

# Dissertation Proposal

Jonny Saunders

February 9, 2022 - Version: draft-3-g4388b19

---

Requirements:

*“defines the plan formulated by the student and the committee regarding (at least):*

- *the topic and scope of the dissertation,*
  - *method and scope of data collection, and*
  - *analysis strategies.*
- 

I see myself as a transdisciplinary researcher: sprawled across worlds, but following a common thread. The closest I’ve come to a simple description of my focus is *communicative systems*: how agents or components interact to create systems of meaning that, reciprocally, structure their interaction. I also believe that the practice of science itself is deeply political, reflective of the systems of belief and power interacting across interlocking institutions and organizations, and have increasingly oriented my work towards understanding and directly counteracting the ways that neoliberalization and commercial capture of science poison every part of our work.

This dissertation will be in three parts, roughly tracing the development of my thoughts over the past seven years. My work here starts and ends with studying a mouse model of **auditory phonetic perception**, comprising behavioral neuroscience, an attempted merger of several strains of linguistic, philosophical, and neuroscientific thought, and in the final few months a touch of neurophysiology. That led, perhaps unwisely, to the development of **Autopilot**: a Python framework for behavioral neuroscientific experiments based in distributed computing, but learning lessons from history and critical technology scholars is mutating into something more like a technical knowledge organization system. Finally in my last year my work has focused on the broader **infrastructure and practice** of science, and weaving between many fields have tried to plot a course away from the subjugation of basic science to surveillance and information capitalism.

## I Phonetic Perception

### I.1 Mice Can Learn Phonetic Categories

I started my work by teaching mice to generalizably discriminate between two sets of naturally produced pitch-shifted consonant-vowel pairs [1]. The idea was to establish phonetic perception in mice as a model that would be useful for twin problems in auditory neuroscience and phonetics: Auditory neuroscientists need better models to move beyond simple stimuli like tone pips and noise to understand how the mammalian auditory system processes the complexity that defines natural sound environments. The long history of phonetics research provides auditory neuroscience with a rich body of work, largely ungrounded in neurophysiological observation, that can inform our experimental designs and structure our predictions. Reciprocally, animal models can augment phonetic research by isolate the specifically auditory component of phonetic perception from the syntactic, semantic, motor, and social influences intrinsic to human perception; and mice provide access to observe and perturb neurophysiological processes unavailable in humans.

The experiment consisted of a two-alternative forced choice task in an arena with three water ports with photosensors to detect when they poked their nose in them. The mice would poke their nose in the center port to hear a consonant-vowel (/CV/) pair beginning with a /b/ or /g/, and then were rewarded if they poked their nose in the correct flanking port corresponding to either consonant. They progressed through several shaping stages that expanded the number of possible /CV/ tokens, adding new speakers and vowels, until they could reliably categorize 20 tokens. The mice then were tested in a generalization task where on a small subset of trials the mice were presented novel tokens. We characterize generalization as the measure of having learned an abstract or at least adaptive notion of the consonant structure, rather than overlearning the training set. We trained two cohorts of mice with different sets of speakers to test whether and how strongly that patterned their responses (vs. learning some “universal” means of categorizing the tokens).

The primary analysis in the paper used a generalized linear mixed model with a logistic link function to predict binary correct/incorrect responses using the type of token (training tokens, generalization tokens, etc.) as a fixed effect nested within each mouse as a random effect. From this we estimated the relative difficulties of different kinds of generalization (eg. to novel speakers vs. novel vowels), and differences in accuracy patterns across mice and cohorts. We tested the predictions of one neurolinguistic model (“Locus equations”) and found no behavioral support for them. Finally we lesioned auditory cortex and found that it returned the mice to chance accuracy – indicating that cortex is at least involved in the computation.

## 1.2 Models & Mechanisms

Given our behavioral results, the next step, had I not taken the long way around by building Autopilot, would be to see how the brain learns and processes the phoneme pairs. Systems neuroscience is in the midst of a quiet methodological upheaval, moving beyond characterizing the properties of individual neurons sampled in isolation to modeling the brain as a dynamic organ whose population activity cannot be extrapolated from its individual components. This move pairs nicely with lessons learned from methodological crises in Psychology and related fields that emphasize iterative formal modeling to inform experimental design and interpretation (eg. [2]). In this case, we can develop formal models to evaluate observed population activity in auditory cortex by working from the broad and multidisciplinary history of phonetic perception research.

Historically, the work of the Haskins lab (eg. [3]), and their acoustic “cue discovery” paradigm (see [4, p. 51]) characterized auditory processing as some mapping function

$$M = f : \mathbf{S} \rightarrow \mathbf{P} \quad (1)$$

from some stimulus space  $\mathbf{S}$  to a fixed [5] perceptual space  $\mathbf{P}$  composed of innate auditory “cues.” The auditory act of phonetic perception, then, is determining the phoneme category  $c_s \in \mathbf{C}$ , the set of possible phonemes in a language, given some perceptual representation  $\mathbf{p}$  such that

$$\mathbf{p} = M(\mathbf{s}) \quad (2)$$

$$c_s = \max(\{p(c_i|\mathbf{p}) : c_i \in \mathbf{C}\}) \quad (3)$$

Finding no set of cues that uniquely identifies a phoneme, they concluded that phonetic perception cannot be purely auditory and instead arrive at perceptual theory that relies on simulating the motor activity of an utterance [5].

Each of the components of this very simple model has been problematized in parallel branches of research: developmental psychologists like Patricia Kuhl studying infant speech acquisition show the clear plasticity of the auditory system [6], de-

scribing a “warped” perceptual space  $\mathbf{P}$  that takes advantage of the “big cuts” of the mammalian auditory system to basic sound features like frequency and energy changes. Learning a language’s phonemes could then be characterized as learning some weight vector  $\mathbf{w}$  as a function of the history of perceptual representations of speech exposure  $\mathbf{p}$  and the category structure of the language  $\mathbf{C}$  that maximizes the separability of the phonetic categories in perceptual space.

$$\mathbf{w} = W(\mathbf{p}, \mathbf{C}) \quad (4)$$

$$c_s = \max(\{p(c_i | \mathbf{p} \cdot \mathbf{w}) : c_i \in \mathbf{C}\}) \quad (5)$$

Cognitive psychologists like Eleanor Rosch studying the nature of cognitive categories emphasize their ill-definedness, that they resemble a Wittgensteinian “game” that does not require a neat mapping from a perceptual space to phonetic category identity[7, 8]. Rather than a “flat” cue space where all cues are evaluated simultaneously, phonologists have described a hierarchy of contrastive features where “lower” cues are evaluated conditionally on the value of “higher” cues[9]. Another perspective dismisses the notion of cues altogether, focusing on the informational content in the temporal dynamics of the whole speech signal[10].

These theories are more compatible with the deeply nonlinear computation of the brain than a simple linear evaluation of cues. The neuroscientific perspective emphasizes the active nature of perception, where the auditory system is not some passive filterbank, but a plastic organ that adapts over multiple timescales to the statistics of its input[11, 12]. This shifts the focus from learning some pre-existing cues to the process by which the brain adapts to create a joint neural-perceptual space that can serve as a flexible basis set for phonetic categorization.

Though I won’t be able to see this experiment through to its completion, in my last months here, I will be working with my labmates to start mesoscopic calcium imaging, measuring a proxy for neural activity, across the course of learning the phonetic discrimination task above. Imaging across the course of learning will let us not only model the computation of phonetic perception as it happens, but how the brain learns to support it. I won’t be able to see the experiment to completion, but I will be responsible for developing the hardware and software to perform the behavior in a head-fixed context, formalizing our models and analytical techniques by reading across the array of disciplines that have studied phonetic perception, and then will be leaving the imaging and data collection to my labmates Sam Mehan and Rocky Penick.

## 2 Autopilot

In our struggle through implementing the speech discrimination task, I learned that the state of neuroscientific behavioral tooling had a big hole in it where an extensible, inexpensive, and distributed framework for behavioral experiments should be — so I have spent the last five years trying to fill it[13]. Autopilot is a software package that also attempts to contend with the social realities of experimental work. To quote the [readme](#):

As a tool, it provides researchers with a toolkit of flexible modules to design experiments without rigid programming & API limitations. As a vision, it dreams of bridging the chaotic hacky creativity of scientific programmers with a standardized, communally developed library of reproducible experiment prototypes.

There is far too much work that has gone into Autopilot to summarize here, but the essence of the design is that by using many small single-board computers that run a full operating system and a high-level language like python<sup>1</sup> that can interact

<sup>1</sup>as opposed to a microcontroller that runs code compiled/generated by the user-facing part of the software, which is the norm in engineering

with fast, low-level compiled libraries you can get all the speed you need while keeping the entire library within reach of anyone who uses it. By using the Raspberry Pi I wanted to allow people to use whatever off-the-shelf components they wanted to, rather than restricting it to a **small number of bespoke parts**, and be able to rely on and contribute to the large community of raspi users.

It's also designed with the knowledge that experimental software is often the first code that many researchers experience, and aims to scaffold that process and give good examples of code structure — a process that is under continual development as I myself was learning to program while writing it, and only feel like I've reached a point of being comfort with best principles recently. Rather than a tool that's a black-box behind the scenes, Autopilot is intended to be understood at whatever level you want to engage with it: if you want to just use it as it is, that's fine, but if you want to dig into it the documentation tries to be exhaustive (and is currently a **336 page book**, though the **HTML docs** are considerably more manageable).

Autopilot is **many things**, a sample of which for illustration include:

- a library of **hardware** controllers, including **cameras**, **GPIO**-powered components, a few **sensors** like a 9-degree motion sensor and a 2D heat sensor, and a few more that I haven't had time to add yet.
- **Networking** modules that make it (increasingly) straightforward to link anything to anything else across computers, stream data, etc.
- A realtime **sound server** with millisecond latency and microsecond timestamp precision, as well as a collection of simple to use **sound** generation classes
- A system for writing behavioral **tasks**, including for experiments that need multiple shaping stages with programmatic **graduation** between them, and the ability to describe reusable, adaptable, and extensible **task prototypes** to avoid writing everything from scratch
- A system of chainable realtime data **transformations** that let you compose common **geometric** algorithms, filters, unit conversions, and computer vision operations into reproducible pipelines that can be **programmatically reconstructed** on any machine. I also had the pleasure of collaborating with the DeepLabCut[14] team to incorporate realtime markerless pose tracking across the network into Autopilot[15]
- An extremely permissive **plugin** system that lets people extend many parts of Autopilot and incorporate it naturally with the rest of the system without needing to directly modify the main library. In the next planned **major version** "many parts" will become "all parts" and allow users to replace, rather than extend them as well.
- A means of preserving exhaustive experimental **provenance**, including every parameter change in every task that an experimental subject ever experiences and all the versions of the code that ran them, making exact experimental replication possible from a single file. Autopilot's separation of the logic of a task from its hardware implementation also makes replication possible without having exactly the same instrumentation.
- A collection of tools to **setup** and configure various parts of the occasionally-cantankerous raspi
- A **GUI** that is at the moment an abomination in need of repair, and a **TUI**
- A densely-linked, semantically tagged **wiki** that integrates technical knowledge about hardware, task design, etc. with the software to use it.
- Dozens of my own hardware designs including **3D CAD**, **2D CAD**, and a few **circuit boards**; along with a few (at the moment incomplete) **guides** to build them

All of which I did myself starting from zero programming experience at the start of the program. I'm quite proud of it.

> DLC collaboration [15]

> Wiki and the future

## 2.1 Interlude: The People’s Ventilator Project

> Description of project, the acute need for scientists to use their expertise for social good, and continuation of work from Autopilot [16]

## 3 Infrastructure

Through my work in Autopilot I became acutely aware of the social dynamics that define infrastructural development. While there is subconscious belief that scientific technology will get, in fits and starts, “better” over time, instead it became abundantly obvious that without specific and strategic intervention, scientific infrastructure is bound to a fate that drains scientific funding and binds every bit of the practice of science to a few massive information conglomerates.

> Research on publisher propaganda networks and capture initiatives

> Field work on researching wiki communities and finding the alt-tech world that is building a better future outside the eyes of the corporate internet.

> Listening to the social scientists about the intrinsic intersection between social and technological systems.

> The “experiment” so to speak is whether it is possible to organize a group of scientists to develop anti-profitable technology and overcome the incentive systems that drive us into serfdom. I will be publicly releasing this and trying a new kind of public contribution

## 4 Resources

Since not everything is freely available or a DOI-indexed PDF...

- [SaundersWehr-JASA2019](#) - repository with code, data, and retypset version of “Mice Can Learn Phonetic Categories”
- [Autopilot Repository](#)
- [Autopilot Docs](#)
- [Autopilot Wiki](#)
- [Autopilot CI](#)
- [Autopilot test coverage](#)

## 5 References

### References

- [1] Jonny L. Saunders and Michael Wehr. “Mice Can Learn Phonetic Categories”. In: *The Journal of the Acoustical Society of America* 145.3 (), pp. 1168–1177. ISSN: 0001-4966. DOI: [10.1121/1.5091776](https://doi.org/10.1121/1.5091776). URL: <https://asa.scitation.org/doi/abs/10.1121/1.5091776> (cit. on p. 1).
- [2] Iris van Rooij and Giosuè Baggio. “Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science”. In: *Perspectives on Psychological Science* (), p. 1745691620970604. ISSN: 1745-6916. DOI: [10.1177/1745691620970604](https://doi.org/10.1177/1745691620970604). URL: <https://doi.org/10.1177/1745691620970604> (cit. on p. 2).
- [3] Jessamyn Schertz and Emily J. Clare. “Phonetic Cue Weighting in Perception and Production”. In: *WIREs Cognitive Science* 11.2 (), e1521. ISSN: 1939-5086. DOI: [10.1002/wcs.1521](https://onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1521). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1521> (cit. on p. 2).
- [4] John J Ohala et al., eds. *A Guide to the History of the Phonetic Sciences in the United States*. URL: [https://escholarship.org/uc/item/6mr8317x#article\\_main](https://escholarship.org/uc/item/6mr8317x#article_main) (cit. on p. 2).
- [5] Alvin M. Liberman and Ignatius G. Mattingly. “The Motor Theory of Speech Perception Revised”. In: *Cognition* 21.1 (), pp. 1–36. ISSN: 00100277. DOI: [10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6) (cit. on p. 2).
- [6] Patricia K. Kuhl. “Early Language Acquisition: Cracking the Speech Code”. In: *Nature Reviews Neuroscience* 5.11 (11), pp. 831–843. ISSN: 1471-0048. DOI: [10.1038/nrn1533](https://www.nature.com/articles/nrn1533). URL: <https://www.nature.com/articles/nrn1533> (cit. on p. 2).
- [7] Eleanor Rosch. “Wittgenstein and Categorization Research in Cognitive Psychology”. In: *Meaning and the Growth of Understanding: Wittgenstein’s Significance for Developmental Psychology*. Ed. by Michael Chapman and Roger A. Dixon. Berlin, Heidelberg: Springer, pp. 151–166. ISBN: 978-3-642-83023-5. DOI: [10.1007/978-3-642-83023-5\\_9](https://doi.org/10.1007/978-3-642-83023-5_9). URL: [https://doi.org/10.1007/978-3-642-83023-5\\_9](https://doi.org/10.1007/978-3-642-83023-5_9) (cit. on p. 3).
- [8] Eleanor Rosch and Carolyn B Mervis. “Family Resemblances: Studies in the Internal Structure of Categories”. In: *Cognitive Psychology* 7.4 (), pp. 573–605. ISSN: 0010-0285. DOI: [10.1016/0010-0285\(75\)90024-9](https://www.sciencedirect.com/science/article/pii/0010028575900249). URL: <http://www.sciencedirect.com/science/article/pii/0010028575900249> (cit. on p. 3).
- [9] B Elan Dresher. “The Contrastive Hierarchy in Phonology”. In: *Contrast in phonology: theory, perception, acquisition* 13 (), p. 11. ISSN: 1718-3510. DOI: [10.1017/CBO9780511642005](https://doi.org/10.1017/CBO9780511642005) (cit. on p. 3).
- [10] Keith R. Kluender, Christian E. Stilp, and Fernando Llanos Lucas. “Long-Standing Problems in Speech Perception Dissolve within an Information-Theoretic Perspective”. In: *Attention, Perception, & Psychophysics* 81.4 (), pp. 861–883. ISSN: 1943-393X. DOI: [10.3758/s13414-019-01702-x](https://doi.org/10.3758/s13414-019-01702-x). URL: <https://doi.org/10.3758/s13414-019-01702-x> (cit. on p. 3).
- [11] C Angeloni and MN Geffen. “Contextual Modulation of Sound Processing in the Auditory Cortex”. In: *Current Opinion in Neurobiology*. Neurobiology of Behavior 49 (), pp. 8–15. ISSN: 0959-4388. DOI: [10.1016/j.conb.2017.10.012](https://www.sciencedirect.com/science/article/pii/S0959438817302325). URL: <https://www.sciencedirect.com/science/article/pii/S0959438817302325> (cit. on p. 3).
- [12] Lori L. Holt. “Temporally Nonadjacent Nonlinguistic Sounds Affect Speech Categorization”. In: *Psychological Science* 16.4 (), pp. 305–312. ISSN: 0956-7976. DOI: [10.1111/j.0956-7976.2005.01532.x](https://doi.org/10.1111/j.0956-7976.2005.01532.x). pmid: 15828978 (cit. on p. 3).
- [13] Jonny L. Saunders and Michael Wehr. “Autopilot: Automating Behavioral Experiments with Lots of Raspberry Pis”. In: *bioRxiv* (), p. 807693. DOI: [10.1101/807693](https://www.biorxiv.org/content/10.1101/807693v1). URL: <https://www.biorxiv.org/content/10.1101/807693v1> (cit. on p. 3).

- [14] Alexander Mathis et al. “DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning”. In: *Nature Neuroscience* 21.9 (9), pp. 1281–1289. ISSN: 1546-1726. DOI: [10 . 1038 / s41593 - 018 - 0209 - y](https://doi.org/10.1038/s41593-018-0209-y). URL: <https://www.nature.com/articles/s41593-018-0209-y> (cit. on p. 4).
- [15] Gary A Kane et al. “Real-Time, Low-Latency Closed-Loop Feedback Using Markerless Posture Tracking”. In: *eLife* 9 (). Ed. by Gordon J Berman et al., e61909. ISSN: 2050-084X. DOI: [10 . 7554 / eLife . 61909](https://doi.org/10.7554/eLife.61909). URL: <https://doi.org/10.7554/eLife.61909> (cit. on p. 4).
- [16] Julianne LaChance et al. “PVP<sub>1</sub>—The People’s Ventilator Project: A Fully Open, Low-Cost, Pressure-Controlled Ventilator”. In: *medRxiv* (), p. 2020.10.02.20206037. DOI: [10 . 1101 / 2020 . 10 . 02 . 20206037v1](https://doi.org/10.1101/2020.10.02.20206037v1). URL: <https://www.medrxiv.org/content/10.1101/2020.10.02.20206037v1> (cit. on p. 5).