# Specific aims

## 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.

# Approach

## Aim 1: To understand the landscape of learning measures

***Introduction***. Human and animal learning has fascinated researchers for centuries, and a variety of tasks and measures has been employed to shed light on the processes that underlie it. Nevertheless, as a result of the mere variety of paradigms and measures employed, it has become increasingly difficult to compare results between studies and to draw general conclusions about learning. The *objective of this aim* is to take a step toward resolving this problem by mapping out the relationships between different commonly used measures, and between tasks that span relevant dimensions for learning. This research is necessary to provide a full understanding of individual measures of learning. In addition, only the joint consideration of different measures will allow us to draw conclusions about the mechanisms that underlie learning in general, and its neural substrates. This knowledge is important because it will allow to shed light on individual differences and can help elucidate the mechanisms underlying aberrant processes in neural or psychiatric conditions.

***Background & Justification***.

*Background*. Instrumental learning has been studied using a variety of experimental designs in animals and humans. Some learning tasks employ deterministic feedback (rewarding all correct responses) and others probabilistic feedback (rewarding only a fraction of correct responses). Some tasks focus on the acquisition of one set of behaviors from beginning to end, whereas others employ reversal learning, the acquisition of new responses that contradict previously acquired behavior. Studies also differ in the type of feedback employed, spanning, for example, positive feedback (reward), negative feedback (punishment), primary rewards (e.g., food), and secondary rewards (e.g., points to win, money).

*Problem*. Most likely, different neural learning processes are involved to different degrees in all these tasks. The striatal dopamine system, for example, has been argued to underlie reinforcement learning, but this has almost exclusively been studied in probabilistic tasks in which the reward history needs to be integrated across many trials. The prefrontal cortex has often been associated with learning, but not all studies show its involvement and it is still unclear which learning processes it underlies. Similarly, the hippocampus has been implicated in certain memory processes that are important for leaning, but its exact role has remained unclear.

Individual studies therefore lack the ability to pinpoint mechanisms of instrumental learning that go beyond the specific task design. The involvement of the striatum in tasks that require the integration of many subsequent outcomes, for example, does not necessitate the involvement of the striatum in all learning tasks. This presents a serious problem to the field of learning research: we are often unable to relate findings from different studies to each other, and we do not know which processes underlie instrumental learning irrespective of task design.

*Proposed solution*. As a step toward resolving this issue, we will conduct a combined study, employing multiple tasks that span several relevant dimensions of learning. The combination of several tasks and measures will allow us to reveal to which degree measures that are assumed to reflect the same processes indeed co-vary across tasks. This research will also shed light on the general learning processes that underlie differences across tasks.

***Proposed research***.

*Experimental design*. We recruited 53 adult human participants to complete four independent reinforcement learning tasks that have previously been used in the learning literature. Together, the tasks spanned crucial dimensions of learning, including tasks with probabilistic and deterministic feedback, with and without reversal, and with and without negative feedback.

*Data analysis*. We will first assess learning independently in each task and then assess correspondences between the tasks. We will focus on typically assessed behavioral markers of learning, such as employed strategy (e.g., win stay-loose shift) and learning accuracy, as well as computational markers of learning, such as learning rate, decision temperature, and forgetting. We will thereby obtain measures of various aspects of learning for each participant. We will then use a data-driven approach, such as factor analysis or principle component analysis, to identify the common factors that underlie differences between participants across tasks and measures.

*Expected results*. We expect that the common factors identified by our method will include aspects of learning that have previously been identified as crucial, such as learning speed and contributions from working memory. Such a finding would give credence to the claims that these factors play significant roles for instrumental learning across tasks. It is also possible that our analyses will reveal additional factors of importance that have not been identified in the previous literature, thereby opening up news ways for future exploration. One such factor might be change detection, which is rarely studied as a contributor to instrumental learning, but might explain substantial variance, especially in tasks that involve reversals. Lastly, our method will help us elucidate which specific measures are associated with which learning factors and to which degrees. For example, not only computational measures of working memory might load onto an expected memory-based factor, but also behavioral strategies that are related to working memory, such as the win stay-loose shift strategy. Taken together, shared variance between measures will shed light on the factors that underlie learning across tasks. Associations between individual measures and these underlying factors will provide insights into the appropriate interpretation of each measure.

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

*Selection of tasks*. One potential problem of the proposed study is the selection of the four tasks that were employed in our study. The tasks might be too specific to encompass all dimensions that are of relevance to instrumental learning. Although it is likely that no collection of tasks will ever be able to span all potential dimensions, we selected the tasks in a way to maximize variation, by including several factors that have previously been identified as crucial to instrumental learning.

*Statistical methods*. Principle component analysis is a standard technique for identifying a small number of factors that underlie variation across many measures, and has been used in many psychological studies. Nevertheless, other methods of data-driven research, such as k-means clustering or independent-component analysis, could be of relevance as well.

*Sample size*. If we find our study to be underpowered in terms of sample size, we will recruit an additional sample from the university’s research participation pool to replicate results.

## 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. It is our *hypothesis* that the creation of structure is driven by two factors, (1) the motivation to explore outcomes that have occurred unexpectedly, and (2) the acquisition of skills during this form of exploration. We will test our hypothesis by using the combined *approach* of computational modeling and behavioral analysis. We will develop a computational algorithm that infers hierarchy based on the proposed mechanisms, and show that this approach leads to improved learning. We will then assess human learning in a hierarchical task to characterize human hierarchical reasoning. Finally, we will bring both components together by using the computational algorithm as a model of human hierarchical learning. This will allow us to assess the mechanisms of structure learning on an individual basis, such as participants’ learning rate or decision noise.

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. Hierarchical reasoning is a key faculty which is compromised in various psychiatric conditions [XX], and mechanistic insight is a necessary step toward appropriate assessment and eventual treatment of such conditions. The acquisition of such knowledge is therefore critical to the development of improved therapeutic strategies for diseases related to abstract reasoning and decision making.

***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 in the cognitive and brain sciences, by proposing a precise computational model of learning and decision making. Nevertheless, research based on the RL framework has encountered some crucial limitations. Specifically, learning is driven entirely by “rewards”, which constitute those outcomes in the environment that the agent tries to maximize (e.g., food, praise, achievement of a goal).

*Problem*. This formulation of RL leads to two problems. First, by definition, learning cannot occur without rewards, but humans show considerable learning in situations that seemingly lack rewards (e.g., learning about an event by reading a book). Second, RL cannot explain how structured learning arises in which simple skills are acquired before complex skills, although the environment’s reward structure 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 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 learning agents have no influence on 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 in environments that would otherwise be sparse in rewards because the agent can add rewards. For example, hearing what happened next might provide an appropriate reward when reading a book. This approach also allows for hierarchical learning because the reward structure can be adjusted to reward more difficult behaviors once simpler ones have been acquired. In our formulation, the agent’s intrinsic “curiosity” is used as the reward signal, whereby curiosity is a function of the novelty of experienced outcomes.

*Rationale of the proposed research*. The goal of our proposed research is twofold. First, we will provide proof of concept for the curiosity-based hierarchical reinforcement learning (CHRL) algorithm. Second, we will investigate the processes underlying human hierarchical learning, and relate them to this algorithm.

*Preliminary results*.

*Problem formulation*. In order to test our theory, we first created a hierarchical learning environment that is abstract enough to encompass learning problems in a variety of domains, ranging from motor skill learning to the acquisition of cognitive skills. In the environment, specific action sequences elicit environmental outcomes (e.g., touching a box might change its color). Crucially, the environment contains action sequences of different lengths, whereby longer sequences are composed of shorter sequences. In other words, simpler sequences (e.g., touching a box) provide the building blocks for more complex action sequences (e.g., touching several boxes in order produces a new box). This environment can be formulated as a semi-Markov Decision Problem.

*Implementation of the agent*. The CHRL agent has two crucial features. First, it constructs its own reward function based on the novelty of each environmental outcome. This reward function can be interpreted as the agent’s “curiosity” about each outcome. Second, the agent acquires skills to achieve specific outcomes, through trial and error RL.

*Results*. When the CHRL agent starts exploring a new environment, it accidentally elicits outcomes by executing short action sequences. The agent develops curiosity for these novel outcomes, and is therefore motivated to acquire the skills necessary to achieve them through trial-and-error learning. As the agent acquires skills, the corresponding outcomes decrease in novelty, and therefore curiosity. At the same time, by virtue of the hierarchical structure of the environment, the execution of sequences of action sequences accidentally elicits novel outcomes. These outcomes increase the agent’s curiosity (Fig. xyz) and motivate new skills learning, which completes the circle. The CHRL agent therefore gradually transitions from acquiring simple action sequences to acquiring more complex action sequences, guided by its novelty-based curiosity. This learning pattern differs from classic RL agents and allowed the CHRL agent to perform a larger number of meaningful action sequences in the hierarchical environment (Fig. xyzA), at all levels of abstraction (Fig. xyzB). Thus, the CHRL agent created a reward function that maximized learning. Over time, the CHRL agent became more efficient at controlling its environment, i.e., at eliciting those responses that it deemed worthwhile.

***Proposed research***.

*Task design*. In order to assess human hierarchical reinforcement learning, we will present research participants with a hierarchical task similar to the one for the CHRL agent. Participants will perform sequences of actions (button presses), one at a time, and observe the resulting outcomes (boxes changing color or appearing). Like above, outcomes will depend on hierarchical action sequences. Crucially, the task does not contain explicit rewards. We expect that human participants will nevertheless acquire action sequences to elicit specific outcomes. The mechanisms behind this is expected to be curiosity-driven hierarchical RL, as implemented in the CHRL agent.

*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 will enroll 40 participants.

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

*Behavioral analysis*. We will first create behavioral markers that reflect whether participants perform the task hierarchically. One 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 pattern of transition probabilities in which actions within the same sequence have higher transition probabilities than actions in different sequences (Desrochers & Badre, 2016; Balleine?). We can then investigate the time course of participants’ acquisition of hierarchical structure, i.e., how shorter action sequences are combined into longer ones, and whether this matches the predictions of our model.

In a separate set of analyses, we will assess which factors contribute to hierarchical learning. We expect that besides the history of actions and outcomes, the fundamental component of RL, other factors might play additional roles that are associated with different memory systems, due to the differences in time needed to execute action sequences of different lengths, and/or directly with hierarchy.

*Computational modeling*. Having characterized human hierarchical learning using behavioral methods, we will next create a computational model of the underlying thought process, based on the CHRL algorithm. We will use state-of-the-art model fitting techniques to create a model that appropriately captures human behavior (Daw book chapter, a couple of Anne’s modeling papers?). This model will provide insight into the components of human hierarchical learning, and will allow us to assess individual differences in learning.

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

*Human learning is not based on CHRL*. A major concern might be that human hierarchical learning is not actually driven by the mechanisms of CHRL, as hypothesized. This possibility is the reason why we are planning to conduct extensive behavioral analyses on the human learning task before model fitting. The behavioral analyses will reveal which components play a crucial role for hierarchical learning, and we will adapt the CHRL algorithm based on these results.

*The CHRL model is too complex*. Another reason for concern might be the complexity of the proposed learning mechanism. Indeed, it is more challenging to fit more complex models, such as the proposed CHRL. Nevertheless, there are several strategies to avoid this problem. We might be able to simplify the algorithm, in concordance with the results of the human behavioral analyses. We can also turn to more sophisticated modeling techniques that allow for the fitting of more complex models.

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

***Introduction***. The ability to reason abstractly underlies feats of human cognition that are unparalleled in other animals as well as state-of-the-art artificial-intelligence systems. One theory about abstract reasoning is based on task sets (TS; cite), behavioral rules that are abstract from specific contexts and allow for flexible, generalizable behaviors. Nevertheless, it is unclear how humans decide which TS is appropriate for a specific context. RL theory, on the other hand, can explain learning and decision making that is tied to a specific context, but lacks the flexibility and generalizability of TS theory. The goal of this aim is to integrate TS and RL in order to develop a computational model of abstract human reasoning and decision making. Such a model is importance because it can be used as a process model of abstract reasoning, and might shed light on differences between individuals with and without mental diseases.

***Background & Justification***.

*The fundamental equation of RL*. Reinforcement learning (RL) explains how agents acquire preferences (“*Q*-values”) through interaction with an environment and how they make decisions based on these *Q*-values (Sutton & Barto). Agents learn a *Q*-value *Q(a|s)* for each action *a* that can be executed in each state *s* of the environment. Selecting actions based on *Q*-values leads to successful behavior because *Q*-values are continuously adjusted based on feedback from the environment, and eventually converge to the true underlying reward structure. *Q*-values are adjusted when the agent receives a reward *r* after takingaction *a* in state *s*, *Q(a|s)=Q(a|s)+α(r-Q(a|s))*. In this equation, *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 brain areas that are sensitive to RL calculations, suggesting that the brain’s learning mechanisms implement a form of RL. For example, activity in the basal ganglia (BG) increases whenever reward prediction errors arise (Schultz, 1997; recent human fMRI study). Activity in the prefrontal cortex (PFC), on the other hand, often scales with *Q*-values. It has therefore been suggested that RL is neurally implemented in a “loop” between BG and PFC. Add stuff about EEG!

*Human hierarchical reasoning with Task Sets (TS)*. RL theory can explain well how agents learn to associate specific actions with specific stimuli. Nevertheless, RL cannot explain how agents select responses to unknown stimuli, or how a whole set of behavioral rules could be reapplied in a new situation. This kind of flexibility and generalization has been formalized in the framework of Task Sets (TS). It contends that humans create TSs, which contain entire behavioral strategies that are abstract from any concrete context (Collins & Frank). For example, the TS of how to use a Windows computer contains the set of behavioral rules that can be applied to any specific Windows computer. Agents equipped with TS therefore have entire behavioral strategies available when entering entirely new contexts.

*Problem*. Stunningly, humans are very good at selecting appropriate TSs, although identifying the optimal TS for a given context requires calculations that are computationally intractable. The fact that the required calculation is intractable suggests that humans instead use a simpler, approximate strategy for TS selection.

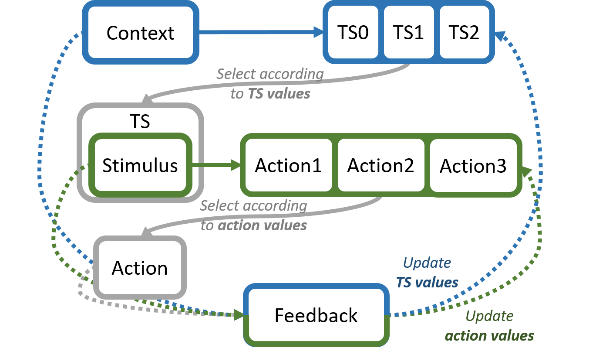
*Proposed solution*. We suggest that the combination of the TS framework with RL can provide a solution to this problem. We propose that humans learn an approximate strategy for TS selection using RL (Collins & Frank). In this proposal, agents learn “TS-values” *Q(TS|c)* for each *TS* in each context *c*, just like they learn Q-values *Q(a|s)* for each action *a* in response to each stimulus *s*. The update rule for TS-values is similar to the update rule for *Q*-values, *Q(TS|c)=Q(TS|c)+α(r- Q(TS|c))*. TS-values are used to select TSs, given a specific context. The selected TS determines which behavioral strategy, i.e., which set of Q-values, guide behavior. Neural evidence suggests that such a solution is indeed implemented in the brain’s PFC-BG learning loop.

Figure 1: 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 will provide behavioral evidence that humans use a combined TS-RL strategy. Second, we will develop a computational algorithm of TS-RL. And third, we will assess the neural mechanisms underlying TS-RL, using EEG in combination with behavioral analyses and computational modeling.

*Preliminary results*. To achieve the first component of our proposed research, we aimed to test whether humans apply RL at the level of TS selection by acquiring *Q(TS|c)*, and at the level of action selection by acquiring *Q(a|s,TS)*. A key prediction of our model is that participants’ choices should be sensitive to both TS-values and *Q*-values.

*Task design*. We designed a behavioral task in which human participants acquire three different TSs, which differ in overall value. After having acquired the TSs, participants enter a test phase in which they indicate which TSs and stimuli they prefer (reflecting learned TS-values and Q-values) and another test phase in which they select actions in a novel context (assessing TS generalization). The two tests assess whether participants acquire TS­-values and whether TS­-values guide generalization, i.e., TS selection in novel contexts.

*Results*. In accordance with our predictions for the first test phase, 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 in novel contexts. This shows that TS values 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 were obtained after controlling for low-level *Q*-values, and have been replicated in two independent datasets. Taken together, the behavioral analyses of human behavior in a hierarchical task support our hypothesis that humans select TSs based on TS-values.

***Proposed research***.

*Task design and data collection*. We will use the same task as in the behavioral experiment described above, adjusting only minor details such as timing parameters for the use with electro-encephalography (EEG). Participants will be recruited through UC Berkeley’s research participation pool (RPP). The task will take approximately 60 minutes to complete, and we are planning to enroll 40 participants.

*Computational modeling*. We will implement a TS-RL agent according to the description and formulas above. We will first verify that the TS-RL agent’s behavior replicates the patterns we observed in humans, as described above. We will then conduct model fitting procedures and identify which model fits the human behavior best.

*Planned analyses and expected results*. The EEG data will be analyzed in three ways, using classic event-related potentials (ERPs), model-based analyses based on computational modeling, and time-frequency analysis. With respect to the former, we expect specific ERP components in certain task conditions (e.g., P3 wave after unexpected outcomes). We also expect ERP differences in trials that differ in TS-values or Q-values, reflecting brain activity at both levels of RL. For the model-based analyses, we will assess whether certain electrodes or time points reflect the trial-by-trial changes in TS-values and Q-values that we will derive for each participant from the model. We expect that Q-values and TS-values affect brain activity during stimulus presentation, and that prediction errors affect brain activity during feedback. For the time-frequency analyses, we will assess whether brain oscillations differ between trial in the ways predicted by the TS-RL model.

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

*Decoding approach fails*. It is a possibility that we will not be able to decode Q-values and TS-values from the EEG signal, although previous research has reported these, for several reasons. If this should be the case, we will still be able to make valid conclusions based on ERP and time-frequency analyses.