# 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 abstract hierarchical thinking in tightly-controlled experiments; verify these models with human behavior; and identify neural signatures of hierarchical thinking.
* 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 is necessary because the field of RL in cogsci can’t advance unless it integrates hierarchical thinking. Computational models are the future, and we need models of hierarchical thinking.

## Specific aims paragraph

**Aim 1. Understand the landscape of reinforcement learning in humans**. [SLCN]

**Aim 2. 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. This process can be modeled within the options framework of reinforcement learning.

**Aim 3. Identify the neural mechanisms underlying hierarchical reasoning**. [Aliens EEG (& Nora EEG?)]

## 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). 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 has provided us with elegant models of simple, non-hierarchical learning and decision making. Nevertheless, a hallmark of human cognition is the ability to represent hierarchical structures and abstract thought (Cohen? Old cognitive scientists? Chomsky? Miller?). For example, humans have the ability to break down tasks into goals and sub-goals (cite), or to reason about problems at different time scales and levels of abstraction. *More research is needed in this field* because despite their importance, these crucial abilities have not yet been integrated into current models of reinforcement learning.

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 extend these findings by explaining the cognitive processes involved in the creation of hierarchical structure, and by shedding light on the neural signatures underlying reinforcement learning 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 led to breakthroughs in the study of simple cognitive and neural processes, with hierarchical structure, a crucial element of human thought that has so far been neglected in this research area.*

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

We expect that the creation of hierarchical RL models will have significant positive impact. These models will further our understanding of complex human thought, and provide precise process models of the underlying cognitive processes. Achieving this goal would consist major progress for the field of cognitive science, and be a step toward integrating it with CS / ML / AI.

Many neurological and psychiatric conditions are associated with problems of abstract hierarchical thinking (e.g., PTSD, schizophrenia, ADHD, depression?). This is a huge problem for our society. Computational models of these functions will lead to the detailed understanding of the underlying mechanisms and can reveal 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

## Aim 2: 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 infer hierarchical 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 combined *approach* of computational modeling and rigorous behavioral analysis. We will first implement a computational algorithm that infers hierarchy based on the proposed mechanisms, to show that this approach leads to superior behavior than other, classical algorithms. We will then collect human data on a task with an underlying hierarchical structure, and analyze the characteristics of human hierarchical reasoning. Finally, we will use our computational algorithm as a model of human hierarchical learning and fit free model parameters in order to reproduce human-like behavior, with the same learning trajectories and underlying 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 2 have been completed, it is our *expectation* that we will possess a computational model of human reasoning, which explains the cognitive processes underlying hierarchical inference. 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 appropriate assessment and eventual treatment of such conditions.

***Background & Justification***.

*Background*. In recent years, reinforcement learning (RL) theory has allowed for breakthroughs in the fields of artificial intelligence and machine learning as well as cognitive and brain science, by proposing a precise computational model of learning and decision making. Nevertheless, the RL framework has some crucial limitations. Specifically, learning is driven entirely by “rewards”, i.e., those signals from the environment that the agent tries to maximize (e.g., food, praise, money, etc.).

*Problem*. This formulation of RL leads to two particular problems. First, traditional RL frameworks cannot explain how agents learn in environments that provide no rewards, or even in environments that are sparse in rewards, because learning cannot occur without reward signals. Nevertheless, biological systems excel at learning in such situations (just imagine high school as an example). Second, RL cannot explain how structured learning processes can arise in which simpler skills are acquired before, and facilitate learning of, more abstract skills, while the (reward) structure of the environment remains the same. Nevertheless, many biological systems, including humans, show this kind of hierarchical learning.

*Proposed solution*. We propose that both shortcomings arise because of the way rewards are usually implemented in RL frameworks, and that both can be resolved if rewards are implemented in a different, more flexible way. In current RL applications, rewards are hard-coded into the task environment, and importantly, the learning agents have no immediate access to them. We propose a formulation in which the reward function is instead created by the agent, and can be adapted flexibly over time. This allows for learning to occur in environments that would otherwise be sparse in rewards, because the agent can add rewards. It also allows for hierarchical learning because the reward structure can be adjusted to reward more difficult behaviors once simpler ones have been acquired. The challenge for this approach is to create an agent that will set the rewards in the places that actually facilitate learning (rather than, for example, an agent hijacking the reward function to achieve maximum reward without any learning). We propose that curiosity can provide the necessary information, as will be explained below.

*Rationale of the proposed research*. The goal of our proposed research is twofold. First, we will provide a proof of concept for our proposed adapted reinforcement learning mechanism. Second, we investigate the processes underlying human hierarchical learning, and specifically test the predictions of our algorithm in human learning.

*Preliminary results*.

*Problem formulation*. In order to test the proposed algorithm, we first created a hierarchical learning environment that would be abstract enough to encompass a broad variety of learning problems, ranging from the domain of motor skills (e.g., learning to grasp objects before stacking objects, before playing the violin) to cognitive skills (e.g., learning to recognize an object before recognizing a scene of multiple objects). We formulated the environment as a semi-Markov Decision Problem (semi-MDP), which is a formal description of an environment’s structure and used in most RL algorithms. In the semi-MDP, the agent selects an action and the environment responds with an outcome, which might be determined by a whole sequence of past actions, then the agent selects another action, the environment responds again, etc. The problem is hierarchical in that all action sequences that elicit responses are composed of shorter action sequences that also elicit (different) responses. In other words, shorter action sequences are the building blocks for longer action sequences. This problem is meant to capture the structure of hierarchical learning problems that humans typically encounter. For example, in the domain of motor learning, simpler motor skills (e.g., grasping) are combined to achieve more complex skills (e.g., stacking objects), which are themselves combined to achieve even more complex skills (e.g., playing the violin), etc.

*Implementation of the agent*. The crucial feature of the RL agent is that it constructs its own reward function, based on curiosity, as mentioned above. Specifically, the agent is maximally “curious” about environmental responses that is has rarely seen and becomes less curious the more often it has seen them. This mechanism assures that at the beginning, the agent is most curious about those environmental outcomes that can be elicited by simple and short action sequences, and acquires the skills necessary to achieve these outcomes using standard hierarchical RL (options framework). As the agent gets more proficient at eliciting these outcomes, it becomes less curious about them and more curious about outcomes that are instead elicited by more complex and longer action sequences, and that were elicited by accident while learning the simpler skills. The agent therefore gradually transitions from acquiring simple action sequences to acquiring more and more complex action sequences, guided by its own, novelty-based curiosity.

*Results*. This agent indeed showed superior learning compared to classic RL agents. It acquired a larger number of meaningful action sequences, which was evident in that the agent consistently elicited a larger number of environmental responses at all levels of abstraction (Fig. xyzA), and also discovered a larger number of meaningful action sequences overall (Fig. xyzB). In other words, the agent became more efficient at controlling its environment, i.e., at eliciting those responses that it deemed worthwhile. The mechanisms underlying this behavior were evident in the changes in the agent’s curiosity over time. As expected, curiosity about the outcomes of basic action sequences increased and decreased very rapidly, whereas curiosity about more abstract outcomes changed more slowly (Fig. xyz). This made sure that the agent set its own goals as for which skills to learn in a way that maximized potential learning.

***Proposed research***.

*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 abstract semi-MDP described above. In the computerized task, participants will perform sequences of actions (button presses), one at a time, and observe environmental responses (objects appearing on the computer screen). Like above, environmental responses will depend on hierarchical action sequences. Crucially, the task does not contain any explicit rewards. We expect that human participants will still acquire those action sequences that are meaningful in the given environment, i.e., that elicit environmental responses, by setting goals (rewards) according to their own curiosity.

*Data collection*. Participants will be recruited through UC Berkeley’s research participation pool (RPP). The task will take no more than 60 minutes to complete, and we are planning to enroll 40 participants for the study.

*Planned analyses and expected results*. We are planning to analyze the data form this experiment in two ways, using traditional behavioral methods as well as computational modeling.

*Behavioral analysis*. We will first create behavioral markers that assess whether participants perform the task hierarchically. One expected marker for hierarchy is a pattern of response times in which actions within the same sequence are executed with short delays, whereas transitions between different sequences are associated with longer delays. Another marker is a patterns of transition probabilities in which actions within the same sequence have higher transition probabilities than actions in different sequences.

Having established behavioral markers of hierarchical performance, we can then investigate the time course of participants’ acquisition of hierarchical structure, i.e., how shorter action sequences are combined into longer ones. We expect that over time, response times will decrease at the borders of short action sequences that are components of longer ones, such that eventually, response times will be distributed homogeneously throughout long action sequences. Such gradual changes in response times within longer action sequences therefore provide insight into the creation of hierarchy. Transition probabilities should show complementary changes, with initially low transition probabilities between actions at the borders of shorter sequences, which increase over time until they reach the levels of the remaining transition probabilities. Changes in response times and transition probabilities should thereby be correlated within the same action sequences, but show no correlation between different ones. In this way, different action sequences can serve as mutual control conditions.

Lastly, we can compare action sequences of different lengths, and thereby characterize individual factors of learning, such as the amount of feedback received for each action sequence, and the delays between two instances of feedback. It is an open question whether the same amount of feedback suffices to acquire short and longer action sequences, or whether longer sequences require more feedback to be learned. We can assess this question by comparing the markers of action sequence acquisition (response times and transition probabilities) between short and longer action sequences, including as predictors the amount of feedback received (i.e., the number of times the corresponding environmental response was elicited), the delays that lie between two executions of a sequence, and the length of a sequence. In this way, we will be able to determine whether the amount of feedback received is the only factor that determines learning (as would be assumed in traditional RL), or whether other factors that might be related more to memory also play a role.

*Computational modeling*. Having characterized human hierarchical learning with behavioral analyses, we will next use the algorithm introduced above as a computational model of the underlying thought process. We will identify the appropriate model using state-of-the-art model fitting techniques (Daw book chapter, a couple of Anne’s modeling papers?), which include steps like the comparison of alternative models, fitting free model parameters to human data using maximum-likelihood estimation, and simulating data using the model to verify that important behavioral trends are reproduced. Once the best algorithm has been identified, we will be able to interpret its structure as a model of human cognitive processes. The model will allow us insights into how human participants’ curiosity changes over time, and what learning rate underlies the acquisition of action sequences. We will also be able to determine which factors influence whether humans specialize, i.e., acquire a small number of very long action sequences, or whether they generalize, i.e., acquire a larger number of shorter action sequences.

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

## Aim 3: Identify the neural mechanisms underlying hierarchical reasoning

***Introduction***.

***Background & Justification***.

*The fundamental equation of RL*. 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 reward structure of the environment. Agents learn a specific *Q*-value for each action *a* in each state *s*, termed *Q(a|s)*. A *Q*-value is adjusted when the environment produces a reward *r* in response to the agent’s action *a* in state *s*, *Q(a|s)=Q(a|s)+α(r-Q(a|s))*. Here, *r-Q(a|s)* is also called the reward prediction error, and *α* specifies the agent’s learning rate.

*Neural foundations of RL*. Research in animals and humans has revealed that specific brain areas are sensitive to specific components of RL calculations. The basal ganglia (BG) show activity that corresponds to reward prediction errors, with larger responses for more unexpected rewards (Schultz, 1997; recent human fMRI study). Specific areas within the prefrontal cortex (PFC), on the other hand, show activity that scales with the size of *Q*-values. Taken together, it has been argued that RL is neurally implemented in a “loop” between BG and PFC.

*Human hierarchical reasoning with Task Sets (TS)*. Humans store behavioral rules as abstract rule structures, or Task Sets (TS) (Collins & Frank). TSs contain rules that go beyond any specific situation (or “context”), and that can be applied to many different contexts. For example, the abstract TS of how to use a Windows computer can be applied to any specific Windows computer. An advantage of storing rules as TS is that agents immediately have a complex behavioral repertoire when entering a completely new situation, which would be a problem for classic RL. For example, an agent would immediately have an idea of what to do with a new Windows computer.

*Problem*. Classic RL can explain how humans and animals learn in simple situations and tasks, but it cannot explain how we learn in more complicated, abstract situations. And although there is evidence that humans use TS (Collins & Frank), it is unknown how humans make the decision which TS to apply to any specific situation. The problem of selecting the right TS for the current context is in fact computationally intractable, demanding research into what cognitive and brain mechanisms allow humans to solve this problem.

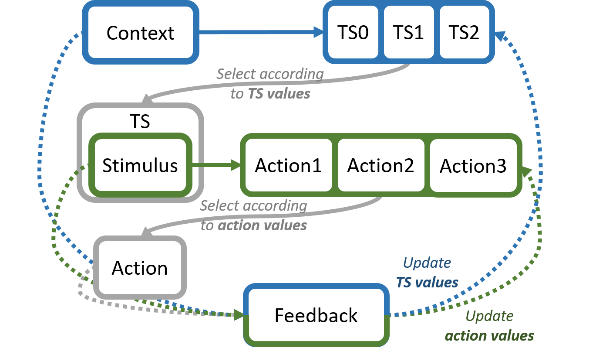
*Proposed solution*. We suggest that the selection of TS is guided by RL mechanisms (Collins & Frank). In this proposal, agents learn TS-values *Q(TS|c)* for each TS in each context *c*, using an updating rule similar to the one above, *Q(TS|c)=Q(TS|c)+α(r- Q(TS|c))*. Agents then select the appropriate TS for a context based on its learned TS-value. Neural evidence suggests that such a solution is indeed implemented in the PFC-BG loops mentioned above.

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

*Rationale of the proposed research*. The goal of our proposed research has three components. First, we want to provide behavioral evidence for the existence of two separable RL loops in humans. Because this evidence is crucial for the success of the remaining projects, we have completed a preliminary study that provides this evidence. Second, we want to develop a computational algorithm of the learning problem. Third, we want to assess the neural mechanisms underlying RL in TS, using behavioral as well as computational methods.

*Preliminary results*.

*Task design*. We designed a behavioral task in which human participants first acquire three TS, and were 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.

*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, and have been replicated in two independent datasets. Taken together, the results support our theory. TS values influence TS selection.

*Computational model*.

* We use *Q(TS|c)=Q(TS|c)+α(r- Q(TS|c))* to update TS values and *Q(a|s,TS)=Q(a|s,TS)+α(r- (a|s,TS))* to update action values.
* Softmax action selection; free parameters: …
* Simulations: behavior is qualitatively similar to humans: learning curves based on values; intrusion errors; generalization
* Gen-rec: we can recover parameter values

***Proposed research***.

*Task design*.

* Same task as above, with small modification to make it appropriate for EEG (jitter ITIs, make feedback stimuli more similar between different rewards?)
* Record EEG

*Data collection*. Participants will be recruited through UC Berkeley’s research participation pool (RPP). The task will take approximately 120 minutes to complete, and we are planning to enroll 40 participants for the study.

*Planned analyses and expected results*. We are planning to analyze the data form this experiment in two ways, using traditional behavioral methods as well as computational modeling.

*Replication of behavioral results*. We’ve already replicated it, so that should be fine.

*Computational modeling*. The hierarchical model will fit better than flat models

*Behavioral analysis of EEG results*.

* Reward-size-based differences in the reward-evoked potentials?
* Value-based differences in the stimulus-evoked potentials? (TS and item values)
* What about errors? (does decoded value agree with selected action?)
* What about cloudy phase? (can we decode when participants have found the correct TS?)
* Competition (can we get the value signals of both images?)
* Generalization (can we get the values of the selected TS?)

*Model-based EEG analysis*. Extract RL signatures from the model and regress against the EEG data.

* RPE signature? (for the mixed RPE? For the pure low-level RPE?)
* High- and low-level values?
* Current TS?

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