key:

- ⚠️ start from scratch
- 🧑‍💻 minor code update
- 🤖 major technical work
- 🖊️ detailed illustration
- 🗺️ minor layout changes
- ✂️ potential merge / cut, or not sure whether it will be in scope

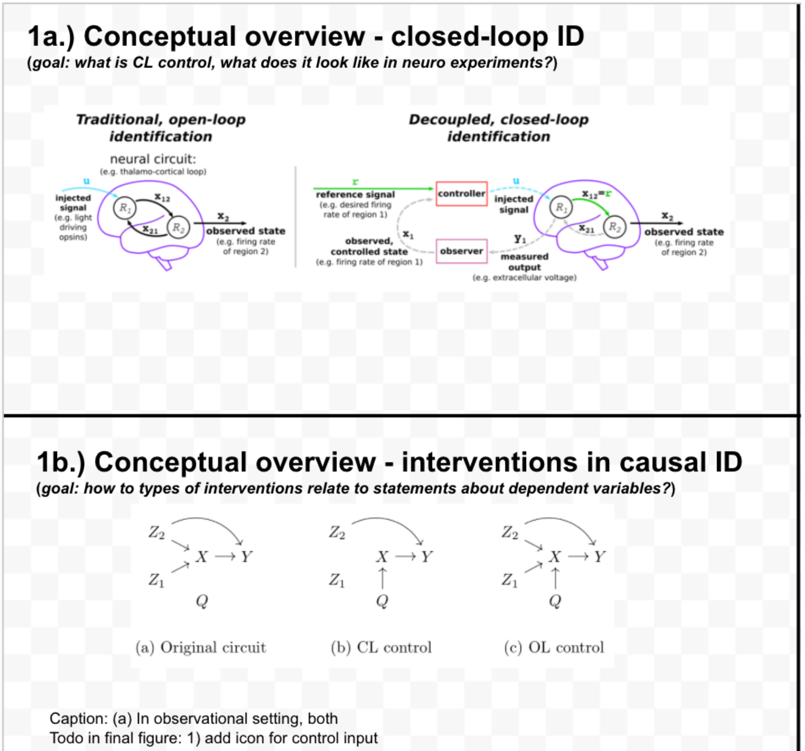
Most important, least complete figures

- [ ] Empirical/simulation methods overview
    - 🗺️ biggest challenge is complexity of layout, deciding what to keep
        - outputs
        - intervention as timeseries
        - correlation matrices or xcorr
        - circuit estimates
- [ ] hypothesis entropy figure
    - 🤖 requires some coding to finish
        - need some infrastructure for computing entropy across many conditions
        - needs some summary statistics across a well-defined hypothesis set
    - 🗺️ will likely require layout iterations
- [ ] improved data efficiency & bias
    - 🤖 🤖 requires major technical work - namely committing to how to turn correlations into circuit inference
    - ✂️ as such, may end up having to cut this
    - ... but if we end up connecting back to the more detailed empirical sims, this will be necessary

bonus figures

- [ ] Theory overview
    - predicting correlation structure
    - predicting intervention's impact on correlation
- [ ] flowchart for steps of intervention experiment

Intro / background

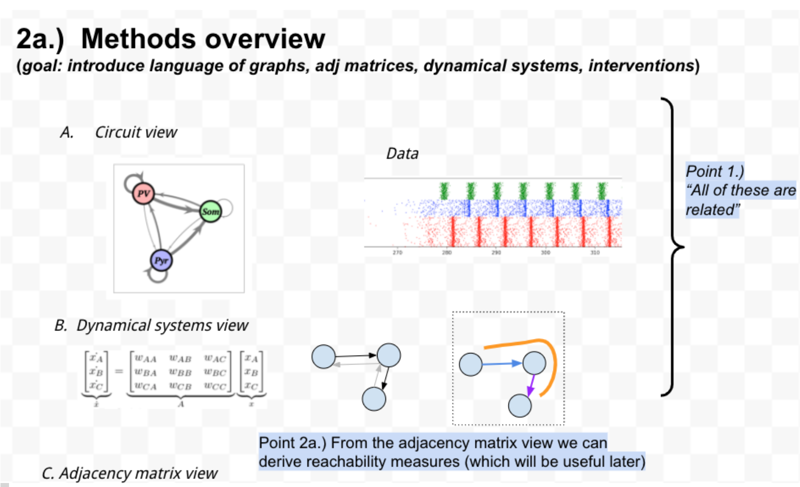


🗺️ needs layout / illustration cleaning up

🖊️ top panel, needs a lot of excess cutting out

✂️ bottom panel, need to decide whether we want to use this language to talk about the effect of intervention

✂️ overall, is this figure redundant with 2 circuit walktrhough?



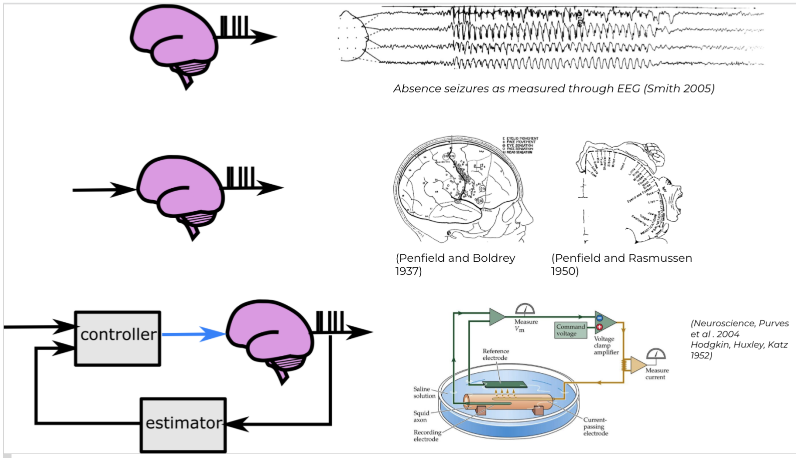
🧑‍💻 needs commitment to concrete time-series data

🗺️ needs layout / illustration cleaning up

but overall, conceptually, close enough to write a caption for

> ✂️ merge with conceptual overview figure

Interventions in Neuro



> *(close to final, but could be significantly cut down / merged with other figure)*
>
> **Figure:** Examples of the role of interventions in discoveries in neuroscience (A) Identifying when a patient is having a seizure, from passive recordings alone (B) through systematic open-loop stimulation experiments, Penfield was able to uncover the spatial organization of how senses and movement are mapped in the cortex [2] (C) Feedback control allows us to specify activity in the brain in terms of outputs. Allows us to reject disturbances, respond to changes
>
> ✂️ merge into causal diagram? - might be cut if the paper ends up being constrained on figures

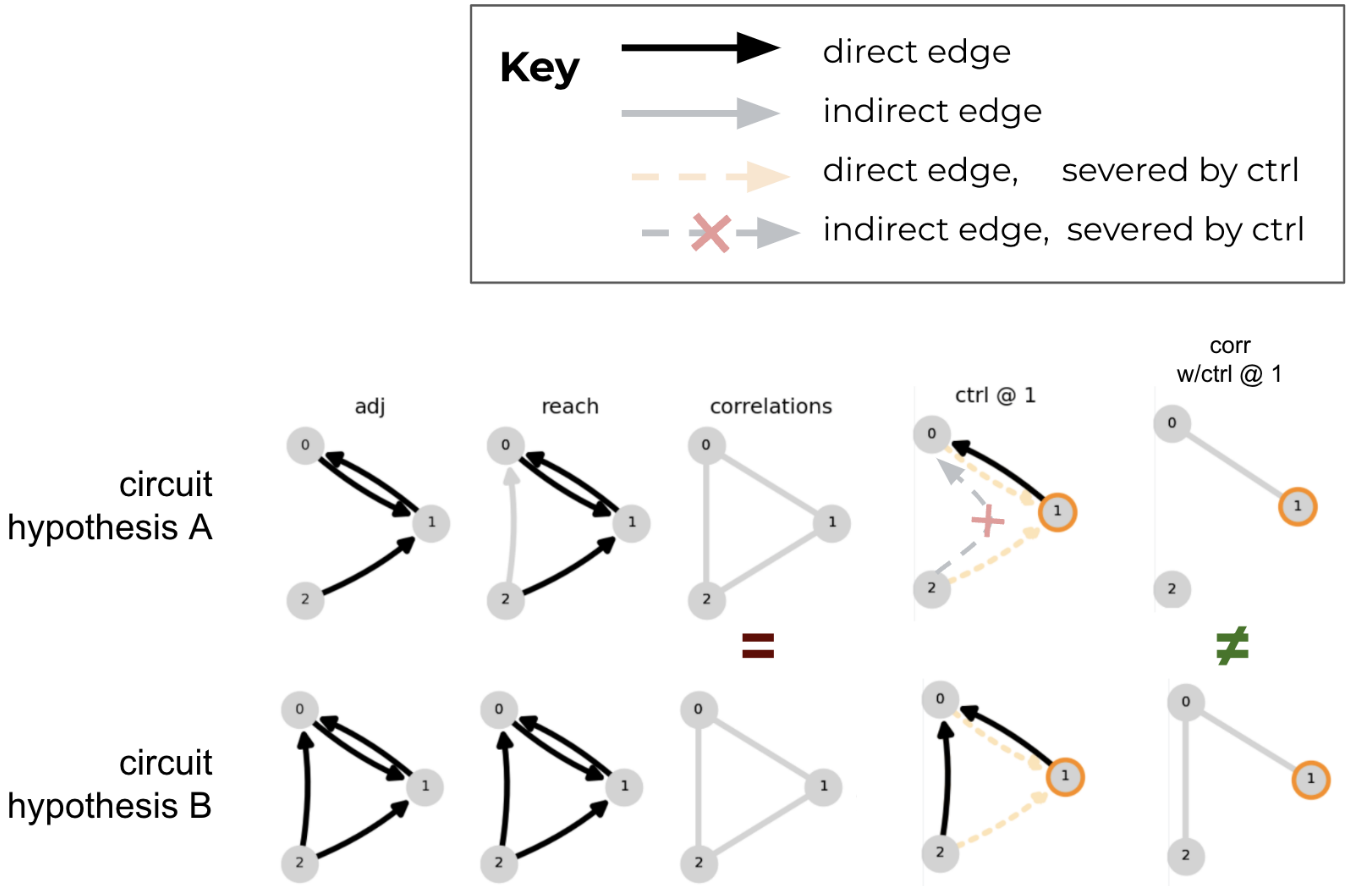


Figure DEMO: Applying CLINC to distinguish a pair of circuits

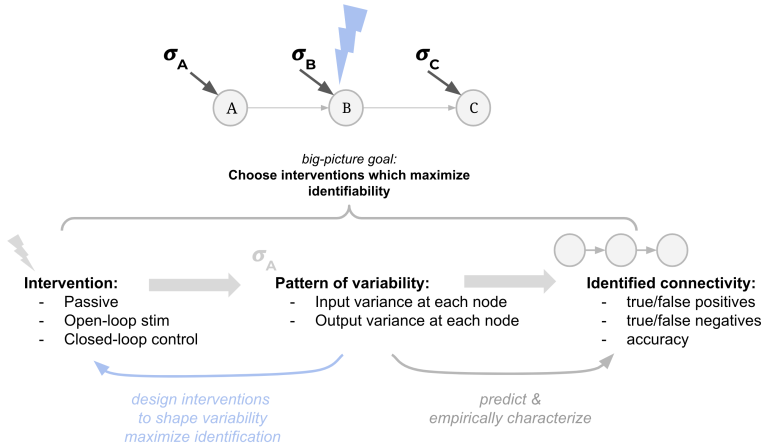
🎉 *(img close to final draft)*

good enough sketch to write a caption for

▸ ↪2,3 circuit versions, straight from code

Theory

⚠️ need theory overview!



> ⚠️ very loose sketch, need to decide what we want here

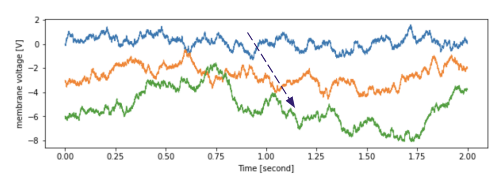
Simulation methods

Network simulations

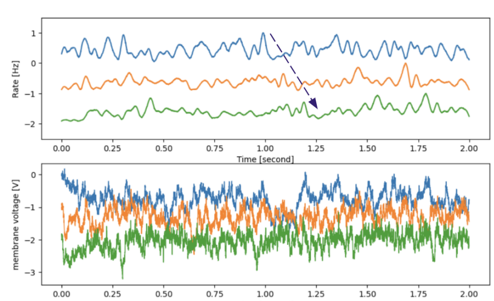
Implementing interventions

Estimating circuits from data

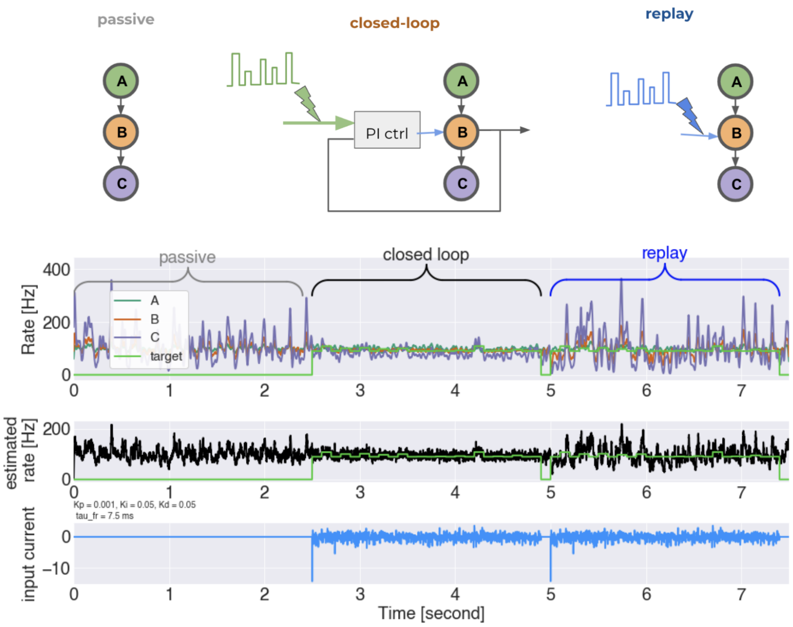
Gaussian network
(delayed interactions)

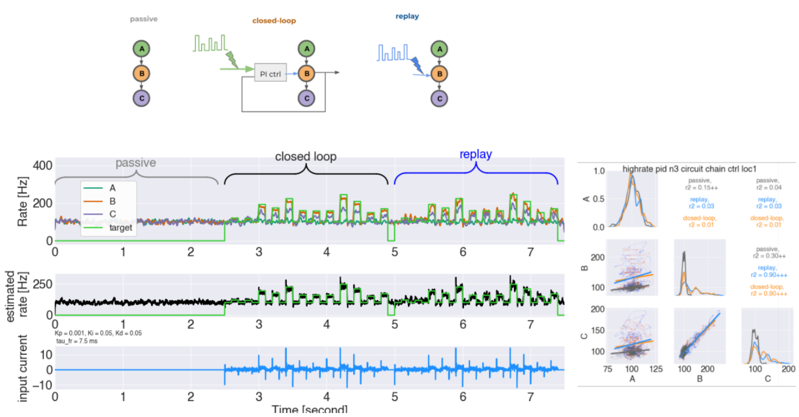


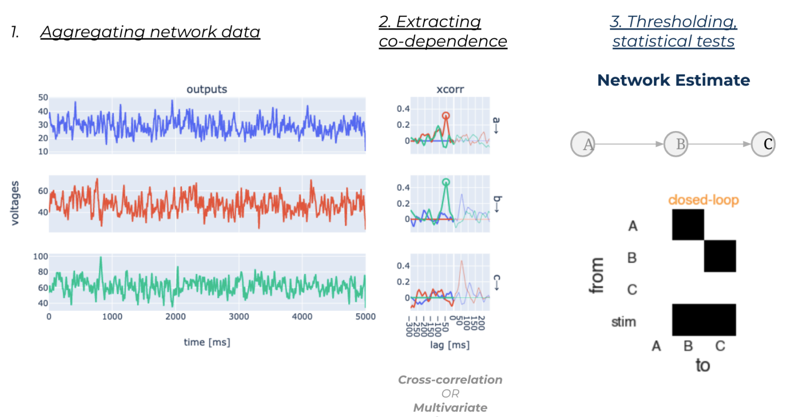
LIF spiking network
(delayed interactions)



small_circuit_scripts/circuit_functions/delayed_gaussian_network.py







> see also google slides ...
>
> 🖌 something that shows neurons in networks on left side
>
> ✂️ not sure whether extracting co-dependence and statistical tests will be in scope for the paper

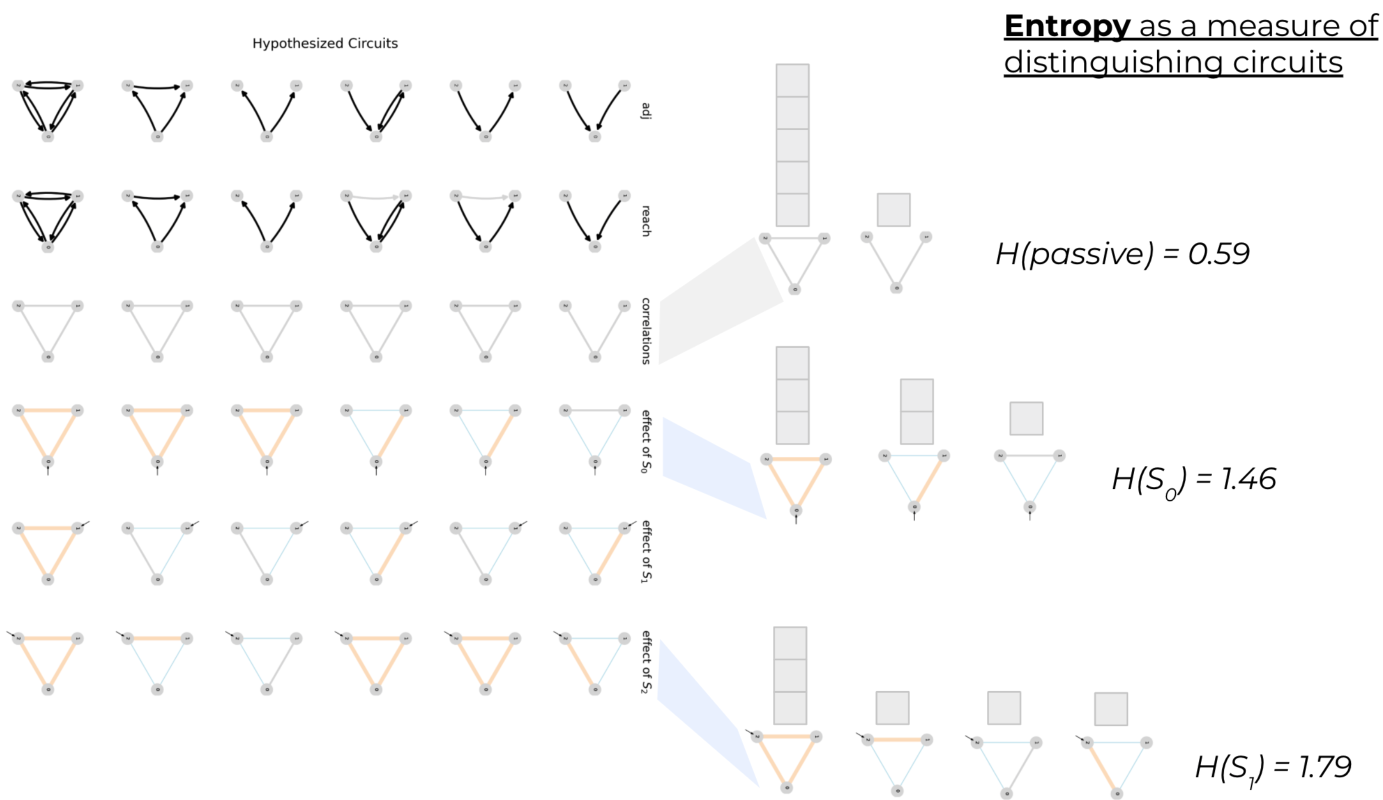
▶ ↪see also, xcorr, predicting correlation

⚠️🚧 **figure request:** flowchart for steps of intervention experiment 🚧⚠️
see section_content/_steps_of_inference.md

Results

Impact of intervention

Intervening provides (categorical) improvements in inference power beyond passive observation



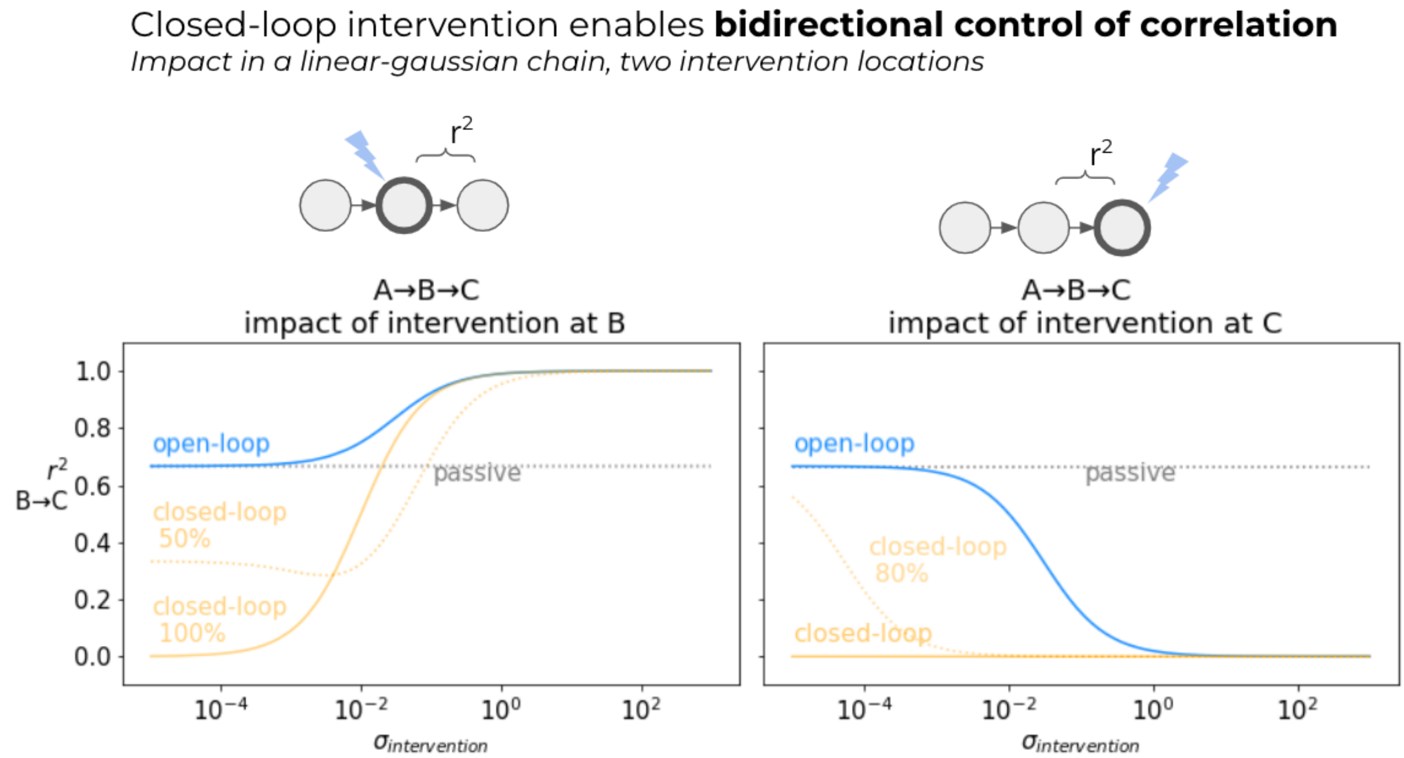
> **Figure DISAMBIG: Interventions narrow the set of hypotheses consistent with observed correlations**
> **(A)** Directed adjacency matrices represent the true and hypothesized causal circuit structure
> **(B)** Directed reachability matrices represent the direct *(black)* and indirect *(grey)* influences in a network. Notably, different adjacency matrices can have equivalent reachability matrices making distinguishing between similar causal structures difficult, even with open-loop control.
> **(C)** Correlations between pairs of nodes. Under passive observation, the direction of influence is

> difficult to ascertain. In densely connected networks, many distinct ground-truth causal structures result in similar "all correlated with all" patterns providing little information about the true structure.
> **(D-F)** The impact of open-loop intervention at each of the nodes in the network is illustrated by modifications to the passive correlation pattern. Thick orange[^edge_color] edges denote correlations which increase above their baseline value with high variance open-loop input. Thin blue[^edge_color] edges denote correlations which decrease, often as a result of increased connection-independent "noise" variance in one of the participating nodes. Grey edges are unaffected by intervention at that location.
> A given hypotheses set (A) will result in an "intervention-specific fingerprint", that is a distribution of frequencies for observing patterns of modified correlations *(across a single row within D-F)*. If this fingerprint contains many examples of the same pattern of correlation (such as **B**), many hypotheses correspond to the same observation, and that experiment contributes low information to distinguish between structures. A maximally informative intervention would produce a unique pattern of correlation for each member of the hypothesis set.
> 🚧 `caption too long`

Stronger intervention shapes correlation, resulting in more data-efficient inference with less bias

```
Explain why closed-loop helps - bidirectional variance control
```

Impact of intervention location and variance on pariwise correlations



🚧 (Final figure will be a mix of these two panels, caption will need updating) **Figure VAR: Location, variance, and type of intervention shape pairwise correlations**

(CENTER) A two-node linear gaussian network is simulated with a connection from A→B. Open-loop interventions *(blue)* consist of independent gaussian inputs with a range of variances σ_S^2. Closed-loop interventions *(orange)* consist of feedback control with an independent gaussian target with a range of variances. *Incomplete closed-loop interventions result in node outputs which are a mix of the control target and network-driven activity*. Connections from sources to nodes are colored by their impact on correlations between A and B; green denotes $dR/dS > 0$, red denotes $dR/dS < 0$.

(lower left) Intervention "upstream" of the connection A→B increases the correlation $r^2(A, B)$.

(lower right) Intervention at the terminal of the connection A→B decreases the correlation $r^2(A, B)$ by adding connection-independent noise.

(upper left) Intervention with shared inputs to both nodes generally increases $r^2(A, B)$, *(even without A→B, see supplement)*.

(upper right) The impact of shared interventions depends on relative weighted reachability $\mathrm{Reach}(S_k \to A)/\mathrm{Reach}(S_k \to B)$, with highest correlations when these terms are matched (see *)

Closed-loop interventions *(orange)* generally result in larger changes in correlation across σ_S^2 than the equivalent open-loop intervention. Closed-loop control at B effectively lesions the connection A→B, resulting in near-zero correlation.

> [1]
>
> but may need to layout panels by parts of the circuit like old sketch:

▶ ↪old sketch

data efficiency & bias ⚠️

⚠️

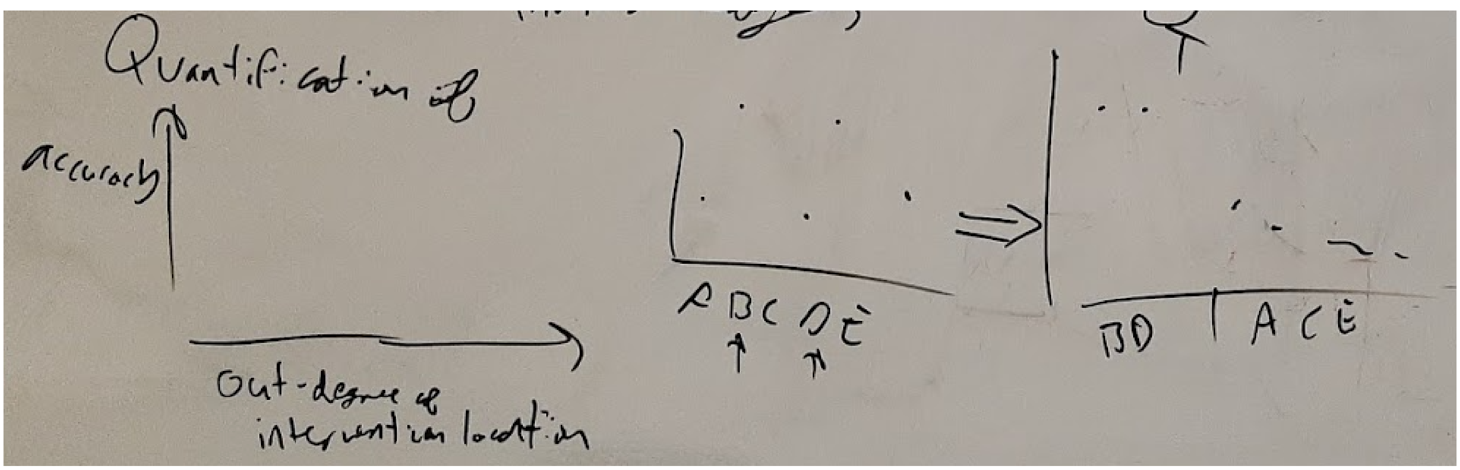
> **Figure DATA: Analysis of simulated circuits suggest stronger intervention facilitates identification with less data** [2]

⚠️

> **Figure DATA: Analysis of simulated circuits suggest stronger intervention facilitates identification with less data** [2:1]
> Explain why closed-loop helps - less bias
>
> - higher infinite-data accuracy (i.e. less bias)
>   - lower bias likely comes from the categorical advantages above
> - breakdown false positives, false negatives

▶ figure sketches

Impact of circuit properties ⚠️

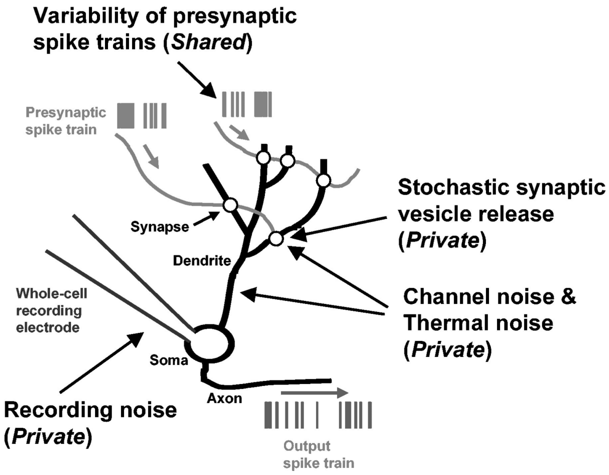


- 🤖 ⚠️ analysis hasn't been written yet! maybe have to be cut

Supplement

shared v.s. private variance

some variant on:

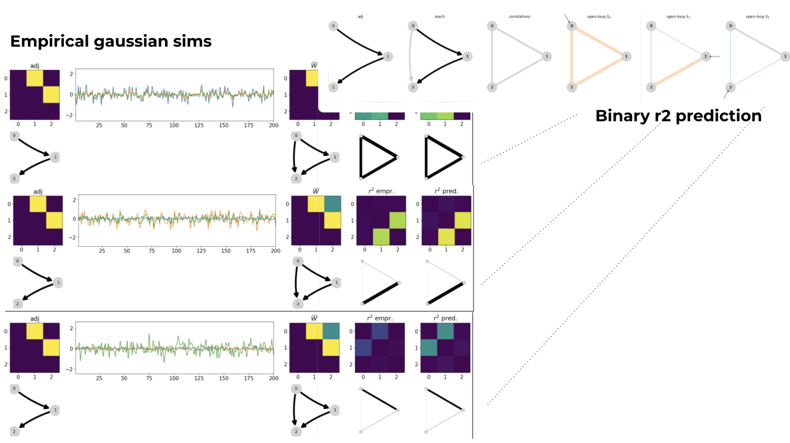


▶ ↪ graph for shared v.s. private sources

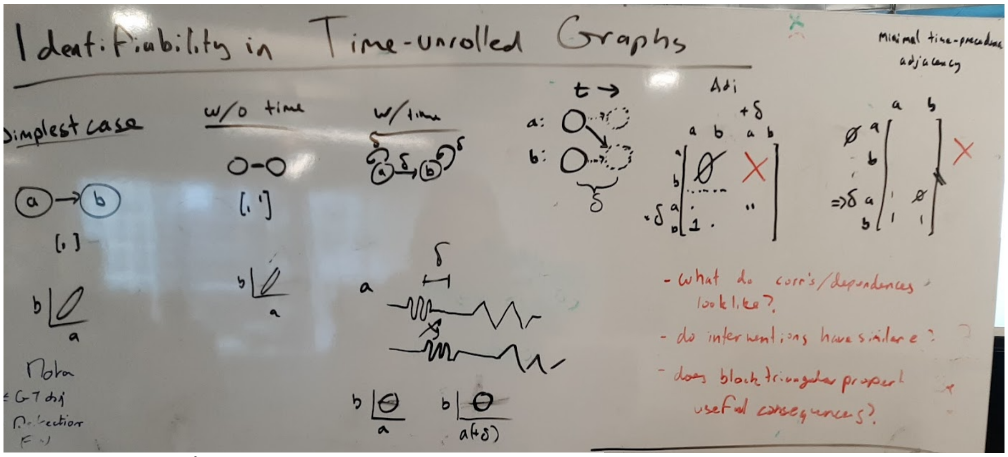
- [ ] **Figure:** illustrate reachability *(skip for now?)* - explain how reachability is derived from adjacency

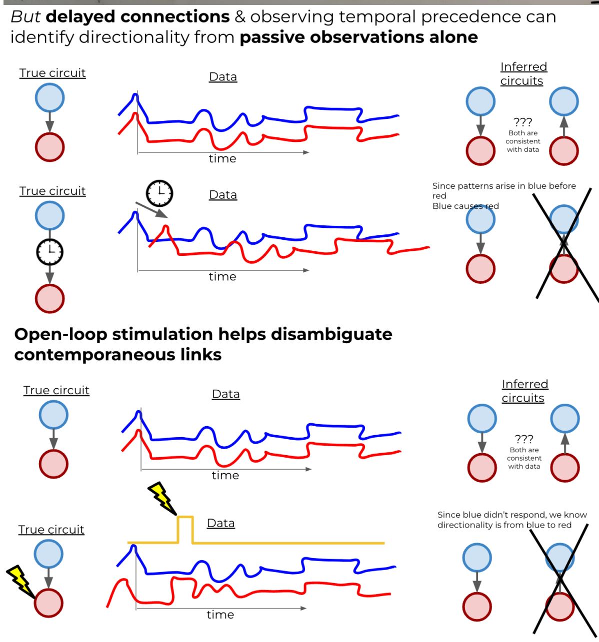
Prediction v.s. emprirical correlation

- predicting correlation across network parameters



Contemporaneous v.s. time-resolvable network effects





1. compare especially to "Transfer Entropy as a Measure of Brain Connectivity", "How Connectivity, Background Activity, and Synaptic Properties Shape the Cross-Correlation between Spike Trains" Figure 3. ↩
2. examples of data efficiency, accuracy figures can be found in: "Extending Transfer Entropy Improves Identification of Effective Connectivity in a Spiking Cortical Network Model", "Evaluation of the Performance of Information Theory- Based Methods and Cross-Correlation to Estimate the Functional Connectivity in Cortical Networks" ↩ ↩