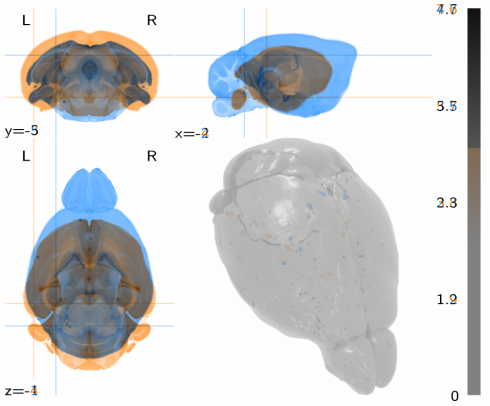


Figure S1: Multifactorial (depth and posteroantior) implant coordinate comparisons of signal intensity in the VTA region of interest. Protocols coded as in tables S1 to S7.



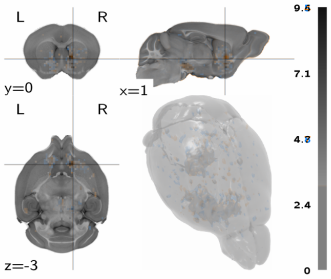
(a) Block stimulation, slices centered on largest cluster.



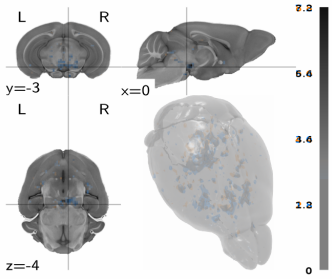
(b) Phasic stimulation, slices centered on largest cluster.

Figure S2: No negative activation patterns are salient upon block VTA stimulation, and no coherent activation patterns of any sort after phasic VTA stimulation. Depicted are t-statistic maps (thresholded at $|t| \geq 3$) of second-level analyses, divided by stimulation category and binning all implant coordinates. Slices are centered on the VTA coordinates (RAS $= 0.5/-3.2/-4.5$) and the largest cluster, respectively. All maps are adjusted for the wild type control stimulation effects.

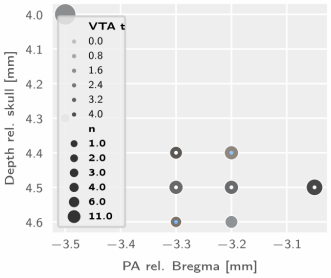
997



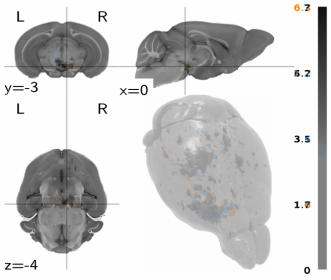
(a) Block stimulation of best implant category group (according to VTA activation), centered on largest cluster.



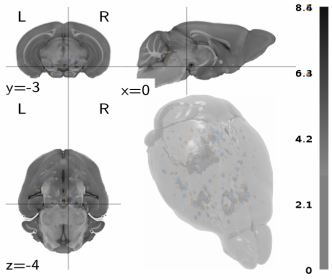
(b) Block stimulation of rejected implant category group (according to VTA activation), centered on VTA.



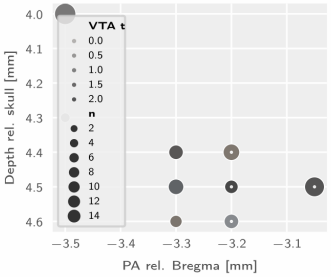
(c) VTA activation-based implant coordinate classification (dotted markers indicate best category).



(d) Block stimulation of rostralmost implant category group, centered on VTA.

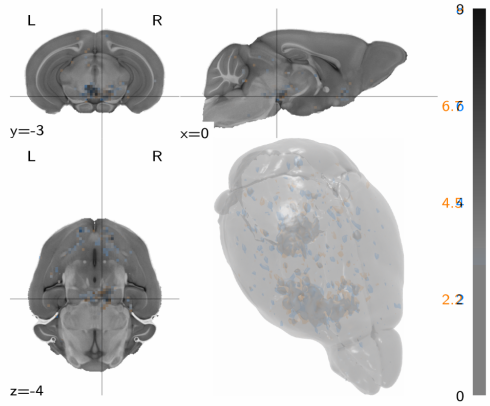


(e) Block stimulation of caudalmost implant category group, centered on largest cluster.

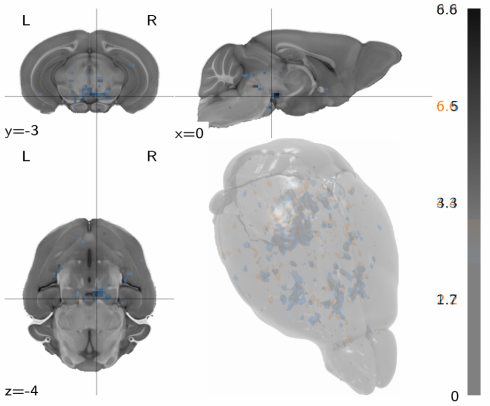


(f) PA-coordinate implant classification (dotted markers indicate rostralmost category).

Figure S3: PA-coordinate-based classification does not show a better projection segmentation than block trial-based classification. Depicted are t-statistic maps (centerd on largest cluster, thresholded at $t \geq 3$) of the second-level analysis for block stimulation protocols, divided into best and rejected **(a, b)**, or rostralmost and caudalmost **(d, e)**. All maps are adjusted for the wild type control stimulation effects.



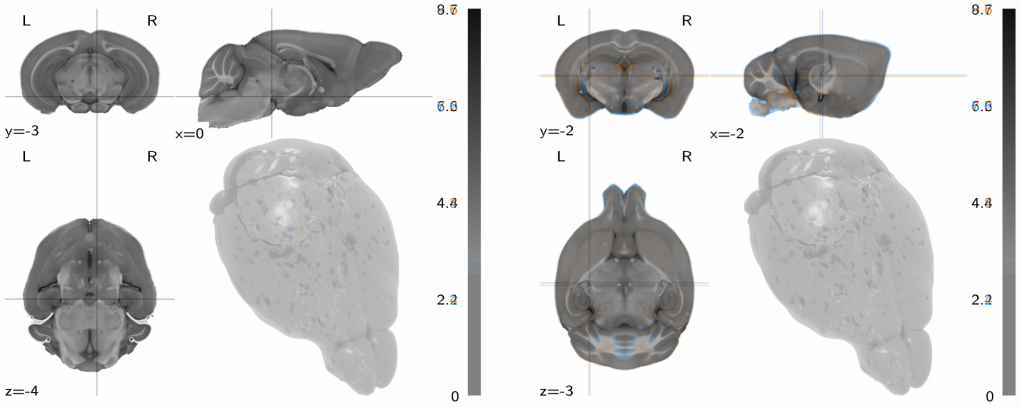
(a) Block stimulation of best implant category group



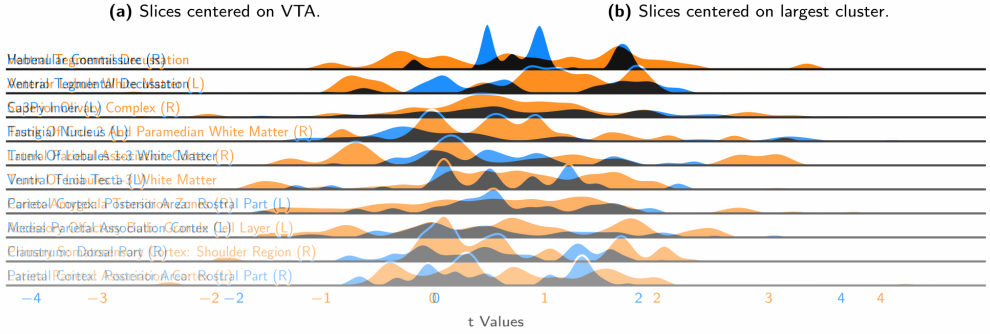
(b) Block stimulation of rejected implant category group.

Figure S4: The uncorrected population-level response to block stimulation does not qualitatively differ from the wild type control corrected results in figs. 3a and 3c. Depicted are wildtype-control-uncorrected t-statistic maps (thresholded at $t \geq 3$) of the second-level analysis for block stimulation protocols, divided by implant category group. Slices are centered on the VTA region of interest.

Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

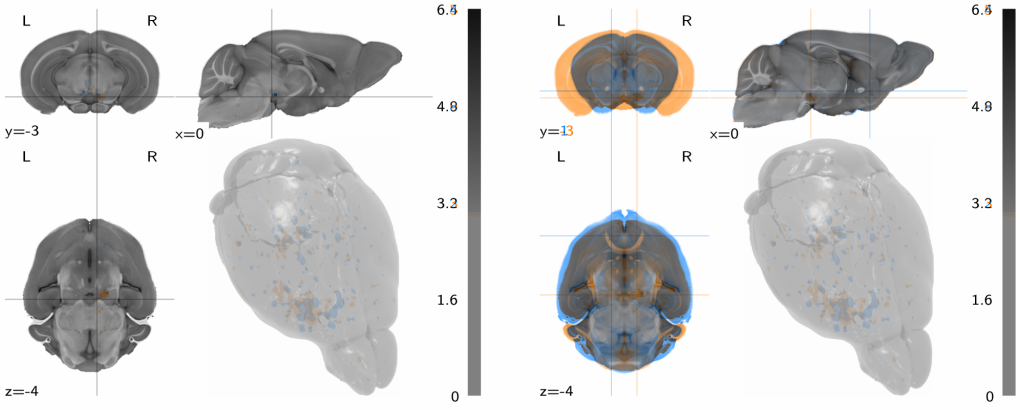


(a) Slices centered on VTA. **(b)** Slices centered on largest cluster.



(c) Distribution densities of t-statistic values in the 10 most strongly activated areas.

Figure S5: Block stimulation in wild type control animals produces no large activation clusters, yet scattered activation hints at some visual excitation and heating artefacts. Depicted are volumetric population t-statistic maps **(a, b)** — thresholded at $t \geq 3$, as well as a break-down of activation along atlas parcellation regions **(c)**.



(a) Slices centered on VTA. **(b)** Slices centered on largest cluster.

Figure S6: Depicted are t-statistic maps (thresholded at $t \geq 3$) of the second-level analysis for block stimulation task VTA seed functional connectivity, observed in the best implant category, corrected for the negative control baseline. Slices are centered on the VTA coordinates (RAS $= 0.5/-3.2/-4.5$) and the largest cluster, respectively. This comparison is only provided for the sake of completeness and analogy with the stimulus-evoked analysis. Conceptually this comparison is not of primary interest, since seed-based functional connectivity attempts to include precisely the baseline functioning of the system into the evaluation.

999

Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

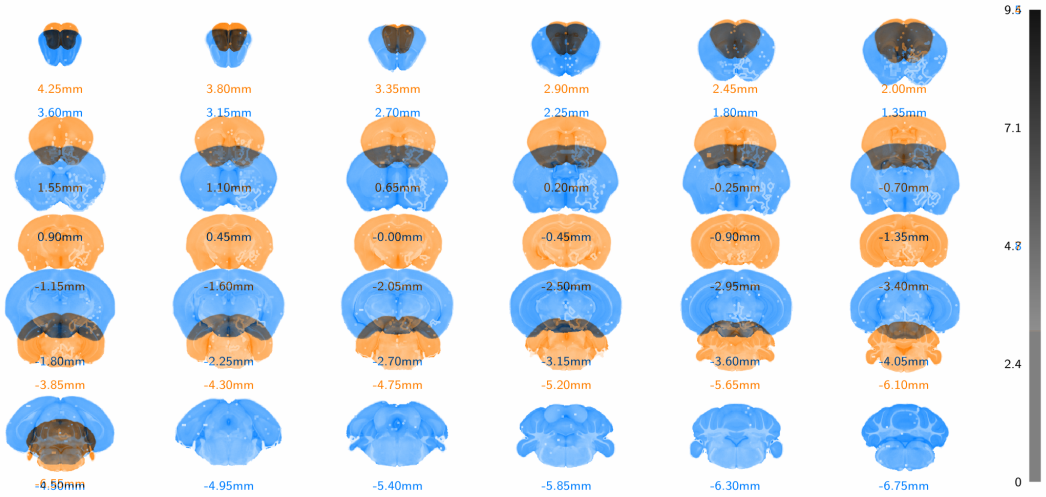


Figure S7: Coronal slice overlay, showing the VTA functional activation t-statistic heatmap (as in **fig. 3a**), and the VTA structural projection outline, both thresholded at $t \geq 3$. Interpretation of this figure as showcasing a complementarity in the patterns is cautioned, as qualitative inspection of thresholded data does not accurately capture variation in the statistic distributions. For statements regarding the compariosn of functional activation and structural projection, figs. 4a to 4c are more suitable.