Outline for Maria’s thesis proposal

Aliens & CHaRLy: Part of the RO1

SLCN project: 2 separate pages

# Specific aims

(1 page)

## Introductory paragraph

* Opening sentence:
  + Abstract, hierarchical thinking is a crucial ability of humans, defines human intelligence
  + Compromised in neural and psychiatric conditions (long-term planning, abstract thinking)
* Current knowledge:
  + RL models have revolutionized our understanding of human cognition and brain function
  + Humans think hierarchically and abstractly (e.g., goals & sub-goals), but most current computational models do not take this into account
  + Only few, but promising, recent advances incorporate hierarchy into RL models:
    - Collins & Frank model about hierarchically-organized BG-PFC loops
    - Options are powerful in ML (Sutton, Precup, Singh)
* Gap in knowledge base / unmet need:
  + The Collins & Frank model, although built on top of plenty of evidence, has not been tested empirically. The options framework, although powerful in ML, has not been tested in humans.

## What, why, who paragraph

* Long-term goal: Understand the computational structure of abstract, hierarchical thinking, a hallmark of intelligence; understand how this is implemented in the brain; how problems with abstract hierarchical thinking can arise (and be cured); and understand the purpose of abstract thinking.
* Objective of this application: Develop specific, testable computational models of simplified tasks that require abstract hierarchical thinking; verify these models with human behavior; and identify neural correlates of hierarchical thinking in these tasks.
* Central hypothesis
  + Humans create hierarchical structure in a process involving curiosity, hypothesis formation, and skill learning, which can be formalized in the options framework
  + Hierarchical thinking is implemented in PFC-BG loops, whose activity can be measured from EEG signals
  + Humans’ use of hierarchical thinking is beneficial because it affords advantages in compositional environments and produces transferable skills
* Rationale (“why do I want to conduct the research?”): This research will produce objective, computational models of human hierarchical thinking. These models will capture the cognitive steps humans take when learning hierarchical representations, will explain brain activity associated with hierarchical thinking, and will shed light on the benefits of hierarchical thinking.

## Specific aims paragraph

Aim 1. Understand how humans build hierarchical reasoning structures on-line. [CHaRLy (& Nora’s sub-goals?)] The working hypothesis for this aim is that the creation of structure is driven by humans’ curiosity about novel events, in conjunction with the motivation to understand the mechanisms underlying such events, which leads to skills learning. Such a process can be modeled within the options framework of reinforcement learning.

Aim 2. Identify the neural mechanisms that underlie hierarchical reasoning. [Aliens EEG (& Nora EEG?)]

Aim 3. Explain the benefits of hierarchical reasoning. [Aliens (and CHaRLy?) in different environments -> when is hierarchy beneficial?]

## Payoff paragraph

* Expected outcomes: The proposed research will produce computational models of hierarchical, abstract thinking that can be used as process models (explaining human abstract thinking), elucidate the neural systems underlying abstract thinking, and shed light on its benefits.
* Generality regarding positive impact: Models will elucidate intelligence, and help cure diseases.

# Significance

(0.5-0.75 pages)

***Part 1: Statement of the problem and need for the research***.

The application of computational models has revolutionized our understanding of human an animal cognition as well as brain function (Schultz, 1997; Huys, …; Daw, ….; Tenenbaum / Griffiths). Reinforcement learning algorithms can reproduce human and animal decision making, and therefore serve as models for the underlying cognitive processes (Huys, Daw, Balleine). In addition, specific brain areas show responses that correspond to the crucial components of reinforcement learning computations, suggesting that the brain implements similar processes (Schultz, 1997; some human study).

Reinforcement learning models have revolutionized our understanding of simple, non-hierarchical learning and decision making. Nevertheless, a hallmark of human cognition is the ability to represent hierarchical structures and allow for abstract thought (Cohen? Old cognitive scientists? Chomsky? Miller?). Humans have the ability to break down tasks into goals and sub-goals (cite), and to reason about problems at different time scales, for example. *More research is needed in this field* because despite their importance, these crucial abilities are missing from most current reinforcement learning models.

A small number of recent studies has aimed to incorporate hierarchy into reinforcement learning models, with promising results (cite a few: casino, sub-goals, etc.). The *contribution of the proposed research* is to explain the cognitive processes involved in the creation of hierarchical structure, and to show that reinforcement learning occurs simultaneously at different levels of abstraction, thereby addressing the above-mentioned need.

***Part 2: Statement of significance***.

*Our proposed research will be significant because it will equip reinforcement learning models, which have provided elegant explanations for simple cognitive and neural processes, with hierarchical structure, a crucial element of human thought.*

***Part 3: Positive effect***.

The creation of such models is expected to have positive impact. It will further our understanding of complex human thought, and provide an explanatory and process model. It will advance research in the cognitive science, and maybe also in CS / ML / AI.

Many neurological and psychiatric conditions are associated with problems of abstract hierarchical thinking: PTSD, schizophrenia, ADHD, depression, etc. This is a huge problem for our society. Computational models of these functions will lead to the detailed understanding of the underlying mechanisms and potential sources of problems. They could be used to diagnose problems and to develop targeted treatments.

They can also be used to learn about the underlying brain function. This knowledge is crucial to diagnose neurological conditions.

# Innovation

(0.5-0.75 pages)

# Approach

(6 pages: 2 per project)

## Aim 1: Understand how humans build hierarchical reasoning structures on-line

***Introduction***. In order to reason hierarchically, humans need to infer the underlying hierarchical structure of a problem first. The *objective of this aim* is to understand how humans create such structure. To attain the objective of this aim, we will test the *hypothesis* that the creation of structure is driven by humans’ curiosity about unexpected outcomes, in conjunction with the motivation to understand the mechanisms underlying such outcomes, which leads to skills learning. We will test our hypothesis by using the *approach* of behavioral analysis and computational modeling. We will collect human data on a task that has an underlying hierarchical structure, and analyze behavioral markers of hierarchical reasoning to assess whether participants form hierarchical structure. We will then create a computational model of the task that aims to produce human-like behavior, showing the same learning trajectories, and making similar decisions. This model will be used to model the underlying human thought processes. The *rationale* for this aim is that successful completion of the proposed research will contribute a missing, fundamental element to our base of knowledge, without which the human ability to create abstract representations cannot be understood. The acquisition of such knowledge is critical to the development of improved therapeutic strategies for diseases related to abstract reasoning and decision making. When the proposed studies for aim 1 have been completed, it is our *expectation* that the created computational model will closely reproduce human hierarchical reasoning. The model will achieve this by combining curiosity and skills learning within the options framework of reinforcement learning. Such a finding would be of importance because it would allow, for the first time, to investigate the mechanisms underlying the creation of hierarchical structure. This key faculty is compromised in various psychiatric conditions, and mechanistic insight is a necessary step toward eventual treatment of such conditions.

***Justification & feasibility***

* Review of relevant literature
  + Explain what an MDP is (?)
  + Explain RL
  + Options framework
  + Something on curiosity?
  + NIPS papers where this came from
* Preliminary studies

Computational modeling of curiosity-driven hierarchical reinforcement learning. Previous research suggests that human hierarchical learning is driven by two main factors, curiosity-driven sub-goal selection, and the learning of sub-policies. We propose that humans develop curiosity about states with novel, unexpected outcomes. Curiosity increases the motivation to seek out these states, i.e., to select them as sub-goals. Humans then learn sub-policies to achieve these sub-goals through trial and error, as specified in the options framework of reinforcement learning. This hypothesis can be tested using a combination of behavioral and computational methods to characterize human hierarchical learning.

The creation of the computational model is a crucial step in every modeling research, and often the source of problems. We have implemented the algorithm in our preliminary work for this proposal and achieve promising results, as detailed below.

*Describing hierarchical learning as an MDP*. As a first step, we aimed to find an abstract description of hierarchical learning that would encompass a broad variety of situations encountered by humans, ranging from motor skill learning to conceptual learning. For example, the description should be broad enough to encompass how basic motor skills (e.g., moving individual fingers) can be combined into complex motor actions (e.g., playing the violin), but also how basic cognitive operations (e.g., recognizing an object) can be combined into more complex ones (e.g., recognizing relationships between objects).

We achieved this by formulating the task as an MDP consisting of discrete states *s*, actions *a*, and outcomes *r*. The states *s* correspond to different states in the environment (e.g., different constellations of objects on a table and the agent’s hands and fingers), actions *a* correspond to the agent’s initial inventory of basic actions (e.g., moving individual fingers), and outcomes *r* correspond to the environment’s responses to the agent’s actions (e.g., an object changes color upon touching; a teacher gives praise in response to the student’s behavior; etc.). Crucially, the environment’s responses *r* are not determined by just the agent’s most recent action, but by action sequences of potentially many steps. These action sequences are constructed in a way that makes the environment strictly hierarchical. The simplest action sequences consist of 1-step sequences. In other words, the agent needs to execute just one action *a* in order to produce *r*. At the next level of complexity, action sequences consist of the combination of two or more of these simplest action sequences, such that the agent needs to execute two or more of the simplest action sequences to produce *r*. Similarly, action sequences at the third level of abstraction consist of two or more action sequences at the second level, and so on. This scheme defines an environment which responds to the agent’s actions in complex ways, and depending on a potentially long history of past actions. Nevertheless, the environment has hierarchical structure, which an appropriate agent can exploit, and eventually understand and control the environment.

*Implementation of the curiosity-driven, hierarchical reinforcement learning agent*. We propose that a curiosity-driven, hierarchical reinforcement learning agent can achieve this. We implemented the agent in the following way. The agent selects actions and observes the environment’s responses. Whenever the agent observes a novel environmental response, elicited by a novel action sequence (i.e., a novel object creates a sound after being touched), the agent’s curiosity about this response jumps up. If the agent experiences the same environmental response again, elicited by the same action sequence, its curiosity drops. In this process, the agent updates its curiosity about all environmental responses it encounters, corresponding to multiple different action sequences. As mentioned before, the agent’s curiosity guides its sub-goal selection, such that the responses associated with the highest curiosity are most likely selected as sub-goals.

Once the agent selects a specific response as sub-goal, it tries to execute the action sequence that leads to this response. To achieve this, the agent first needs to learn the appropriate action sequence. This process is formalized through the options framework of hierarchical reinforcement learning. In brief, the agent learns the correct action sequences by acquiring a separate policy for each sub-goal, through trial-and-error reinforcement learning. Crucially, the agent can use less abstract action sequences as building blocks for more abstract action sequences, such that even very long action sequences require the sequential execution of just a small number of (abstract) actions. This capacity to reuse already-learned action sequences (so-called “options”) when learning new policies give the agent flexibility and direction at the same time. This feature constrains the agent in important ways, making sure that its behavior is not purely random, but meaningfully adapted to its environment. It also gives the agent flexibility in that it does not pre-determine entire action sequences, but their meaningful components. We next show that this algorithmic design led to specific behavior advantages for the agent, which would also be expected for humans.

*Behavior of the agent*. We first assessed how the behavior of our curiosity-driven, hierarchical agent differed from the behavior of other agents that were not curious and/or not hierarchical. We found that the CHRL agent explored its environment more efficiently than the other agents, which led to its better ability controlling it, in accordance with our expectations. The CHRL agent elicited a larger number of different environmental responses (Fig. xyzA), which shows that it explored a larger number of meaningful action sequences. The CHRL agent also elicited a larger number of responses overall, particularly at higher levels of abstraction, which shows that it selected its actions more efficiently with regard to maximizing environmental responses by action in a more goal-driven way than other agents (Fig. xyzB).

Finally, we assessed the motivation of the CHRL agent by assessing the RL values underlying its behavior. We saw that curiosity about environmental responses to basic actions increased and decreased very quickly, leading to a sharp bump in curiosity (Fig. xyz). Curiosity about responses to action sequences at higher levels of abstraction changed less rapidly, which shows that the agent developed interest more slowly, but also lost interest more slowly. While curiosity is high for an environmental response, the agent is more likely to select it as a sub-goal, and therefore learn the associated policy. In accordance with this, we also saw that the agent learned policies fastest for the least abstract action sequences, and took more time to acquire abstract action sequences (if these were acquired at all; data not shown). Interestingly in this framework, the agent is not required to acquire perfect policies on a lower level of abstraction before developing curiosity and selecting goals on higher levels. The agent always learns at all levels of abstraction simultaneously, even though the learning is usually greatest at just one or two levels. This is in accordance with human hierarchical learning. For example, simple motor sequences like grasping are pretty fast, stacking objects more slowly, and playing soccer at the level of a world championship very slowly.

***Research design***.

Characterization of human hierarchical learning.

Task design: In order to assess human hierarchical reinforcement learning, we will present research participants with a behavioral task that is an instance of the MDP explained above. In the task, participants will be allowed to perform sequences of actions, one action at a time, and observe the environmental responses after each. Participants will select actions through button presses; the environmental responses will consist in shapes, symbols, and patterns that appear on the computer screen for a brief amount of time following the action. Like above, environmental responses depend on action sequences, rather than just individual actions.

More specifically,

* The environmental response to each action sequence will be represented by a different symbol (sequence-symbol associations random or indicative of level?) -> each must be unique so that each is novel the first time around
* Symbols will be presented at random locations on the screen or according to level in the hierarchy?
* Symbols will stay on the screen until participants selects the next action -> needs to be able to move fast
* How many basic actions, how many levels, how long the action sequences?
* The same environment for everyone of a different one for every one?

Test conditions: Participants will interact with this general paradigm in different conditions, each of which will allow us to test specific hypotheses about hierarchical learning.

* Just explore, i.e., no rewards
  + show that learning occurs in the absence of rewards
  + determine which mechanisms underlie this learning? HRL suggests that sub-goal achievement acts as a pseudo-reward; test this.
* Transfer paradigms: after spending some time in one game, be transferred to a different game, with different background color, different symbols, at different locations for the responses; assess learning curves
  + Action sequences, at different levels, are either totally different from the ones in the original world; identical; or mostly identical but 1 out of many things has changed
  + Question: does interference impact different levels differently? Prediction: more basic levels will be affected more because they are more petrified, less flexible.
* Reach specific goals (collect as many objects of a certain kind as possible)
  + See how intrinsic (pseudo-rewards) and extrinsic (rewards) motivation are combined; how do extrinsic rewards affect sub-goal selection and curiosity?

Data collection: Data collection will be completed within one 60-minute session per participant. Participants will be recruited through UC Berkeley’s research participation pool (RPP). We are planning to enroll 40 participants for the study.

Planned analyses and expected results:

Behavioral analyses.

* Evidence for the formation of discrete action chunks
  + Slower RTs between chunks than within chunks
  + Larger transition probabilities between actions within the same action sequence than across action sequences
  + Transition between item1 and item2 in a chunk increases the transition probability between imte2 and item3 in the same chunk (etc.)
* Evidence for hierarchical learning: lower-level chunks are combined into higher-level chunks
  + Look at higher-level sequences that are performed more often:
    - Transition probabilities are higher between two actions when these actions are the last and first actions of two chunks that occur in the sequence, than when they are not
    - RTs are shorter in the same cases
    - Both change gradually over time; plotting this change should reveal something about how the sequence is learned (faster change -> more learning)
    - These changes should be correlated between transitions within the same chunk, but should not be correlated to transitions within other chunks (that have been picked equally often, and are at the same level of hierarchy)
  + Look at action sequences at different levels:
    - Are lower-level ones learned sooner than higher-level ones? (RTs and transition probs are indices of learning)
    - How many repetitions are necessary to learn lower-level ones compared to higher-level ones? Does this depend on how proficient participants are at the building blocks? Or does it depend on WM?
    - We can assess whether people prefer to develop a small number of very abstract skills, or to acquire a broad of set of less abstract skills.

Computational modeling.

In order to create a good computational model of human behavior in this task, we will take a standard model fitting approach. The most relevant steps will be the following (for a full description, see Daw book chapter). We will first create alternative models to the one presented above, in which crucial elements, such as curiosity and sub-goal learning, will be subsequently removed. We will then fit each model to participants’ behavioral data and calculate model fits. The winning model is the one that is able to reproduce human behavior most closely. We will also conduct behavioral analyses on the simulated data in order to verify that the signatures of human behavior mentioned above are reproduced.

Once a model with a good fit to human data is found, which also reproduces relevant signatures of human behavior, it can be interpreted. We expect the hierarchical model to be the winning model, which would suggest that both curiosity and sub-goal selection are involved in human hierarchical learning. The model also allows introspection into the hierarchical learning process. We can assess whether model-based curiosity matches up with how often participants perform each sub-goal, and whether the accuracy of model-based policies matches up with behavioral markers of policy knowledge (short RTs within; item1->item2 predicts item2->item3; etc.)

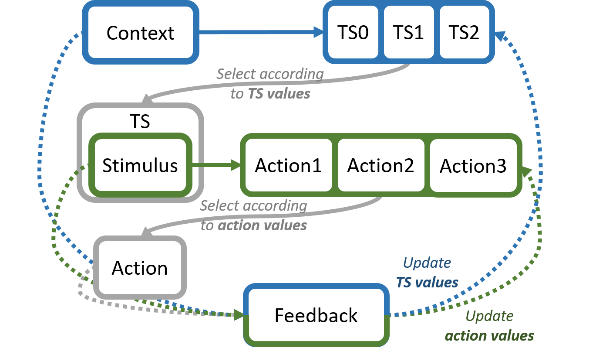
Potential problems & alternative strategies:

* The current model might not fit -> this is normal, it always requires tweaking -> analyze the behavioral data carefully to inform the model
* Curiosity and/or sub-goal selection might not be relevant, i.e., the winning model might be a different one than the one expected -> this is fine; we expect this model to perform best based on previous research, but if another model is better, we will learn that other processes are indeed more relevant.
* We can’t find any model -> we still have the behavioral markers that can tell us many details about human hierarchical learning

## Aim 2: Identify the neural mechanisms that underlie hierarchical reasoning

***Introduction***.

***Justification & feasibility***.

Review of relevant literature.

* Specifically, the brain’s Basal Ganglia show responses that correspond to reward prediction errors, and the Prefrontal Cortex shows responses that correspond to so-called *Q*-values.
* Reinforcement learning theory explains how agents acquire preferences ( “*Q*-values”) through interaction with their environment and how they make decisions based on these *Q*-values (Sutton & Barto). Reinforcement learning agents adjust their *Q*-values continuously in order to approximate the true contingencies of their environment. *Q*-values are adjusted based on a reward signal *r* that the environment produces in response to the agent’s action *a* in state *s*, *Q(s,a)=Q(s,a)+α(r-Q(s,a))*, whereby *r-Q(s,a)* is the reward prediction error, and *α* is the agent’s learning rate.

Figure : The hierarchical reinforcement learning model proposed by Collins & Frank.

* Collins and Frank (…) have proposed that the Prefrontal Cortex and the Basal Ganglia are connected through multiple “loops”, allowing for reinforcement learning at multiple levels of abstraction. The theory is based on ample neural as well as behavioral evidence (see Collins & Frank, …), but has not been tested formally yet. One *contribution of the proposed research* is to apply this model to human decision making in a hierarchical task.
* Another, formal, approach to hierarchical reinforcement learning is the options framework, developed as a tool for artificial intelligence (Sutton, Precup, Singh). Many powerful artificial-intelligence algorithms have been shown to have counterparts in human cognition (cite actor-critic; feature-based RL), but the options framework has not been tested yet in humans. The proposed research will also apply this model to human cognition.

Preliminary studies and results.

Analyzing human behavior during hierarchical reasoning.

Previous research has shown that humans condense behavioral knowledge into latent underlying rule structures, called Task Sets (TS), which can be applied in a number of different contexts. This allows for the generalization of a newly-acquired rule in one contexts to all contexts that are associated with the same underlying TS, and for the fast generalization of entire TS to novel situations. It remains an open question though how humans decide which TS to execute in response to a certain context, given that there is no strict one-to-one mapping between contexts and TS (because each TS can be associated with a number of contexts).

*Predictions*. We suggest that, just like TS are acquired through RL, the selection of TS is also guided by RL. This view has been proposed in a novel theory and is supported by a large collection of neural and behavioral evidence. In order to test this theory, we designed a behavioral task in which human participants first acquired three TS, and where then asked to indicate their preferences between the TS, and to select a TS for a novel context. Crucially, we had designed the task such that the three TS differed in RL values. The two tests therefore allowed us to assess whether participants acquired values for TS, as predicted by the theory, and whether they used these values to select TS in potentially new situations.

*Task design*. We designed the behavioral task in the following way.

*Results*. In accordance with our predictions, participants showed a preference for those contexts that had been associated with higher-valued TS compared to contexts associated with lower-valued TS (Fig. xyz). This shows that participants were sensitive to TS values, a novel prediction of our theory. Participants also preferred higher-valued TS to lower-valued ones in the generalization test of the experiment (Fig. xyz), i.e., participants preferentially applied higher-valued TS to novel contexts. This shows that TS values indeed guided action selection. Another piece of evidence for the influence of TS values comes from the analysis of participants’ errors. We found that the value of a TS was associated with the number of intrusion errors from a given TS, such that responses that were correct in higher-valued TS were more often incorrectly applied in other TS than responses that were correct in lower-valued TS. In addition, learning speed was associated with TS values, such that higher-valued TS were acquired faster than lower-valued ones. All of the results remained after controlling for the values of the stimuli themselves.

Taken together, the results support our theory. TS values influence TS selection.

* Behavioral markers (CogSci results)
* Computational model (what I have + what I’m planning)

***Research design***.

* Study 1: Behavioral markers and computational model of the creation of hierarchical structure
  + Task; RL values at different levels
  + Behavioral analysis
  + Computational model
  + Detailed expectations
* Study 2: Brain activity associated with the creation of hierarchical structure

***Expected outcomes***.

* Behavioral markers (CogSci results)
* Computational model (what I have + what I’m planning)
* EEG analyses (what I’m planning)

***Potential problems & alternative strategies***.

## Aim 3: computational purpose of hierarchical thinking

Rationale

Procedure: Aliens task and CHaRLy

Predictions / hypotheses: Hierarchical algorithms perform better than non-hierarchical algorithms in environments with certain characteristics (compositionality; transferability of skills). Do humans adjust their use of hierarchy to the situation, or do they always reason hierarchically?

Data analysis methods:

* Computational simulations
* Human behavior
* Human computational modeling

Alternative outcomes

# SLCN adult part

Background

Methods

Hypotheses / expected results