UNIVERSITÉ DU LUXEMBOURG

Towards a Computational Model of General Cognitive Control Using Artificial Intelligence, Experimental Psychology and Cognitive Neuroscience

PhD Defense
Morteza Ansarinia
January 30, 2023

Thank you!

I’m very much interested in a higher-order human capacity that enables goal-directed behavior.

This goal-directed behavior is unmatched by any other species and still unmatched by AI and robots.

Towards a Computational Model of General **Cognitive Control** Using Artificial Intelligence, Experimental Psychology and Cognitive Neuroscience

Processes that generate and monitor plans,
in pursuit of evolving goals,
often in noisy environments.

Badre (2020), Cohen (2017)

The scientific concepts that are closest to this capacity are cognitive control, executive function, self-regulation, and attentional control, to name just a few.

There are differences between these terms, but I will ignore that for now and stick to the term cognitive control.

<CLICK>

The term cognitive control as described by Badre and Cohen is an umbrella term that represents “processes that generate and monitor **plans**, in pursuit of evolving **goals**, often in **noisy** environments.”

Cognitive control happens, and it happens all the time.

It is present in almost everything we do. **Consider for example cooking a pizza…**

Cooking a pizza involves **reaching a goal**, which is… well… the pizza**.**

<CLICK>

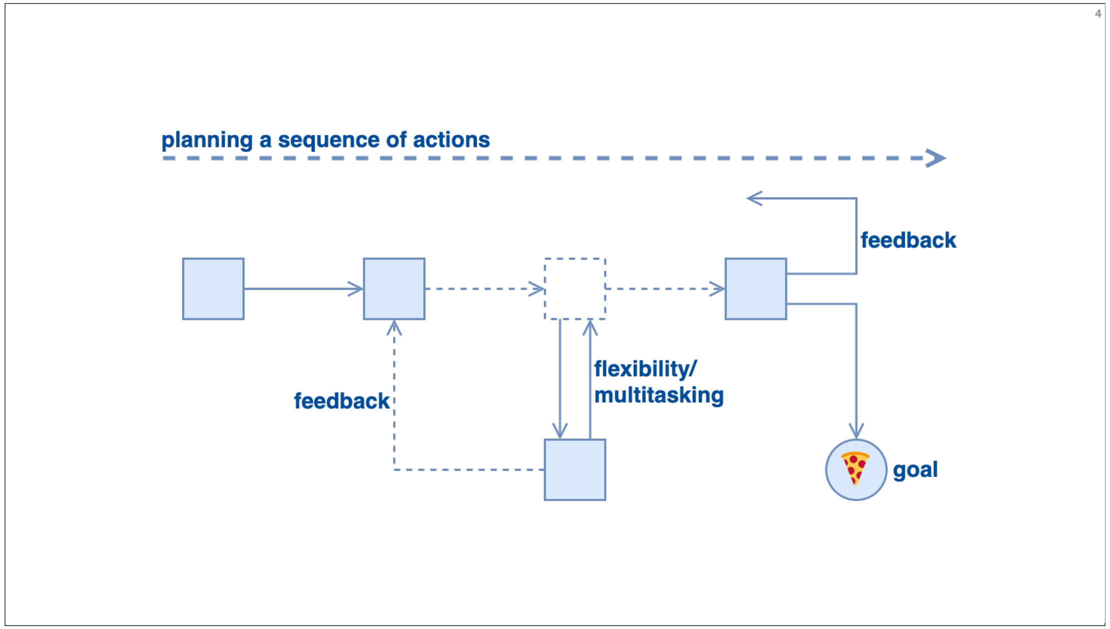
It **requires** planning, multitasking, handling feedbacks, and ignoring noises…

<CLICK>

Properly cooking a Pizza may also benefit from **transfer effect,**

that is if we learn to cook pizza in the oven, it may improve our ability to cook stovetop pizza,

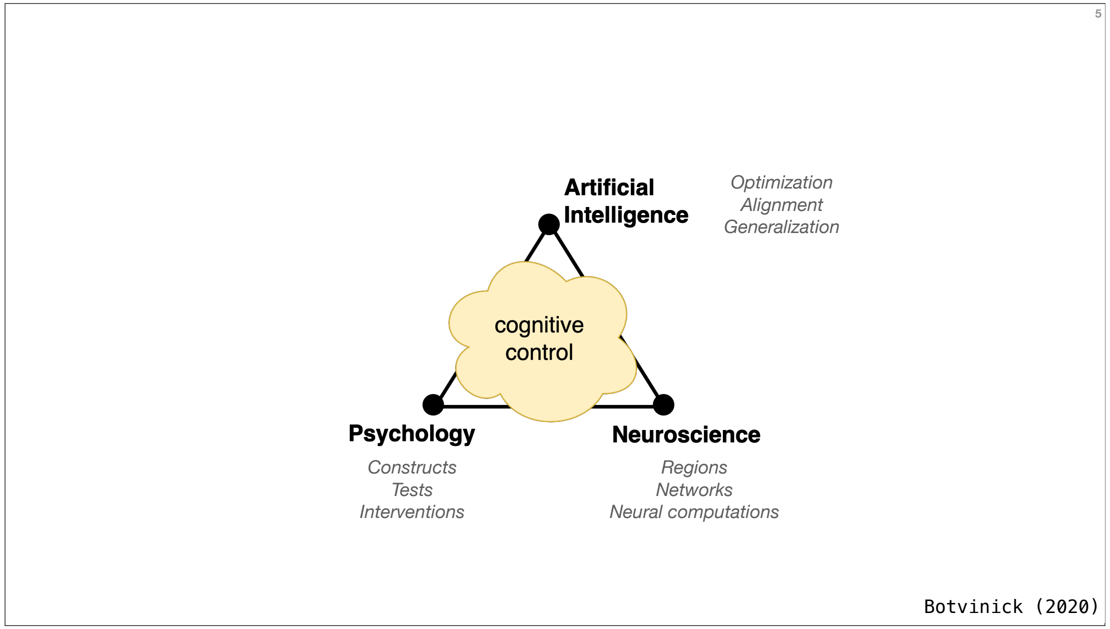
More formally, cooking pizza can be shown like this…



where **boxes** are tasks and the **goal** is the pizza,

It can be seen from here how present this cognitive control is in our life, and how sometimes complicated it can be,

What makes cognitive control even more important… and interesting… is that it is the concern of multiple disciplines.



<CLICK>

In **psychology**, for example, psychologists have been developing many theories and tasks to define and quantify it,

And for its critical role in daily functioning, long-term achievement, and psychological health, the possibility to enhance it is interesting,

So psychologists developed many cognitive training interventions to enhance it,

<CLICK>

In **neuroscience**, we want to know how the brain enables such ability that broadly generalize,

Neuroscientists for example identified many brain functions relevant to cognitive control including regions, networks, and neural computations,

<CLICK>

And in **AI**, control can be simply thought as an optimization to reach a goal,

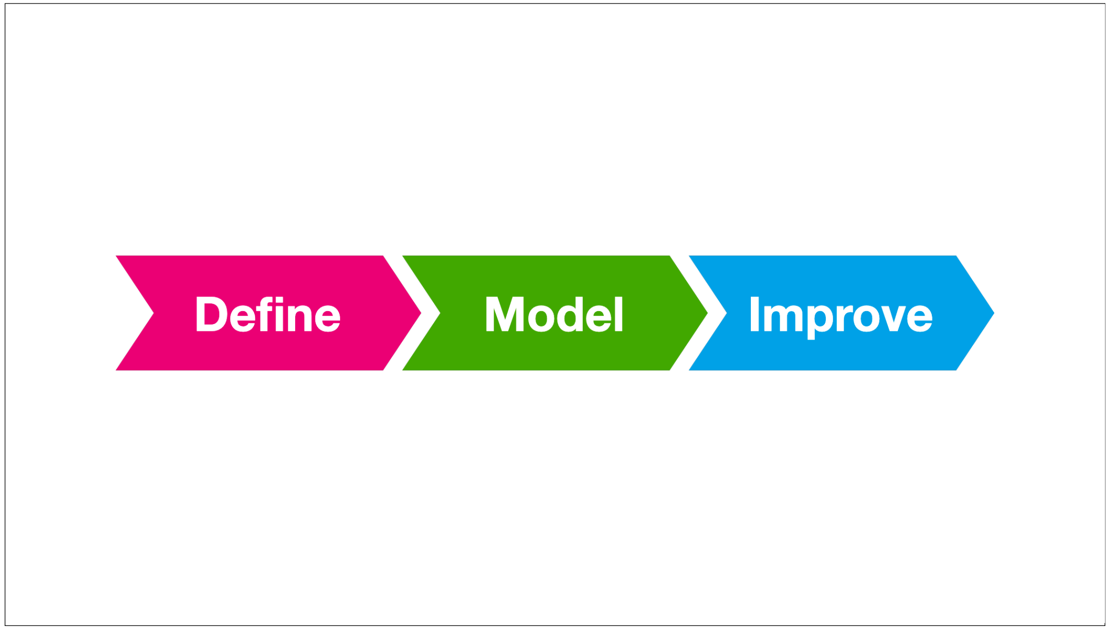
there are also two AI challenges that are relevant to cognitive control:

- *First is alignment*, where we want artificial agents to align to human capacities and intentions,
- and *generalization*, where we want AI agents to broadly transfer prior knowledge,

<CLICK>

Overall, there is practical value in bridging the gap between these disciplines,

And multidisciplinary synergy is the core of my thesis.



My thesis work about cognitive control can be organized in three categories,

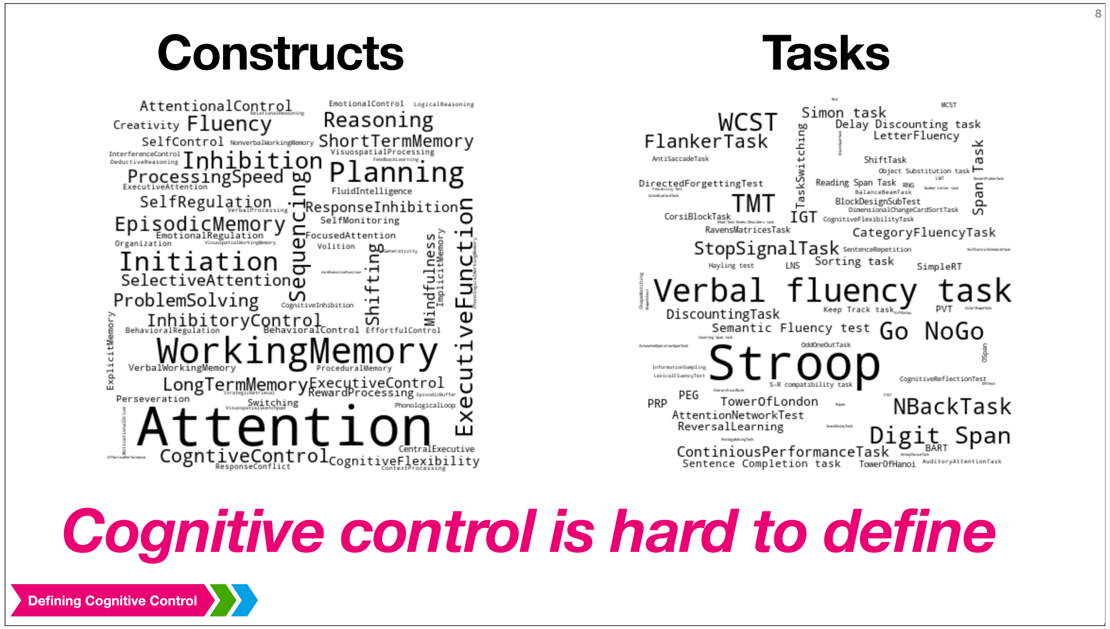
defining cognitive control, modeling it, and training or improving cognitive control.

Let's begin by ... [defining]

Defining Cognitive Control

Chapter 1

What do researchers mean by cognitive control…



To illustrate the challenge of defining,

Here are some relevant constructs about cognitive control. It involves many concepts,

Including attention, working memory, problem solving, and the rest,

<CLICK>

There are also many tasks that are used tap into and quantify different aspects of cognitive control,
Including stroop, trail-making-test, flanker test, and the rest,

the relationship between the two… constructs and tasks… is not always clear,

<CLICK>

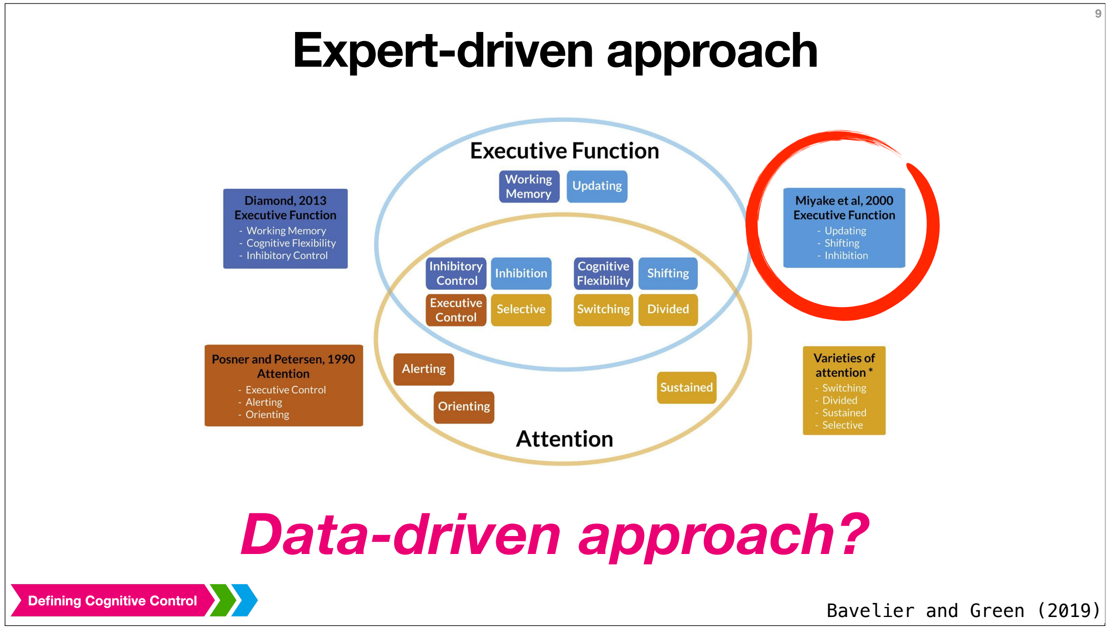
So cognitive control is in fact hard to define,

Which makes it difficult to have an objective and cohesive framework,

…

There are multiple ways to create a knowledge framework of cognitive control.

Currently, we mostly rely on expert-driven knowledge models, Like cognitive atlas or this…



Human experts manually review the literature or devise experiments, to create models like this,

These knowledge maps are useful but,

Problem is that they are subjective and biased, and unable to keep up with the growing literature about cognitive control,

<CLICK>

For example, here, Miyake and Friedman proposed three components of executive functions which was developed in 2000.

And after that it was updated in 2015,

But people still cite the original older one from 2000,

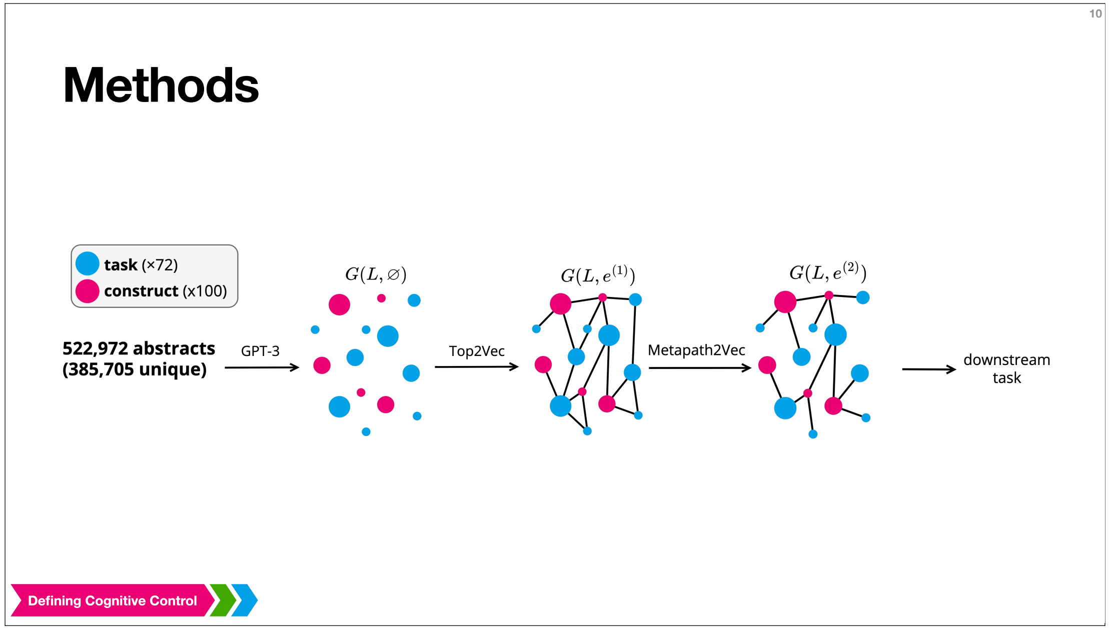
<CLICK>

A possible solution is to use an objective data-driven approach,

We recent advances in language models in AI, we can parse large body of scientific texts and create a knowledge map that relates underlying ideas in a quantitive way.

So, what I suggest is to use natural language processing from AI and create a representation of cognitive control tasks and constructs in a cohesive framework,

Here is how I've created a knowledge model of cognitive control…



I’ve created an ontology, or more precisely a lexicon of 172 task names and constructs names,

And then looked up PubMed and collected about half a million scientific abstracts,

<CLICK>

I’ve then used GPT-3 to encode those texts into numerical vectors. So we have half a million vectors that each represent a paper,

<CLICK>

I’ve then used a Topic modeling to extract underlying ideas from those documents,
So after this step all the tasks and constructs, will be represented in a shared space of topics,
And we can use measures of similarity to relate them.

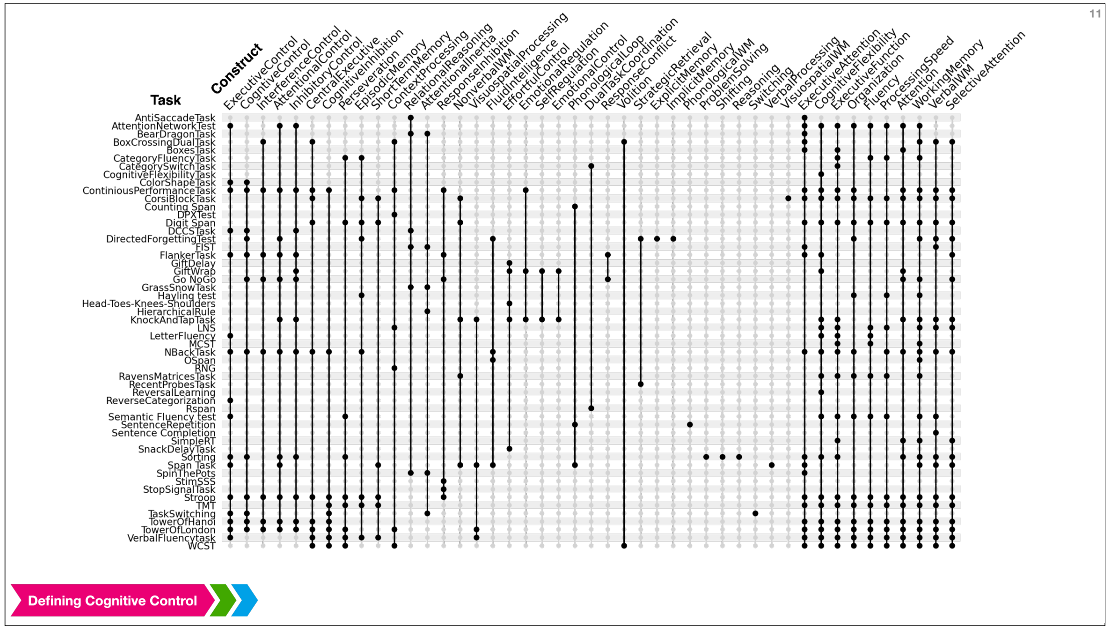
<CLICK>

Finally, I’ve created a graph that relates tasks and constructs based on their similarity, The graph reground tasks on constructs and the other way and provide a semantic representation for their relationships,

<CLICK>

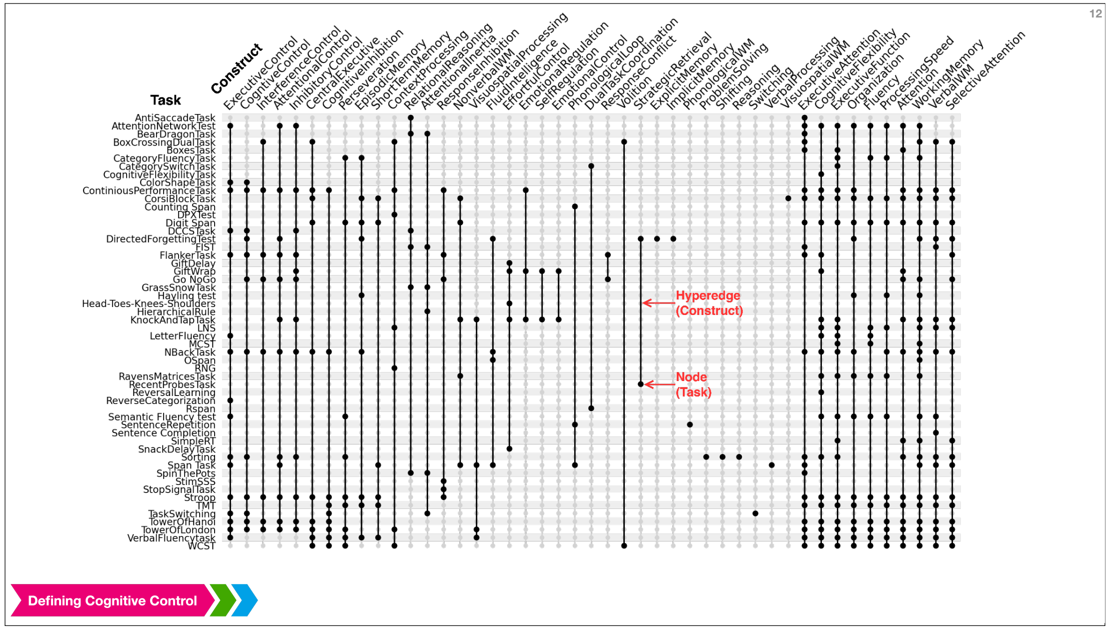
The graph was then used for downstream tasks like clustering or measuring distance between tasks and constructs…

Here is the graph…



On the top, you see the constructs,

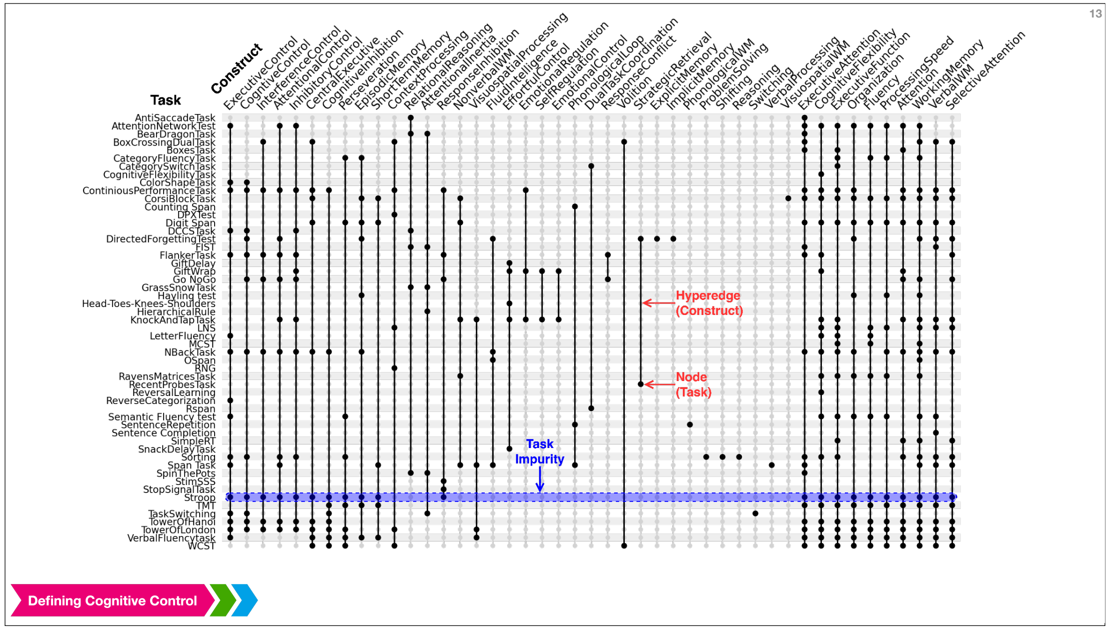
And on the side the tasks…



Vertical lines are edges, or hyper edges, of the graph, that is constructs.

And they connects relevant tasks, the black dots.

The structure of the graph is important, because it shows **two impediments in defining cognitive control.**

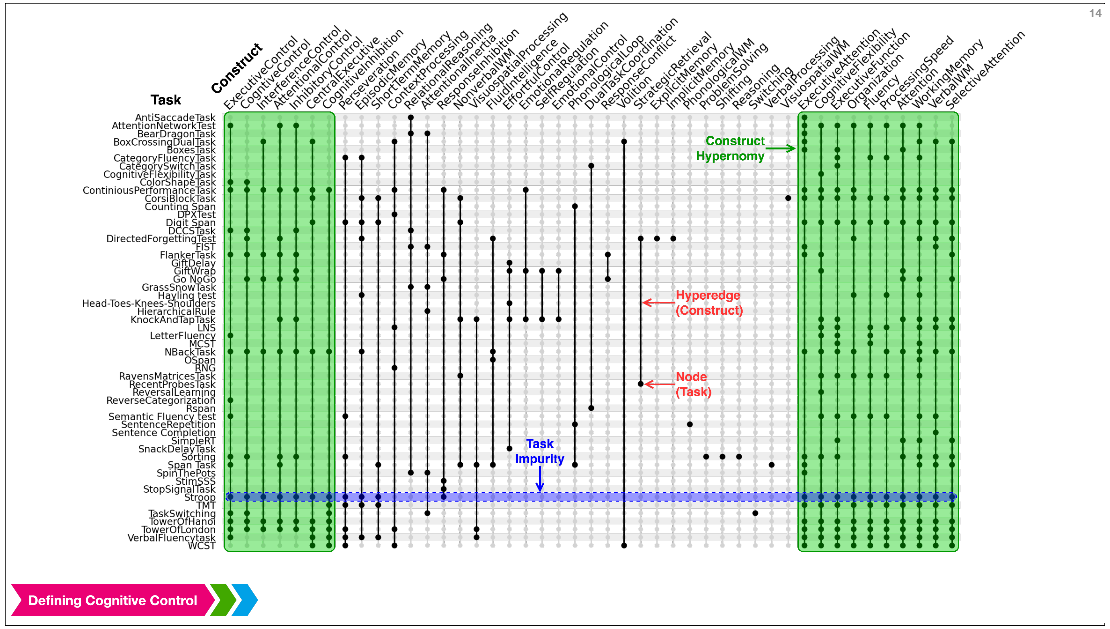


The first one, is task immunity: That is one task taps into multiple constructs,

For example Stroop task here, loads into Attention, Inhibition, Fluency, WorkingMemory and more…

Task impurity makes it difficult to isolate a single constructs using only one task…

And the second challenge is …



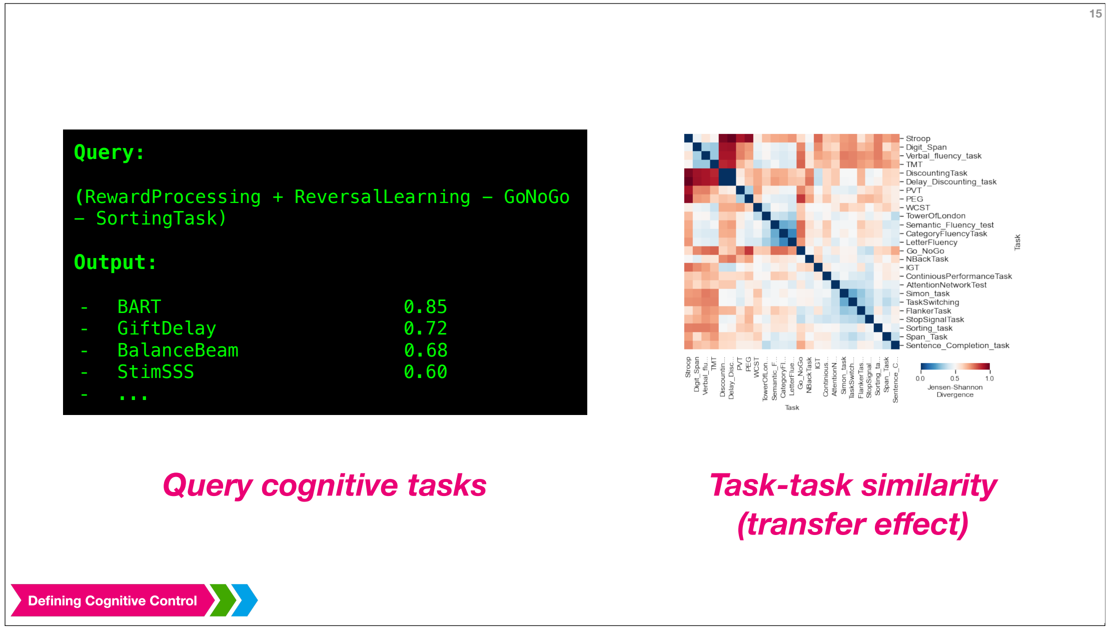
… what we call Construct hypernomy.

Construct hypernomy, shown here in green boxes, is that “different constructs are measured with the same set of cognitive tasks,”
Depending on the context and experimental design,

This makes it difficult to see for example what is the difference between Executive Control and Cognitive or Control and Central Executive

This data-driven quantitative knowledge graph can enable many applications in cognitive control research.

For example…



It helps with selecting a battery of tasks for an experiment

Here, for example, we query the graph for the tasks related to RewardProcessing and ReversalLearning

But we don't want GoNoGo and sorting tasks…

The results are BART, GiftDelay and the rest.

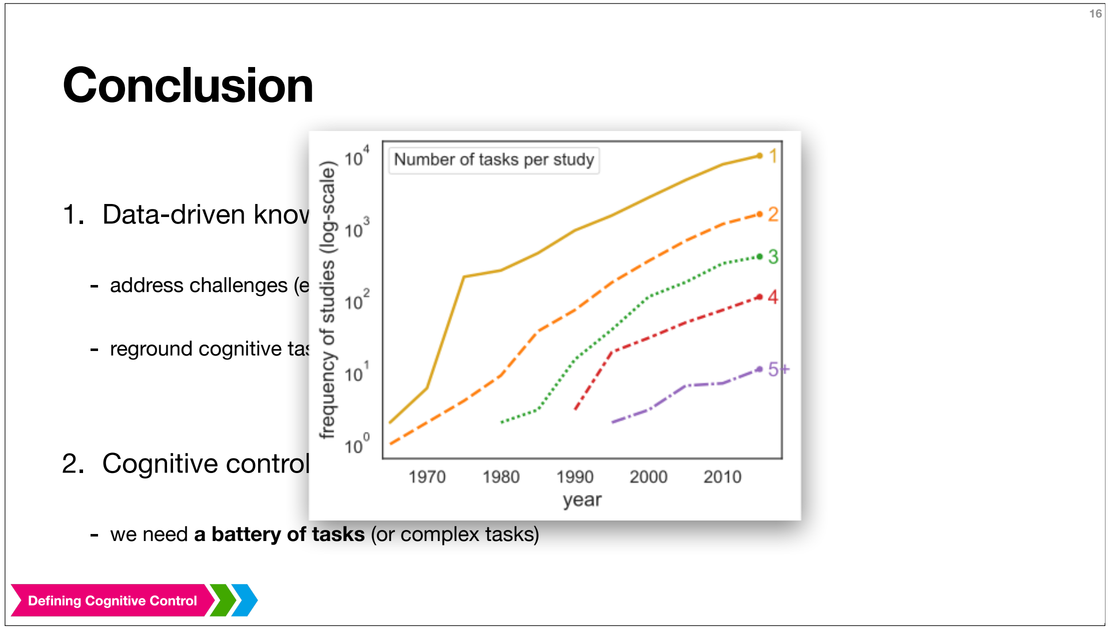
It also gives us a ranking about how relevant the results are to our query.

<CLICK>

Or, we might be interest to quantify how similar cognitive tasks are based on the scientific texts…

Which can provide a principled way to study transfer effect a priori. That is prior to conducting experiments or computational modeling

So to recap part one, …



There is a value in data-driven knowledge models such as ontologies that are automatically extracted from scientific texts,

They improve current frameworks of cognitive control and help to identify and address several challenges such as construct hypernomy and task impurity.
The model also reground tasks (or methods) on constructs (or theories), and constructs on tasks in a single cohesive framework.

The model also shows that cognitive control is contextual and involves multiple tasks and constructs,

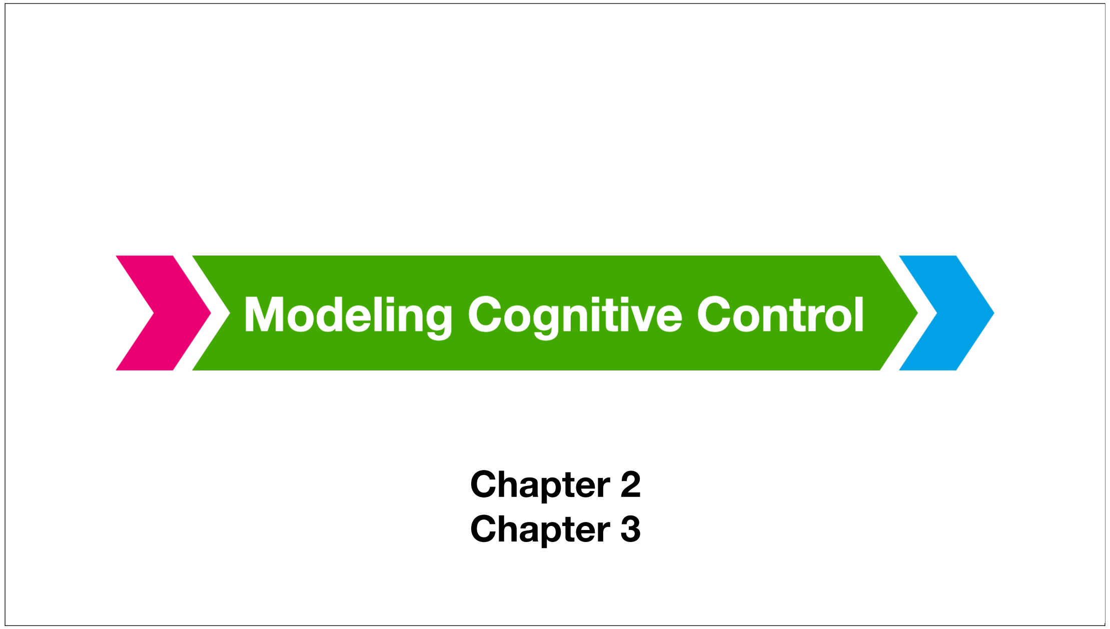
So rather than individual tasks, we need for a battery of tests (or complex tasks) to properly tap into cognitive control.

<CLICK>

This single-task design is in fact a problem in a number of studies, and most studies use only one task.

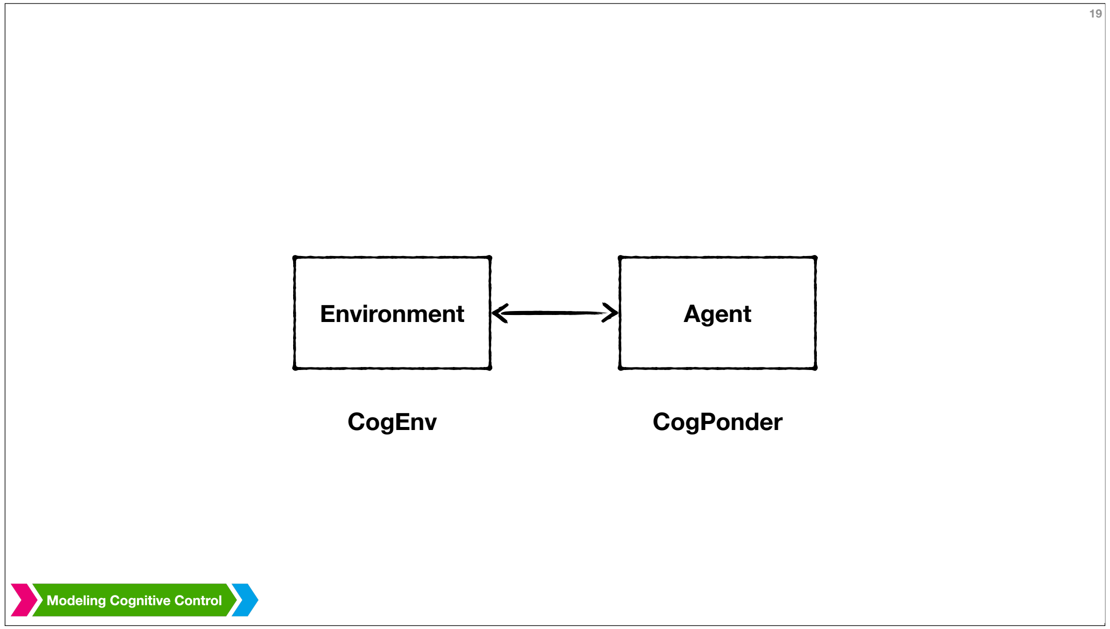
Now let's see how this inspires computational models of cognitive control…

The focus of the second part of my thesis is computation models of cognitive control…



A key step towards this is to make it easier to directly compare artificial agents and humans.

There are many cases for single task computational models but less for multiple tasks or principled models for complex tasks



To give a background…

In modern psychology, as George Miller of Harvard university said, the goal is to find systems that explains what humans do as agents in the environment around us.

The environment-agent duality is also a generic notion that we use in AI.

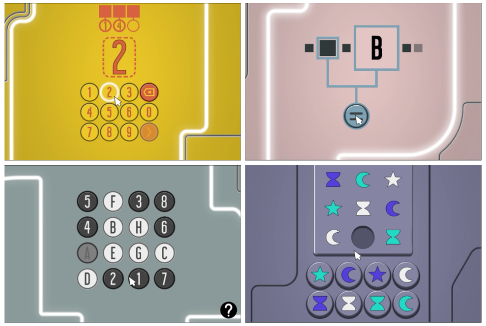
So, how can we build a model of cognitive control that is

One) constrained by human capacities, and
two) can be used in artificial agents; That is the model would interoperable and actually generates actions rather than merely explaining the patterns in the data.

We also want the model to be scalable so it can perform complex tasks or a battery of tasks…

And last but not least, we want the model to functionally separate control, and facilitate the study of cognitive control

First, let me show the environment that we created… CogEnv…



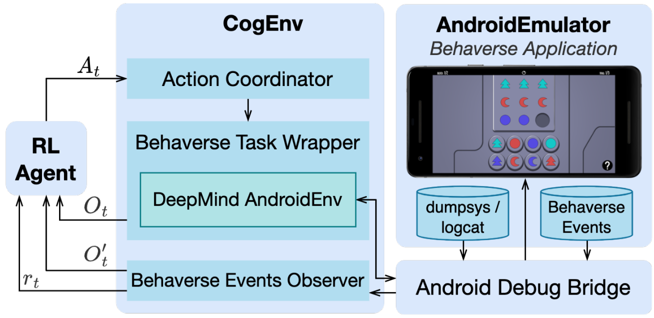
There are currently cognitive batteries and environments that can be used for humans…
like psychToolkit, or ExpFactory, or the other ones

And environments that can be used for artificial agents… like DeepMind PsychLab, and the rest,

To make it work for BOTH humans and artificial agents,

We used Behaverse, a software the provides common cognitive tests, **and…**

CogEnv



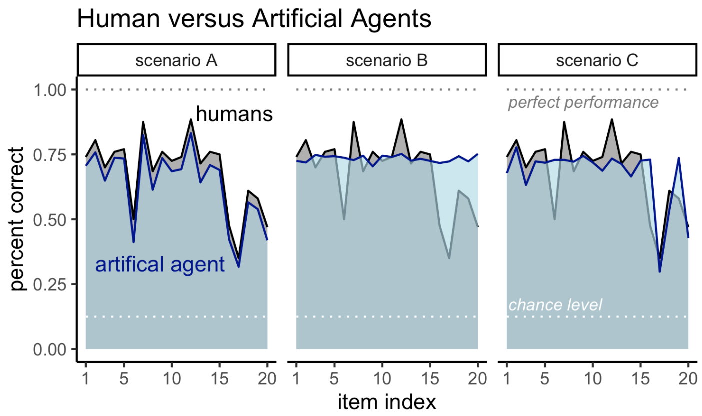
… created a virtual environment where AI agent can perform the same tasks as in in human studies of cognitive control.

This CogEnv wraps Behaverse and allows to directly compare cognitive performance in humans and AI agents.

From technical point of view, the data will be in the same format for all the cognitive tests and agent types.

Why is it useful? Well, we can expose humans and AI agents to the same set of tasks and use the same analysis pipeline to study them.

For example…



… it allows to do diagnostic studies.

We can develop scenarios to test hypothesis about underlying computations in human.

Here, three scenarios are expected about the results we may collect during an experiment.

In scenario A, on the left, agents and humans perform similarly in different items of a given test.

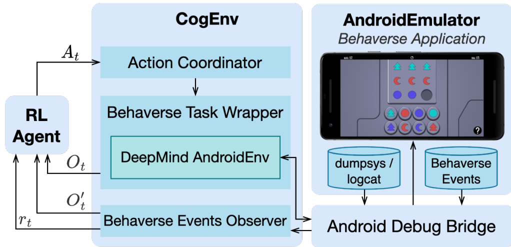
So they may use the same underlying computations.

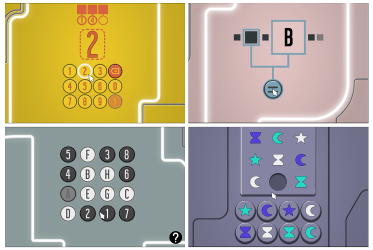
In scenario B, agents outperform humans but they underlying computations are different.

And in scenario C, agents outperform in some items, but are similar in the other ones.

So it may inform us about what kind of differences are there between the two types of agents

CogEnv



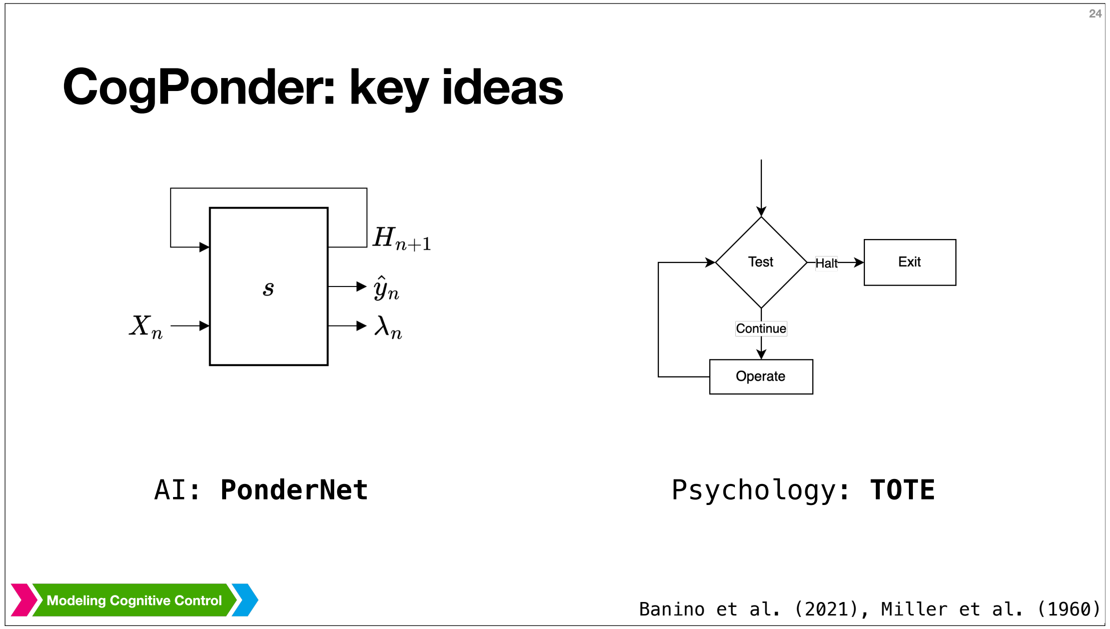


CogEnv, as presented here, was more technical and we create a Python package to enable these kinds of research.

How about the agent, how can we build a generic architecture for an agent that is interoperable and scalable?

More importantly, we are interested in cognitive control, so we also want to separate the act of control from the controlled act.

What would such a model look like? Interoperable, Scalable, and functionally separate control from the other parts of the model



There are two key ideas that may inspire us,

The first one is PonderNet framework.
Recently, people in AI proposed an adaptive deep-learning framework called PonderNet,

PonderNet is a recurrent model that not only produces responses for a given input, here X and y, but it also decides when to stop computing Or the recurrence loop. So for some inputs, the model is faster.

This model is similar to a model of control from psychology, called TOTE

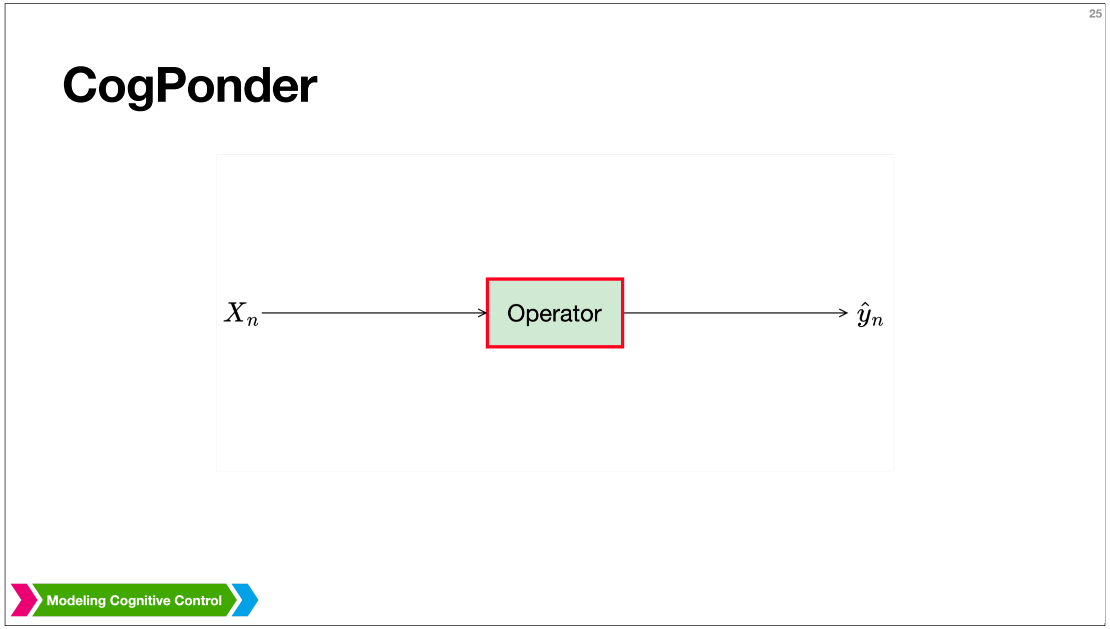
In tote, TEST, OPERATE, TEST, EXIT…

The model iterates over an operation until it reaches some condition, upon which it halts and produces response

This idea is for example used in DDM, a computational model for speeded binary decision, where upon reaching a threshold, a decision in made.

The problem here is that PonderNet does not align to human behavior, and TOTE and DDM are limited in terms of number of choices and scalability.

I thought, we can combine the two ideas and create a general model of cognitive control for explaining human behavior and then align AI agents to human behavior.

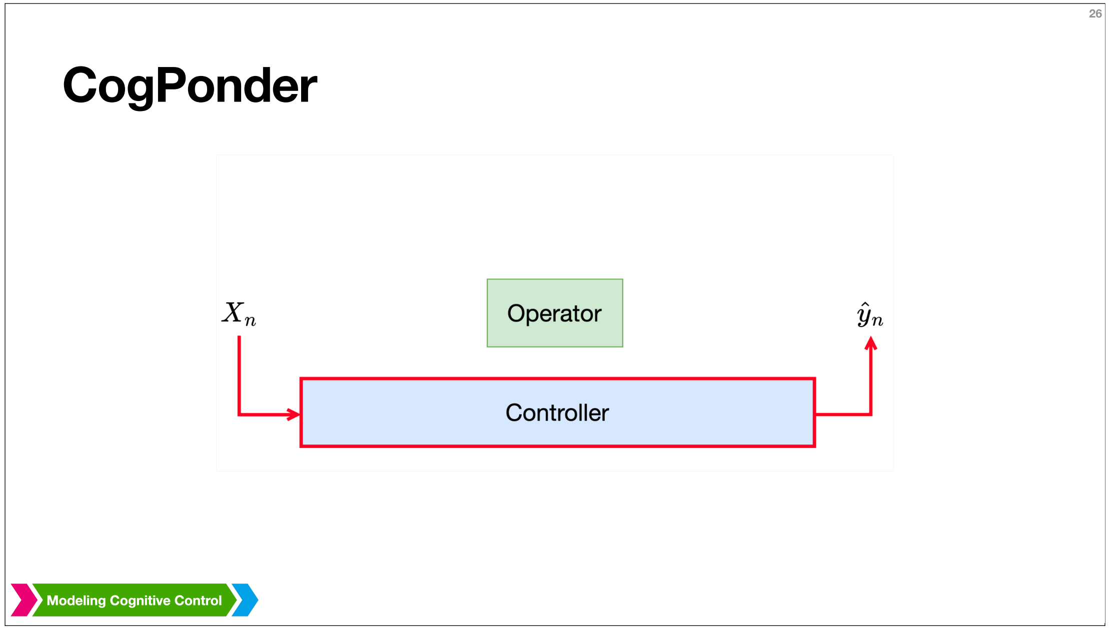


Here is how to build the model, which we call CogPonder

We start by a black box, operator here, that maps inputs X to output y.

We do not impose any constraints on the formats of the inputs and outputs. Any off-the-shelf model would work.

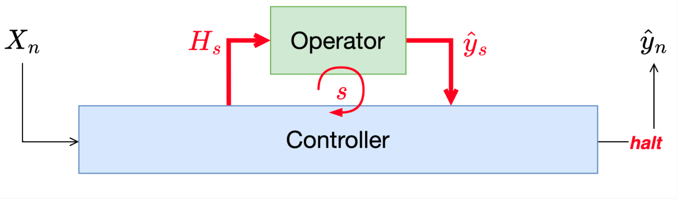
As long as the operator supports a task, it works. So it would work for simple cognitive control tests like Stroop and N-back AND complex tasks like video games.



We now intercept the input and output with a controller that functionally separates control mechanisms.

Inputs now instead go into the controller.

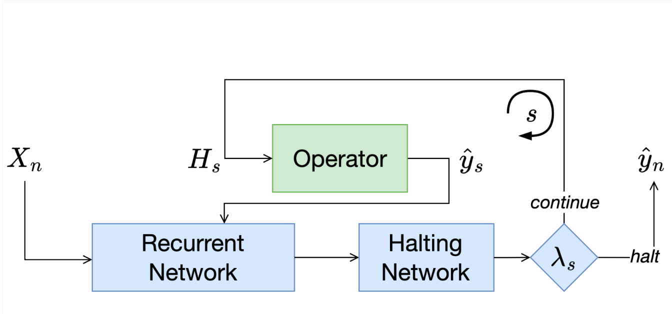
CogPonder



And the controller augment the operator.

Inspired by TOTE, we add a loop that terminates upon reaching some halting condition, which would result in response and response time.

CogPonder



Modeling Cognitive Control

Now,

we re-organize the inputs/outputs arrows, and replace the controller with the PonderNet halting networks.

From machine learning perspective, if the operator is differentiable then the whole model is differentiable.

Lambda on the bottom right is the probability of halting and it is learned from the data.

And upon halting, the controller selects the response from what operator provided.

This is the model, and it can be trained on the data, human data.

Evaluating CogPonder

Aligning to human behavior

$$\boldsymbol{L}_{\textbf{total}} = \sum_{s=1}^{S} \mathcal{L}_{\text{response}}(\hat{y}_n, y_n) p_s + \beta \boldsymbol{L}_{\text{time}} + \beta KL(p_s || d)$$

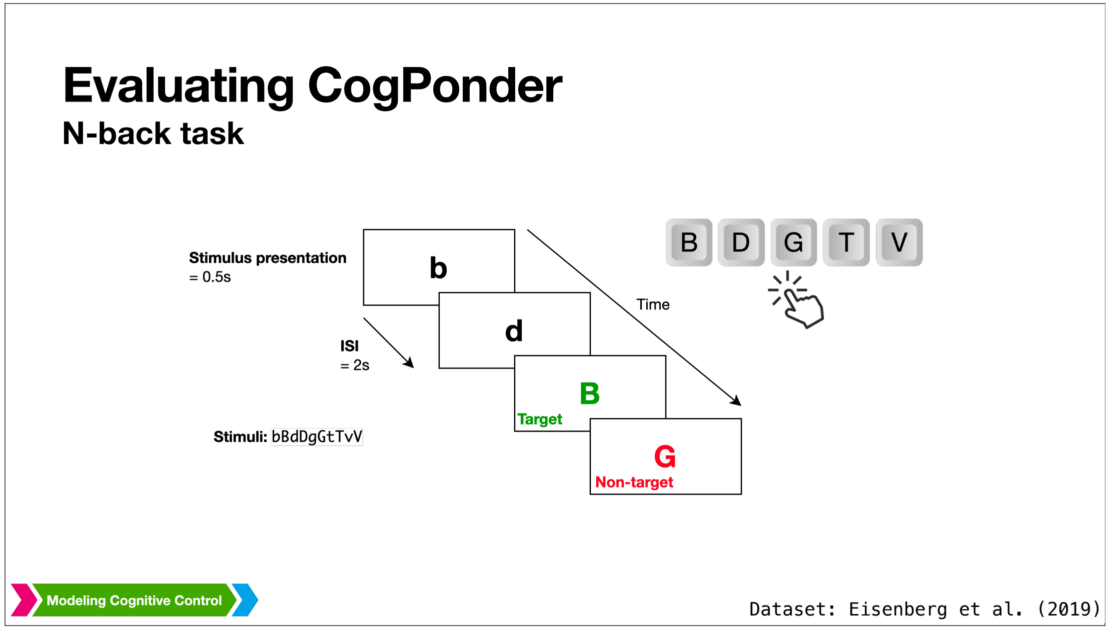
$$p_s = \lambda_s \prod_{j=1}^{s-1} (1 - \lambda_j)$$

Modeling Cognitive Control

The loss function for fitting the network would be like this:

<click>

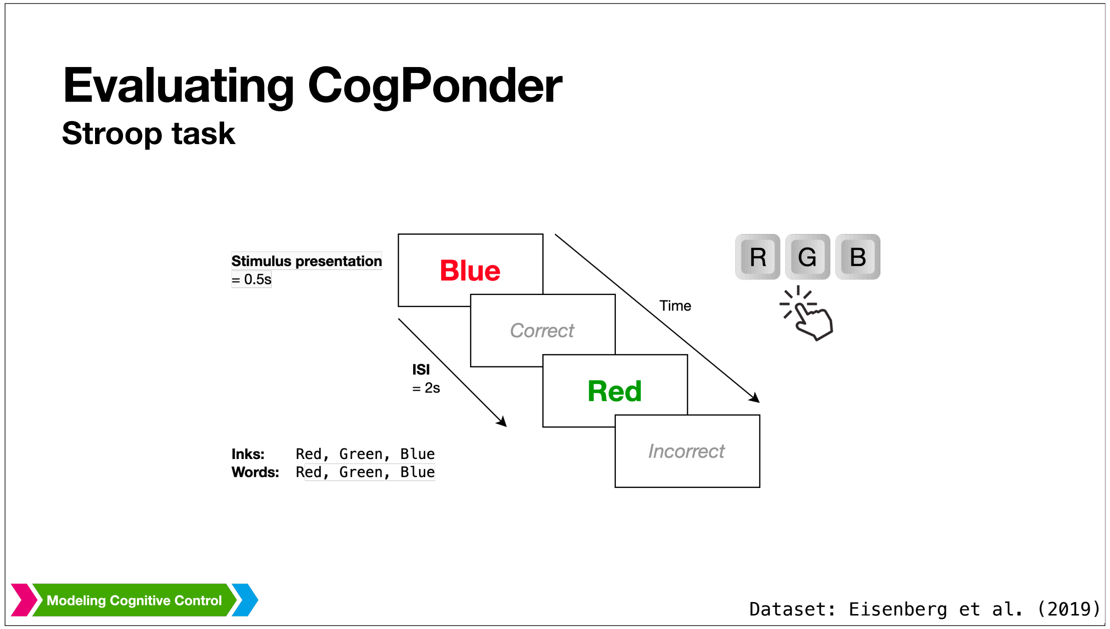
As a proof-of-concept to see if the CogPonder works, I trained an agent on human performance in two cognitive cognitive tests: N-back and stoop.



I used self-regulation ontology dataset, where human participants performed multiple tasks.

<CLICK>

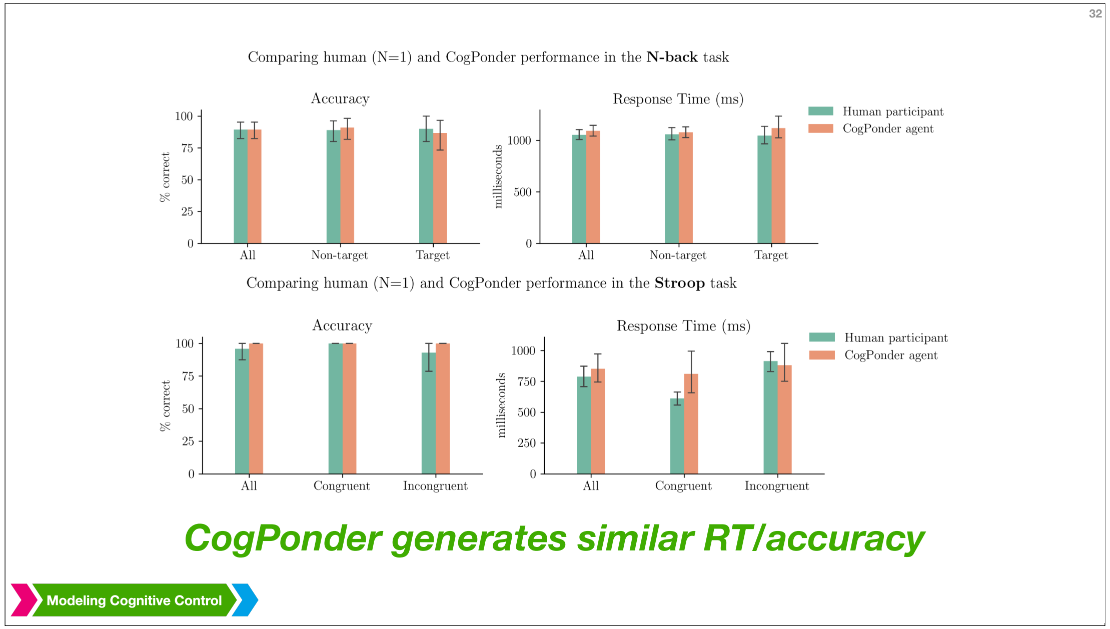
In the 2-back task, which involves working memory, participants are presented with a stream of letters and they are instructed to report for each letter whether it is the same letter as the one presented 2 letters ago.



In the Stroop task involves cognitive control, and participants are presented with a name of a color written in ink that is either congruent or incongruent with the word. They are asked to respond with the ink color.

For both tasks,

We used 75% of the trials to train the CogPonder agent,
and then used the remaining 25% to compare human performance against CogPonder agent.

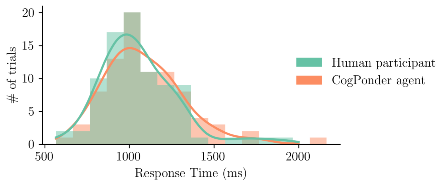


This initial results compares performance in one human participants and CogPonder,

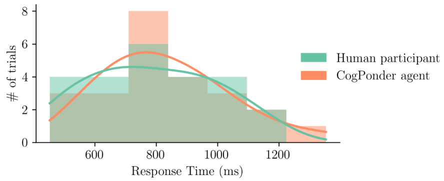
Human and AI agent perform similarity in terms of accuracy and response time in the N-back (on top) and Stroop (on the bottom) across different experimental conditions.

And the plots show performance only in the 25% test trials.

Comparing human (N=1) and CogPonder response time distributions in the **N-back** task



Comparing human (N=1) and CogPonder response time distributions in the **Stroop** task



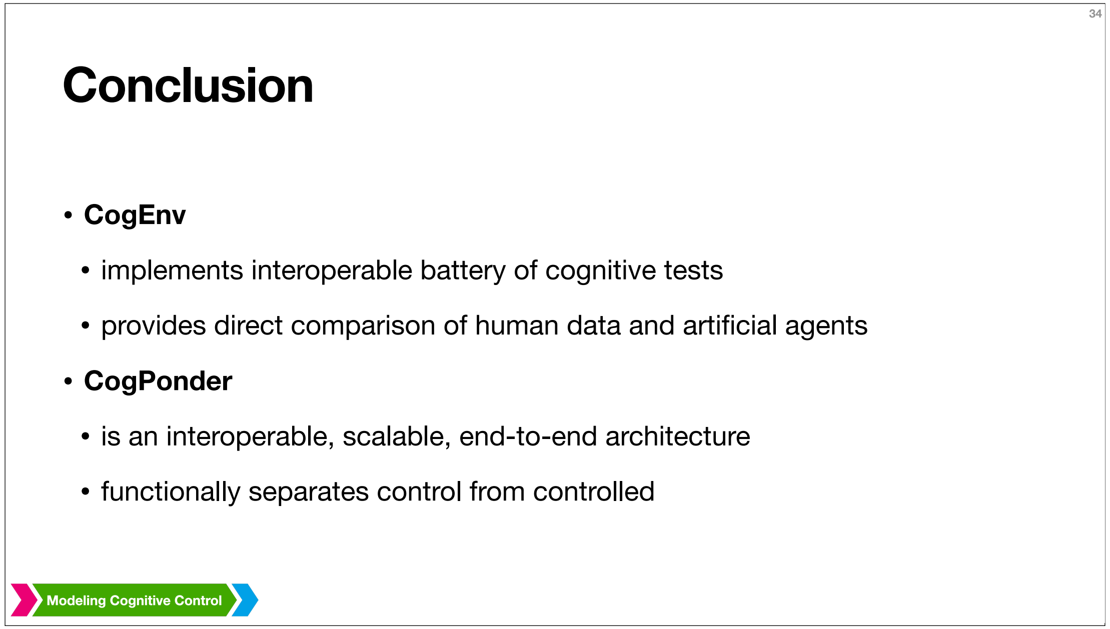
CogPonder generates similar RT distributions

Modeling Cognitive Control

And the CogPonder agent also generates similar response time distributions to human,

On the top for the N-back and on the bottom for the Stroop.

Overall, CogPonder generates similar behavior to human,



To summarize this part,

CogEnv implements an interoperable battery of cognitive tasks that can provide direct comparison of human data and AI agents,

And CogPonder is an again interoperable scalable architecture that functionally separates the act of control from controlled act

—

obviously the results are limited in a number of ways.

But I think there is a value in having a shared account of human cognition and artificial agents,

Such general computational models would ground cognitive control in tractable computations,
And for example provide shared benchmarks for humans and artificial agents,
and maybe inspire new tasks and new architectures that are inspired by human cognition.

Modeling Cognitive Control

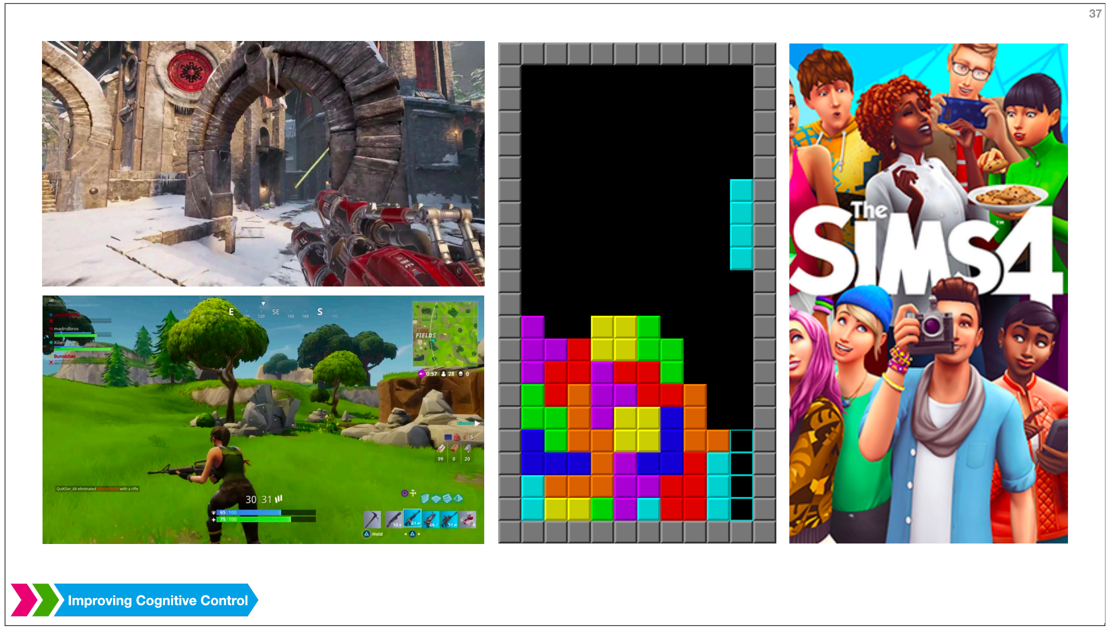
Improving Cognitive Control

Chapter 4
Chapter 5

And the last part of my thesis is about improving or training cognitive control.

This is important in AI because it relates to generalization.

This is also important in psychology because improving cognitive control could lead to improvements in real life.



Training cognition in a way that would generalize to many cognitive tests AND to real life is very hard.

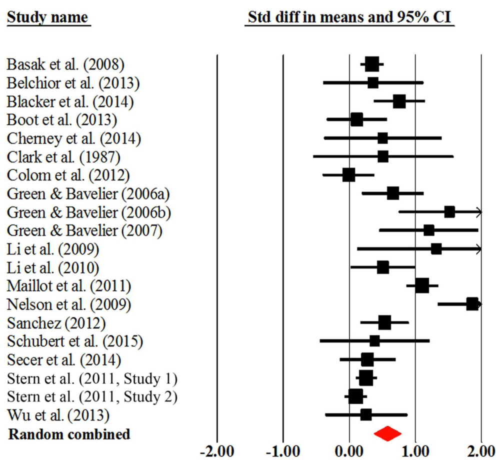
There is however evidence in psychology that transfer effect can be achieved by playing specific kinds of video games,

That is action video games,

<CLICK>

in contrast to other kinds of games like Tetris, a puzzle game

or the sims, a social simulation game



Overall efficacy of action video games training on all cognitive outcomes in adults

Improving Cognitive Control

Wang et al. (2016)

In fact, this meta-analysis across a wide range of cognitive tests by Wang and colleagues shows the overall effect of cognitive training with action video games on broad set of cognitive tests in adults…

The same effect has been observed in young adults and old adults, separately.

What explains AVGP transfer effects

H1) Domain-general? E.g., cognitive control

H2) Domain-specific? E.g., bottom-up attention, visual processing

H3) Task-specific? E.g., contrast perception

Improving Cognitive Control

The main hypothesis in the field is that action video gaming leads to broad cognitive benefits because it improves domain-general abilities, mainly cognitive and attentional control.

If this is true, then studying the effects of action video games becomes relevant to my thesis,

But there are also alternative hypotheses…

<CLICK>

Domain-specific hypothesis attributes the effects to improvements in specific cognitive domains, such as bottom-up attention or visual processing

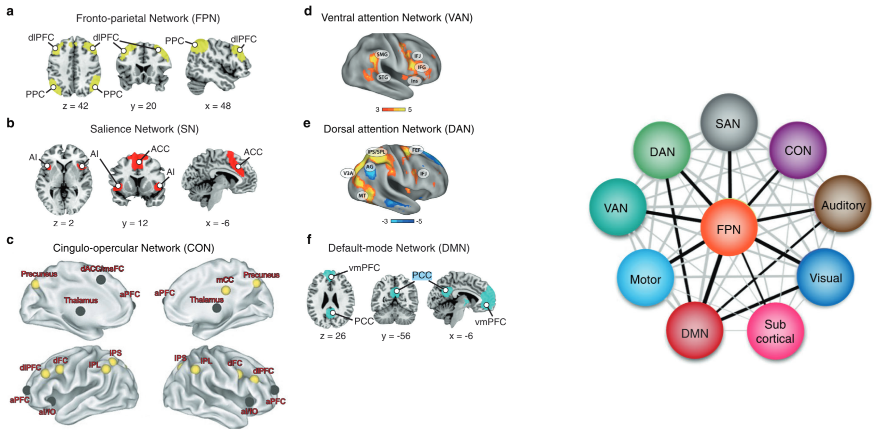
This is because these are the mechanics commonly used in action video games.

<CLICK>

or to task-specific improvements that does not readily transfer to other tasks. This includes for example contrast perception.

We want to tell apart these hypotheses and see which one is most plausible.

Cognitive control networks in the brain



Improving Cognitive Control

Menon and D'Esposito (2022)

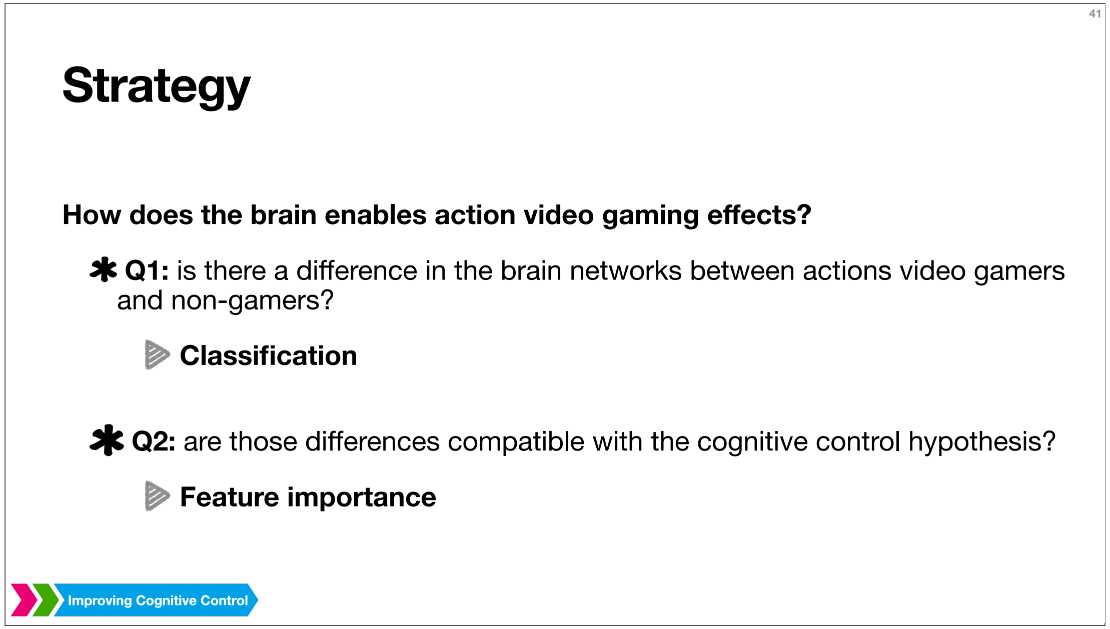
To answer this question we turn to the neuroscience of cognitive control which has identified multiple brain networks that are responsible for cognitive control related functions,

In particular, some of these networks are associated to domain general abilities like FPN in panel A for adaptive control, CON in C for maintaining control, and VAN in panel D for top-down attention.

<CLICK>

While these six networks isolated individual networks, a collections of those networks and connectivities between them have been linked to cognitive control more generally like connectivities between FPN and CON to lower-level SMN.

These networks are often observed during both rest and task. So we can study how action video gaming impacts those networks and connectivities between them.



My strategy was to look at the resting-state data and test

One) if there a difference in the functional connectivities between actions video gamers and non-gamers?

and if so, what features of the brain connectivity are altered. This is the second question.

For the first question, confirming the difference between two groups, I can use machine learning classification pipeline,

and for the second question, to find out the most diagnostic brain networks and connectivities, we can use feature importance analysis, like permutation feature importance or more recent ones like SHAP values.

Predictions

- **H1)** Domain-general?
  - ➡ Interaction between networks
- **H2)** Domain-specific?
  - ➡ Within specific networks
- **H3)** Task-specific?
  - ➡ No effect

under the domain-general cognitive control hypothesis we would expect a collection of networks or their relationships to differ,

under the domain-specific hypothesis we would expect specialized networks like dorsal attention network to differ between the two groups,

finally under the task specific we expect to see no difference at all in the resting state data or within sensorimotor network.

Data

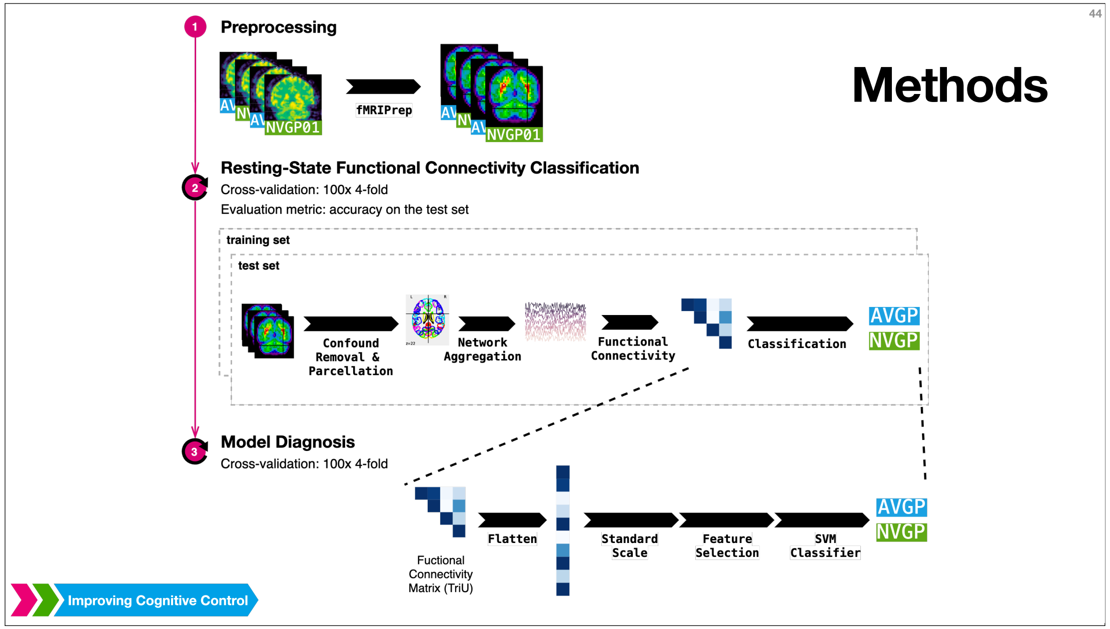
- 7'30" resting-state fMRI
- N = 32
- 16 habitual action video gamers

Improving Cognitive Control

Dataset: Föcker et al. (2018)

For the dataset, Julia Föcker and Daphne Bavelier kindly shared with us unpublished resting-state fMRI scans that accompanied their 2018 task fMRI study.

The data included about 7 minutes of resting-state data for 32 participants, 16 of which were habitual action video gamers and 16 were non-gamers.

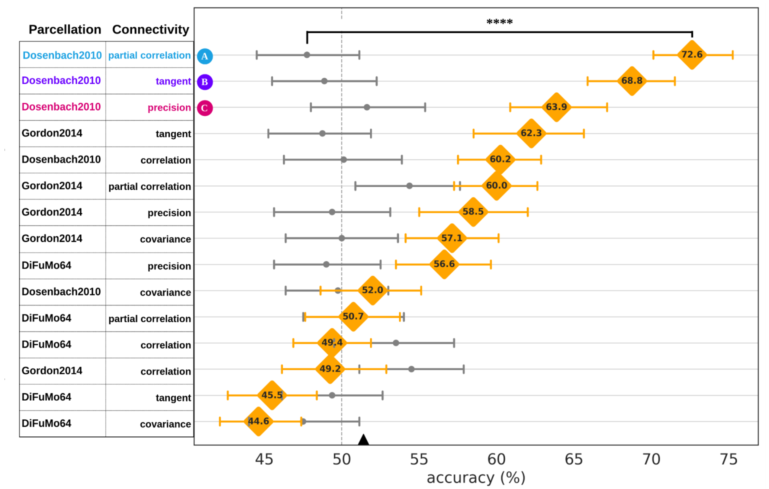


I've then used a fully reproducible pipeline that included three steps:

Preprocessing with fMRIprep

classification pipeline for functional connectivity using SVM,

And model diagnostic using permutation importance analysis



There is a signal in rs-fMRI to classify action video gamers

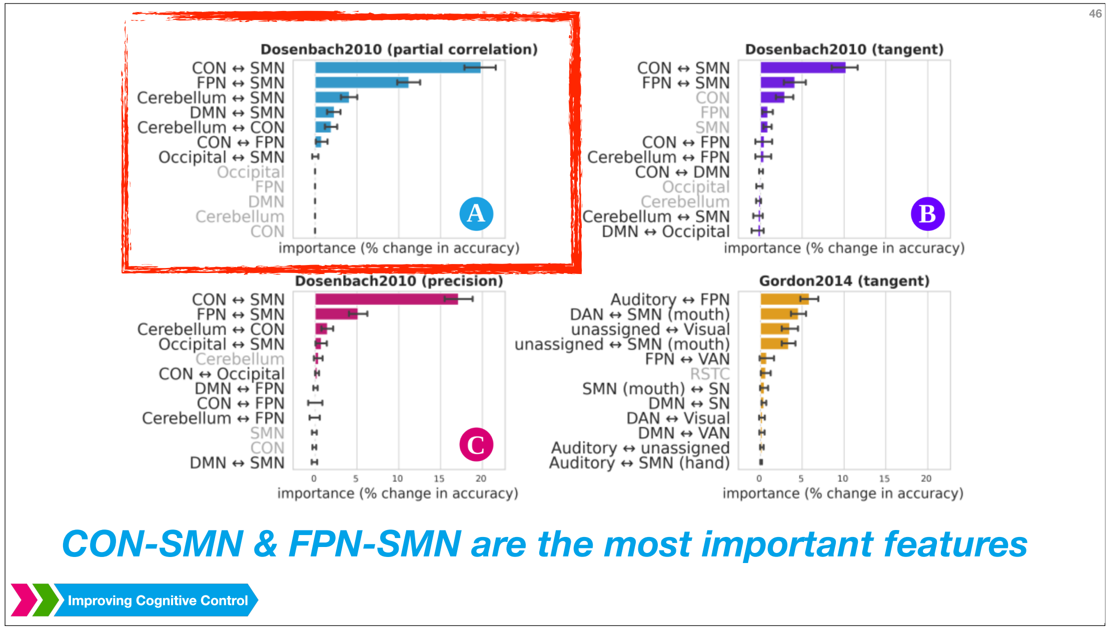
Here is the results:

On the x-axis you see the predictions accuracy on the 20% test split of the data.

And y-axis shows different models that we tested, including choices of parcellations and connectivity metrics.

The results show that classifying AVGP from non-gamers is indeed possible using functional connectivity.

with the best model on top reaching an accuracy level of 72 percent.



And the next results is the permutation feature importance results.

It can be seen the difference is not with specialized networks but distributed in the interaction between multiple control networks.

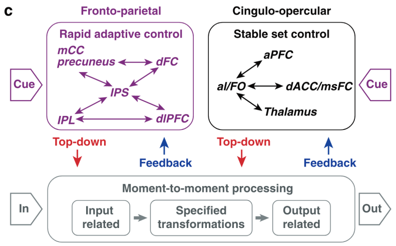
<CLICK>

Here for the best model, it can be seen that interaction between cingulo-opercular network and sensorimotor network and then between fronto-parietal network and sensorimotor network is the most diagnostic feature for the classification.

The key results are also robust to the different choices of connectivity metric. **In conclusion …**

Conclusion

- Resting-state fMRI contains useful information about cognitive control.
- Action video games training is characterized by differences across multiple distributed brain networks associated with cognitive control.



Improving Cognitive Control

Dosenbach et al. (2008)

The results are important for both practical and theorical reasons.

They show that resting-state connectivities contain useful information about cognitive control, and the information can be used for many different things.

Including the effect of actions video games training, which shows that the training is characterized by differences across multiple distributed brain networks that are associated with cognitive control, rather than isolated networks.

More specifically, the results support the dual network framework of cognitive control as suggested by Dosenbach 2008,

<CLICK>

which implies two separate control networks enable cognitive control in the brain.

Fronto-parietal network which is associated with rapid adaptive control,
And cingulo-opercular network which is associated with maintaining control

And the connectivity between these two control networks and lower-level sensorimotor network enables cognitive control

This is important because it may contribute to the development of more effective cognitive training programs or AI models that enable broad brain-inspired control.

So to recap my thesis, I've tried to define cognitive control, develop a computation account of cognitive control, and then study the impact of cognitive training in the brain.

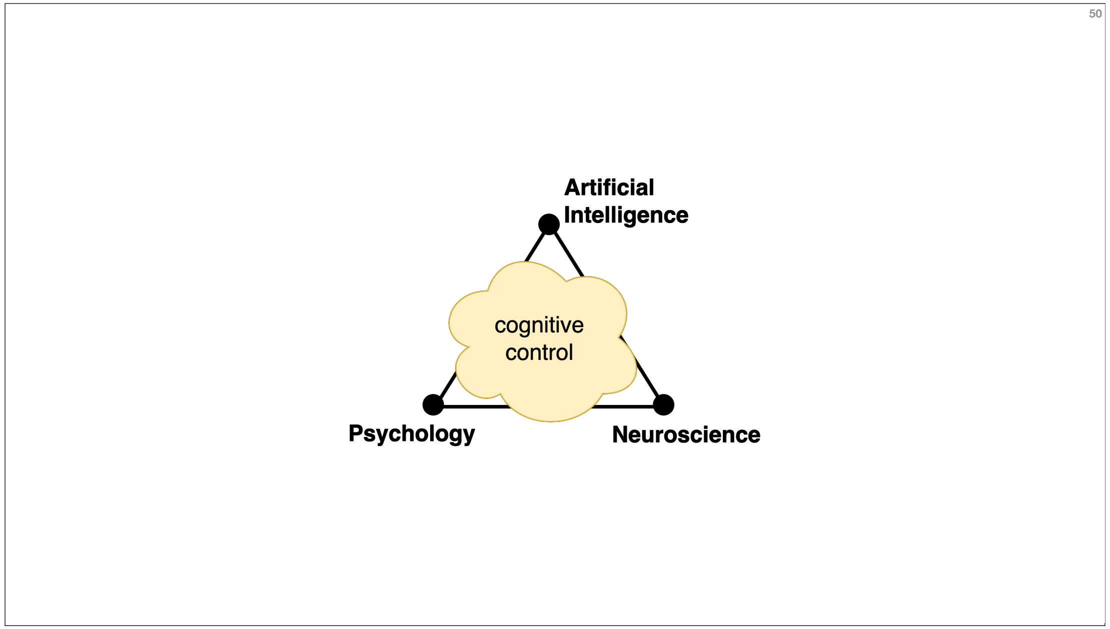
General conclusion

To get a better understanding of cognitive control…

- use data-driven knowledge models applied on scientific texts (adopt ontologies)
- study human behavior across a battery of cognitive tests and on complex tasks
- develop virtual environments that support the direct comparison of human behavior and computational models (e.g., RL agents)
- develop cognitive control computational models that actually perform tasks and provide scientific value
- study the neuroscience of cognitive control as it provides insights on computational mechanisms
- study the impact of cognitive control training as it may reveal the mechanisms underlying transfer (generalization)

In conclusion, my contributions imply that to get a better understanding of cognitive control …

<CLICK>



Back to this triangle, there is a value in the synergy between AI, psychology, and neuroscience.

And I’ve tried my studies to be the lines that connects these dots.

<CLICK>

ChatGPT

“Cognitive control refers to the ability to regulate and direct one’s thoughts, emotions, and actions in order to achieve specific goals or tasks. It involves the use of executive functions, such as attention, working memory, and reasoning, to monitor and modify one’s behavior in order to adapt to changing situations. This ability is essential for many everyday activities, such as problem-solving, decision-making, and goal-directed behavior. Cognitive control can be impaired in certain neurological and psychiatric disorders, such as attention-deficit/hyperactivity disorder (ADHD) and schizophrenia.”

Or as an alternative we can ask ChatGPT, as Pedro did, to explain cognitive control for us.

List of publications

Ansarinia, M., Schrater, P., & Cardoso-Leite, P. (2022). Linking Theories and Methods in Cognitive Sciences via Joint Embedding of the Scientific Literature: The Example of Cognitive Control. CogSci 2022: Annual Meeting of the Cognitive Science Society, Toronto, Canada. https://arxiv.org/abs/2203.11016

Ansarinia, M., Clocher, B., Defossez, A., Schmück, E., & Cardoso-Leite, P. (2022) CogEnv: A Reinforcement Learning Environment for Cognitive Tests. 2022 Conference on Cognitive Computational Neuroscience, San Francisco, CA.

Ansarinia, M., Cardoso-Leite, P. (Submitted to ICML2023) CogPonder: Towards a Computational Framework of General Cognitive Control.

Cardoso-Leite, P., Ansarinia, M., Schmück, E., & Bavelier, D. (2021). Training cognition with video games. The Oxford Handbook of Developmental Cognitive Neuroscience.

Ansarinia, M., Föcker, J., Lepsien, J., Bavelier, D., Cardoso-Leite, P. (in prep) Neural Correlates of Habitual Action Video Games Playing in Control-related Brain Networks.

Ansarinia, M., Lepsien, J., Cardoso-Leite, P. (Submitted to MPI-CBS). Neural Determinants of Action Video Game Training. Max Planck Institute for Human Cognitive and Brain Sciences.

Ansarinia, M., Mussack, D., Schrater, P., & Cardoso-Leite, P. (2019). A Formal Framework for Structured N-Back Stimuli Sequences. 2019 Conference on Cognitive Computational Neuroscience, Berlin, Germany.

Ansarinia, M., Mussack, D., Schrater, P., & Cardoso-Leite, P. (2019). A Multi-Objective Optimization Algorithm to Generate Unbiased Stimuli Sequences for Cognitive Tasks. Bernstein Conference 2019, Berlin, Germany.

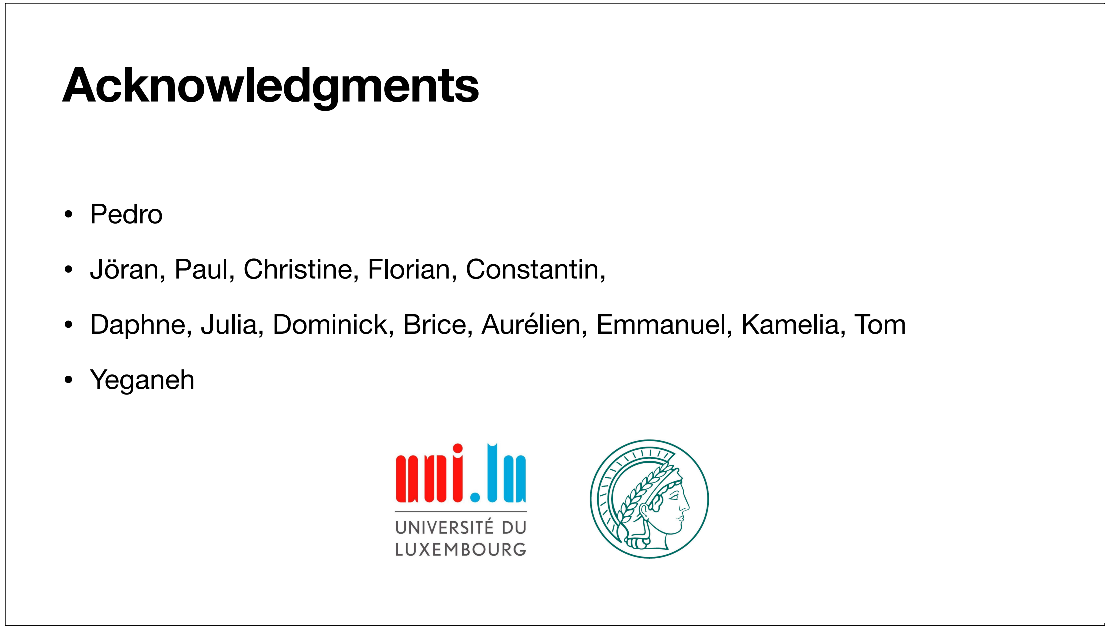
Defossez, A., Ansarinia, M., Clocher, B., Schmück, E., Schrater, P., & Cardoso-Leite, P. (2020). The structure of behavioral data. https://arxiv.org/abs/2012.12583

Ansarinia, M., & Cardoso-Leite, P. (2022a). Weighted Metapath2Vec Graph Embedding (Version v0.1.4) [Computer software]. https://doi.org/10.5281/zenodo.7096229

Ansarinia, M., & Cardoso-Leite, P. (2022b). Nilearn Extra: Drop-in extra features for Nilearn (Version v0.1.1) [Computer software]. https://doi.org/10.5281/zenodo.7096267

Ansarinia, M., & Cardoso-Leite, P. (2022c). UpSet2D: Hypergraphs Visualization Package (Version v0.1.4) [Computer software]. https://doi.org/10.5281/zenodo.7096256

Here is a list of my publications that goes in my thesis



And with that I finish may presentation,

I just want to thank my supervisor Pedro. I always felt safe and valued with him, and whenever I needed help he was there.

Thanks also to the Jury members and committee members, my mentors, lab mates, collaborators, and my beautiful wife, Yeganeh.

I also want to thank University of Luxembourg, and Max Planck Institute for Human Cognitive and Brain Sciences for making this research possible.