Check for updates

A Unified Theoretical Framework for Cognitive Sequencing

Tejas Savalia[1], Anuj Shukla[1] and Raju S. Bapi[1,2]*

[1] Cognitive Science Lab, International Institute of Information Technology, Hyderabad, India, [2] School of Computer and Information Sciences, University of Hyderabad, Hyderabad, India

The capacity to sequence information is central to human performance. Sequencing ability forms the foundation stone for higher order cognition related to language and goal-directed planning. Information related to the order of items, their timing, chunking and hierarchical organization are important aspects in sequencing. Past research on sequencing has emphasized two distinct and independent dichotomies: implicit vs. explicit and goal-directed vs. habits. We propose a theoretical framework unifying these two streams. Our proposal relies on brain's ability to implicitly extract statistical regularities from the stream of stimuli and with attentional engagement organizing sequences explicitly and hierarchically. Similarly, sequences that need to be assembled purposively to accomplish a goal require engagement of attentional processes. With repetition, these goal-directed plans become habits with concomitant disengagement of attention. Thus, attention and awareness play a crucial role in the implicit-to-explicit transition as well as in how goal-directed plans become automatic habits. Cortico-subcortical loops basal ganglia-frontal cortex and hippocampus-frontal cortex loops mediate the transition process. We show how the computational principles of model-free and model-based learning paradigms, along with a pivotal role for attention and awareness, offer a unifying framework for these two dichotomies. Based on this framework, we make testable predictions related to the potential influence of response-to-stimulus interval (RSI) on developing awareness in implicit learning tasks.

Keywords: implicit sequence learning, explicit sequence knowledge, habitual and goal-directed behavior, model-free vs. model-based learning, hierarchical reinforcement learning

1. INTRODUCTION

Cognitive Sequencing can be viewed as the ability to perceive, represent and execute a set of actions that follow a particular order. This ability underlies vast areas of human activity including, statistical learning, artificial grammar learning, skill learning, planning, problem solving, speech and language. Many human behaviors ranging from walking to complex decision making in chess involve sequence processing (Clegg et al., 1998; Bapi et al., 2005). Such sequencing ability often involves processing repeating patterns—learning while perceiving the recurrent stimuli or actions and executing accordingly. Sequencing behavior has been studied in two contrasting paradigms: *goal-directed* and *habitual* or under the popular rubric of *response-outcome* (R-O) and *stimulus-response* (S-R) behavior. A similar dichotomy exists on the computational side under the alias of *model-based* vs. *model-free* mechanisms. The model-based vs. model-free computational paradigm has proved vital in designing algorithms for planning and learning in various intelligent

system architectures—leading to the proposal of their involvement in human behavior as well. In this article, we use another dichotomy on the learning side: *implicit* vs. *explicit* along with a pivotal role for attention and awareness to connect these dichotomies and suggest a unified theoretical framework targeted toward sequence acquisition and execution. In the following, the three dichotomies will be described along with a summary of the known neural bases of these.

1.1. Habitual vs. Goal-Directed Behavior

Existence of a combination of habitual and goal-directed behaviors is shown in empirical studies on rats and humans. In the experiments to study these behaviors two phenomena have been used to differentiate: *outcome devaluation*—sensitivity to devaluation of the goal and *contingency degradation*—sensitivity to an omission schedule. Outcome devaluation is achieved by satiating the rats on the rewarding goal; making the reward less appealing whereas contingency degradation is achieved by omitting a stimulus within a sequence of stimuli leading to the goal. Results demonstrate that overtrained rats and humans seem to be insensitive to both the phenomena. That is, even though the outcome of following a path is devalued or a stimulus in the sequence is omitted, habits lead rats to follow the same path, thus relating overtraining to habitual or stimulus-response (S-R) kind of control (Adams and Dickinson, 1981; Killcross and Coutureau, 2003). On the other hand, moderately trained rats have had little or no difficulty adapting to the new schedule relating this behavior to a goal-directed or response-outcome (R-O) kind of control (Dickinson, 1985; Balleine and Dickinson, 1998; Tricomi et al., 2004; Balleine and O'Doherty, 2010; Dolan and Dayan, 2013). Based on this proposal of two contrasting mechanisms, quite a few notable neuroimaging studies have attempted establishing the neural substrate related to the two modes of control. fMRI studies related to outcome devaluation point to two sub areas of the ventromedial prefrontal cortex (vmPFC)—medial orbito-frontal cortex (OFC) and medial prefrontal cortex (PFC) as well as one of the target areas of the vmPFC structures in the human striatum, namely, the anterior caudate nucleus, to be involved in goal-directed actions (Valentin et al., 2007; Balleine and O'Doherty, 2010). Studies aimed at finding the neural substrate for habitual behavior suggest an involvement of the subcortical structure, dorsolateral striatum (Hikosaka et al., 1999; Yin and Knowlton, 2006; Tricomi et al., 2009).

1.2. Model-Free vs. Model-Based Paradigm

In order to understand the learning process in goal-directed and habitual paradigms, two contrasting computational theories have been proposed. Goal-directed learning and control processes have been related to a *model-based* system (Doya et al., 2002; Khamassi and Humphries, 2012) whereas the habitual paradigm to a *model-free* system (Dolan and Dayan, 2013). Typically, a goal-directed system uses its view of the environment to evaluate its current state and possible future states, selecting an action that yields the highest reward. A model-based mechanism conceives this as building a search tree leading toward goal states. Such a system can be viewed as using the past experiences to understand the environment and using this view of the environment to predict the future.

In contrast, a habitual behavior can be viewed as a repetition of action sequences based on past experience. Acquisition of a habit can be viewed as learning a skill—correcting the residual error in the action sequence leading to the goal. Typically skills are acquired over time, are sensitive to the effectors used for acquisition (Bapi et al., 2000) and there seems to be a network of brain areas involved at various stages of skill learning (Hikosaka et al., 1999). A model-free system conceives the residual error as a sort of prediction error called the *temporal difference* (TD) error (Sutton, 1988). Dopamine has been noted to be a key neurotransmitter involved in encoding the prediction error and thus leading to the view that dopamine plays a crucial role in habit learning (Schultz et al., 1997). The influence of dopamine is in no way limited to habitual behaviors. Role of dopamine in functions mediated by the prefrontal cortex such as working memory, along with the observation that there are wide projections of dopamine to both the caudate and the putamen and studies manipulating dopamine levels in the prefrontal cortex affecting goal-directed behavior indicate its involvement in the goal directed mechanism as well (see Dolan and Dayan, 2013 for a review).

Further work has been directed at establishing a connection between the dichotomies—the behavioral dichotomy of goal-directed vs. habitual and the computational dichotomy of model-based vs. model-free. For example, Daw et al. (2005) suggested an uncertainty based competition between the two behaviors—the computationally simpler process at any point acting as a driver process. Another interesting aspect of such a combination comes from the *hierarchical reinforcement learning* (HRL) framework (Sutton and Barto, 1998; Sutton et al., 1999; Botvinick et al., 2009; Botvinick, 2012). An important aspect of acquisition of sequences is the formation of chunks among the sequence of stimuli. The striatum has been emphasized to be involved in the chunking process, the chunks then selected and scaled by the output circuits of the basal ganglia (Graybiel, 1998).

1.3. Implicit and Explicit learning

For the past three decades, there has been significant interest in sequence learning. A large body of experimental evidence suggests that there is an important distinction between *implicit* and *explicit learning* (see for example Howard and Howard, 1992; Shea et al., 2001 studies). Howard and Howard (1992) used a typical serial reaction time (SRT) task (Nissen and Bullemer, 1987), where the stimuli appeared in one of the four locations on the screen, and a key was associated with each location. The participants' job was to press the key as soon as possible. The stimuli were presented in a sequential order. With practice, participants exhibited response time benefits by responding faster for the sequence. However, their performance dropped to chance-level when they were asked to predict the next possible location of the stimulus suggesting that the participants might have learned the sequence of responses in an implicit manner and thus can not predict the next move explicitly. Another example is the study by Shea et al. (2001), where participants were given a standing platform task and were asked to mimic

the movements of the line presented on a screen. The order of the stimuli was designed in such a way that the middle segment was always fixed whereas the first and the last segments varied but participants were not told about the stimulus order. It was found that the performance of the middle segment improves over time. During the recognition phase participants fail to recognize the repeated segment, pointing to the possibility that they may have acquired this via implicit learning. A recent study done with a variant of SRT task called oculomotor serial response time (SORT) task also suggests that motor sequence can be learned implicitly in the saccadic system (Kinder et al., 2008) as well and does not pose attentional demands when the SORT task is performed under dual-task condition (Shukla, 2012). Another way of differentiating implicit vs. explicit learning would be to see whether an explicit instruction was given about the presence of a sequence prior to the task. An instruction specifying the presence of a sequence in the task would, in turn, drive attentional learning. Without such explicit prior knowledge, however, it may take more number of trials for the subjects to become aware of the presence of a sequence—requiring them to engage their attention toward the sequence, turning the concomitant learning and execution explicit. Apart from these studies, there is a large body of clinical literature which confirms the distinction of implicit and explicit learning. Most of the clinical evidence comes from the *artificial grammar learning* (AGL) paradigm, where patients learned to decide whether the string of letters followed grammatical rules or not. Healthy participants were found to learn to categorize grammatical and ungrammatical strings without being able to verbalize the grammatical rules. Evidence from amnesic patients points toward implicit learning being intact in patients even though their explicit learning was severely impaired (Knowlton et al., 1992; Knowlton and Squire, 1996; Gooding et al., 2000; Van Tilborg et al., 2011). Willingham et al. (2002) suggested that activation in the left prefrontal cortex was a prerequisite for such awareness along with activation of the anterior striatum (Jueptner et al., 1997a,b). Results of the positron emission tomography (PET) study of Grafton et al. (1995) when participants performed a motor sequence learning task under implicit or explicit learning task conditions suggest that the motor cortex and supplementary motor areas were activated for implicit learning whereas the right premotor cortex, the dorsolateral cingulate, anterior cingulate, parietal cortex and also the lateral temporal cortex were associated with explicit procedural memories (Gazzaniga, 2004; Destrebecqz et al., 2005).

It has been established that the brain areas involved in working memory and attentional processing are more active during explicit learning as compared to implicit learning. Further, the findings of functional magnetic resonance imaging (fMRI) studies suggest that the prefrontal and anterior cingulate cortex and early visual areas are involved in both implicit and explicit learning (Aizenstein et al., 2004). However, there is a greater prefrontal activation in case of explicit processing than implicit which is consistent with the findings from attention literature suggesting that prefrontal activation is associated with controlled and effortful processing (Aizenstein et al., 2004). However, the neural bases of implicit and explicit learning are still inconclusive. For example, Schendan et al. (2003) used fMRI

to differentiate brain activation involved in implicit and explicit processing. Their finding suggests that the same brain areas are activated in both types of processing. More specifically, the medial temporal lobe (MTL) is involved in both implicit and explicit learning when a higher order sequence was given to the participants. Furthermore, Pammi et al. (2012) observed a shift in fronto-parietal activation from anterior to posterior areas during complex sequence learning, indicating a shift in control of sequence reproduction with help of a chunking mechanism.

In this section we discussed the three dichotomies that have stayed mostly distinct in the literature. While there have been many significant attempts at combining goal-directed behavior with model-based mechanism and habitual behavior with model-Free mechanism, we attempt to add the third implicit vs. explicit dichotomy to devise a unifying framework explaining both learning and execution.

2. COMPUTATIONAL EQUIVALENTS

In this section we present how explicit learning and goal directed behavior can be related to a model-based mechanism whereas implicit learning and habitual behavior can be related to a model-free system. Indeed, there have been previous such attempts at bringing together the contrasting paradigms (Doya et al., 2002; Daw et al., 2005; Dezfouli and Balleine, 2012; Dolan and Dayan, 2013; Dezfouli et al., 2014; Cushman and Morris, 2015).

2.1. Goal Directed Behavior As Model-Based Mechanism

A goal directed behavior can be viewed as keeping the end-point (goal) in mind and selecting the ensuing actions accordingly. This kind of learning and control can be explained by a simple markov decision process (MDP) framework. Typically, an agent estimates its environment and calculates the value of its current state and possible future states. This estimation can be described by the Bellman equations:

$$V(s_n) = \sum T(s_n, a, s_{n+1})[R(s_n, a, s_{n+1}) + \gamma V(s_{n+1})] \quad (1)$$

Here, $T(s_n, a, s_{n+1})$ is the transition probability: the probability of the agent landing in state s_{n+1} if it takes an action a from state s_n. $R(s_n, a, s_{n+1})$ denotes the reward the agent gets on taking an action a from state s_n and lands in state s_{n+1}, γ the discount factor enabling the agent to select higher-reward-giving actions first. Such a system can be viewed as building a search tree and looking forward into the future, estimating values of the future states and selecting the maximum valued one (denoted as $V^*(s_n)$).

$$V^*(s_n) = \max_a \sum T(s_n, a, s_{n+1})[R(s_n, a, s_{n+1}) + \gamma V(s_{n+1})] \quad (2)$$

This kind of behavior requires the agent to know the transition probabilities along with the reward function. While these values are not directly available in the environment, the agent builds these gradually over time while interacting with the environment—learning the transition probabilities iteratively. These values, in effect, form a model of the environment, hence

deriving its name. Rewards in all the future states are estimated by propagating back the rewards from the final goal state. The transition probabilities can be initially equal for all actions and subsequently getting refined with experience.

2.2. Habitual Behaviors As Model-Free System

Habitual behavior can be viewed as a typical S-R behavior, where the end-goal does not influence the current action selection directly. Instead, previous experiences of being in a particular state are cached (Daw et al., 2005). This can be conveyed by a model-free system through the well established temporal difference (TD) learning. TD learning follows the following update rules.

$$p_k = R(s_n, a, s_{n+1}) + \gamma V(s_{n+1}); \qquad (3)$$
$$V(s_n) = (1 - \alpha)V(s_n) + (\alpha)p_k$$
$$V(s_n) = V(s_n) + \alpha(V(s_n) - p_k) \qquad (4)$$

Here α is the learning rate. p_k encodes a sample evaluation of state s_n when the agent enters state s_n for the kth time in the form of a sum of two terms—the first term indicating reward the agent would receive from the current state and the second term computing discounted value of the next state s_{n+1} that it would enter, the value being returned is the agent's version of the value function $V(\cdot)$. The last term of Equation (4) refers to the prediction error signal. The definition of terms such as $R(\cdot)$ and $V(\cdot)$ are as defined earlier in Bellman equations. This system can be viewed as looking into the past—making a small adjustment to optimize performance and taking the next action. There is no explicit model of the system, the agent learns on-line—learning while performing.

Our proposed architecture attempts a combination of the two contrasting paradigms. We suggest that implicit learning and control can be viewed in a similar way as habitual behavior and in turn both can be modeled using a model-free computational system. Similarly, explicit learning and control seem to have similar requirements as goal-directed behavior and in turn both can be understood as using a model-based computational system. We aim to exploit the hierarchical reinforcement learning architecture and chunking phenomenon to propose how these contrasting dichotomies can be combined into a unified framework in the next section.

3. UNIFIED THEORETICAL FRAMEWORK

In a model-based mechanism—searching in a tree of possible states—as one looks further ahead into the future, the search tree starts expanding exponentially, making such a search computationally infeasible. Whereas in case of a model-free mechanism, the system has to be in the exact same state as it was before to enable an update in its policy. To make such an update account for something substantial, there have to be enough samples of a particular state which might take a larger number of trials exploring the entire state space. The respective inefficiency of the individual systems (Keramati et al., 2011) and

the evidence of existence of both as part of a continuum allow us to formulate a hybrid scheme combining both the computational mechanisms to explain sequence acquisition and execution in the brain.

In an attempt to formulate a unifying computational theory, we add the learning factor—implicit learning conceived as model-free learning whereas explicit learning conceived as model-based. One such idea in computational theories that suits our needs is the hierarchical reinforcement learning (HRL) framework (Sutton and Barto, 1998; Sutton et al., 1999). The HRL framework gives an additional power to the agent: the ability to select an "option" along with a primitive action for the next step. An option is a set of sequential actions (a motor program) leading to a subgoal. The agent is allowed to have separate policies for different options—on selection of an option, the agent follows that option's policy; irrespective of what the "external" policy for the primitive actions is.

We propose that learning within an option—the policy of the primitive actions within an option occurs in a model-free way. The most granular set of actions a human performs are learned by a habitual mechanism and implicitly. As one moves to learning of a less granular set of actions the roles start to change—a habitual model-free learning gradually transforms to a goal-directed one. At some point, one becomes aware of the recurring patterns being experienced and the attentional processes thereafter enable a shift from implicit state to explicit learning. Indeed, in the serial reaction time studies, it has been observed that as the subjects became aware of the recurring pattern or sequence, their learning might have moved from implicit to explicit state. We attribute this conversion to explicit learning to the formation of *explicit motor programs* or *chunks*. Chunks are formed when the subject becomes aware of the sequential pattern and implicit, model-free learning then turns into an explicit and model-based learning process. One interesting theory that can be used to explain the chunking process is the *average reward Reinforcement Learning* model (Dezfouli and Balleine, 2012).

As depicted in **Figure 1**, a similar analogy can be applied in control or performance of sequences with a change in direction in the process described above. The most abstract, top level goals are executed explicitly in a goal-directed way using model-based mechanism, the goal directed mechanisms gradually relinquishing control as the type of actions proceed downward in the hierarchy. At the finest chunk-level, subject loses awareness of the most primitive actions executed and those are then executed entirely in a habitual, model-free, implicit manner.

In neural terms, the ventromedial prefrontal cortex (vmPFC) along with the caudate nuclei may be involved in the goal-directed part, the dorsolateral striatum and dorsolateral prefrontal cortex (DLPFC) may be engaged in the habitual part—dopamine providing the prediction error signal while the anterior regions of the striatum and left prefrontal and medial frontal cortex playing a role in attentional processes. Gershman and Niv (2010) suggested a role for the hippocampus in task structure estimation which could be extended to estimating the world model and hence the transition probabilities required for the model-based system. Neural correlates for the options framework are detailed in **Figure 1**.

FIGURE 1 | Actor critic architecture for learning and execution. Input from the environment passes to the goal-directed mechanism to select a chunk. Action selection at the upper-level is enabled by engagement of attention in a goal-directed, model-based manner whereas at a lower level (without attentional engagement) this process implements habitual and model-free system. Action selection within a chunk occurs on a habitual, model-free basis. Neural correlates of various components of this framework are suggested here. Based on Botvinick et al. (2009). vmPFC, Ventromedial PreFrontal Cortex; DLS, DorsoLateral Striatum; VS, Ventral Striatum; DA, Dopaminergic error signal; HT+, Hypothalamus; DLPFC, DorsoLateral PreFrontal Cortex; OFC, OrbitoFrontal Cortex; MTL, Medial Temporal Lobe.

In this section we presented our unifying framework for combining the three dichotomies. In the subsequent section we attempt to specify the roles of response-to-stimulus interval (RSI) [or more generally, inter-stimulus interval (ISI)] and prior information along with the pivotal role of attention in switching between the two contrasting mechanisms in the explicit vs. implicit and goal-directed vs. habitual dichotomies.

4. ROLE OF RESPONSE-TO-STIMULUS-INTERVAL (RSI) AND PRIOR INFORMATION IN THE UNIFIED FRAMEWORK

A model-based search leading to explicit learning is typically slower—subject is required to deliberate over possible choices leading to the goal. In contrast, subject does not need to think while performing an action habitually or learning implicitly—a model-free mechanism does not deliberate, it performs an action based on an already available "cache" of previous experiences and updates the cache as it proceeds further. Based on this, we propose that *response-to-stimulus interval* (RSI) [or more generally, inter-stimulus interval (ISI)] plays a key role in serial reaction time (SRT) experiments. Larger RSIs allow the subject enough time to form a model of the system, deliberate over the actions and hence this kind of learning and control corresponds to a model-based (explicit) system. On the other hand smaller RSIs do not allow the subject to form an explicit model and as is well known from the literature of serial reaction time experiments, subjects do remain sensitive to (implicitly acquire) the underlying sequential regularities (Robertson, 2007). This sort of implicit learning can be explained with temporal difference (TD) learning, where the error signal leads to an adjustment in action selection keeping the general habitual control the same.

Further, knowledge (prior information) about the existence of sequential regularities in the SRT task leads to the learning and control being explicit and model-based. This can be said to engage attentional processes in our proposal. With attentional engagement, habitual control ceases to exercise control over behavior. While we propose attention-mediated arbitration between model-based and model-free systems, Lee et al. (2014) suggest that such mediation is driven by estimates of reliability of prediction of the two models. Emerging awareness of the presence of a sequence plays a similar role in mediating learning as explicit attentional processes. The complete architecture is depicted in **Figure 2**.

Our proposal relies heavily on the *hierarchical chunking* mechanism and engagement or disengagement of attention to the underlying repeating pattern or sequence. While learning which begins implicitly and in a model-free manner, eventually as the formation of chunks proceeds up the hierarchy, at some point, the size of chunk—defined in terms of the time it takes to execute the set of actions within the chunk, crosses a threshold thus engaging the attentional resources of the subject. At this point explicit model-based learning starts taking control. Similarly, during control (or execution) of a sequence, the top most selection of chunks happens via a goal-directed, model-based

mechanism, on proceeding down the chunk hierarchy after the point of crossing some chunk size threshold, the subject no longer pays attention to the execution—it goes on in a habitual, model-free manner. Learning or execution of a set of actions within a chunk proceeds in a habitual, model-free fashion, – which at "attentive" level in the hierarchy can be explained by a habitual control of goal selection as suggested by Cushman and Morris (2015).

Attention engagement or disengagement occurs when the chunk size is equivalent to a certain *temporal window*. Such a temporal window includes the RSI for a typical SRT task. For instance, larger RSIs need fewer physical actions to reach the threshold size of the temporal window during bottom-up learning and hence cause attentional engagement toward the underlying sequential pattern sooner than in case of a trial with smaller RSIs. Based on this proposal, it will be interesting to empirically investigate the impact of varying the size of temporal window and studying resultant influence on the awareness levels of the presence of an underlying sequence in the standard SRT task. According to our proposal, implicit (associative) learning in the lower-level of the hierarchy proceeds without engagement of attention. Further we propose that as the response-stimulus interval (RSI) increases the width of the

FIGURE 2 | Role of temporal window in engagement/disengagement of attention during learning and execution. The left panel refers to sequence execution (performance) where the flow is from top-to-bottom, attention gets gradually disengaged as you go down the hierarchy. The right panel shows the acquisition (learning) of sequences where the flow is from bottom-to-top, attention gets gradually engaged as you go up the hierarchy. The temporal window determines when to switch between the two mechanisms. For example, for an action worth 1 unit of time with the temporal window size of 5 units with RSI of 3 units; a two-action chunk would lead to attention engagement/disengagement. Lesser RSI would require more number of actions chunked together to engage/disengage attention toward the underlying task.

temporal window available for integration of information related to the previous response and the subsequent stimulus increases. Thus, increasing the temporal window allows deliberative and reflective (analytical) processes to kick in, enabling a transition to explict (awareness-driven) top-down mechanisms. This prediction can be verified experimentally and seems to be supported by preliminary evidence from the work of Cleeremans et al. (see Cleeremans 2014).

Such a hierarchical chunking mechanism for behavior generation has been suggested by Albus (1991), albeit from intelligent control architecture perspective. According to Albus (1991), a hierarchy producing intelligent behaviors comprises seven levels covering at the lowest level, the finest reflex actions, and spanning all the way up to long term planning of goals. At each higher level, the goals expand in scope and planning horizons expand in space and time. In neural terms, the motor neuron, spinal motor centers and cerebellum, the red nucleus, substantia nigra and the primary motor cortex, basal ganglia and prefrontal cortex and finally the temporal, limbic and frontal cortical areas get involved in increasing order of the hierarchy.

5. COMPARISON WITH OTHER DUAL SYSTEM THEORIES

Many dual system theories related to goal-directed vs. habitual behavior or implicit vs. explicit learning have been proposed in the recent past. For example, Keele et al. (2003) suggest a dual system where implicit learning is typically limited to a single dimensional or a unimodal system whereas explicit learning involves inputs from other dimensions as well. Our model incorporates this duality in a different sense and does not distinguish the dichotomy between different modalities. Inputs from multiple modalities are treated as actions in an abstract sense and when a bunch of such actions crosses the threshold (acquisition or execution time), this would lead to attentional modulation (engagement in the case of acquisition or disengagement in the case of performance). A similar idea has been discussed by Cleeremans (2006) who suggested that a representation obtained from exposure to a sequence may become explicit when strength of activation reaches a critical level. Formation of the chunks is, however, assumed to be driven by bottom-up, unconscious processes. These chunks become available later for conscious processing (Perruchet and Pacton, 2006). We concur with the suggestions of Keele et al. (2003) on the neural correlates of implicit and explicit learning; learning in the dorsal system being implicit whereas that in the ventral system may be related to explicit or implicit modes. However, we emphasize that the ventral system—when learning is not characterized as a uni- or multi-dimensional dichotomy— would be more related to explicit learning. Daltrozzo and Conway (2014) discuss three levels of processing: an abstract level storage for higher level goals, followed by an intermediate level encoding of the actions required to reach the goal and a low level acquiring highly specific information related to the exact stimulus and associated final motor action (Clegg et al.,

1998). Our model reflects such a hierarchy by breaking down the actions into a finer set of sub-actions—where the top most abstract actions or goals are decided by a goal-directed, model-based system whereas the more concrete actions are executed by a habitual, model-free system. Walk and Conway (2011) suggest a cascading account where two mechanisms interact with each other in a hierarchical manner—concrete information being encoded in a modality specific format followed by encoding of more domain-general representations. We incorporate such an interleaving phenomenon by suggesting that the actions within a chunk are carried out in a habitual, attention-free manner; the selection of such a chunk being goal-directed and attention-mediated. Thiessen et al. (2013) discuss a dual system involving an extraction and integration framework for general statistical learning. The extraction framework is implicated in conditional statistical learning—formation of chunks or associations between events occurring together. On the other hand, the integration framework is implicated in distributional statistical learning—generalization of the task at hand. We can relate the extraction framework to the implicit, habitual process and the integration framework to a goal-directed mechanism that involves creation of the model of environment using information from potentially multiple sources. Batterink et al. (2015) present evidence suggesting that though there does not seem to be an explicit recognition of statistical regularities, the reaction time task, which is deemed 50% more sensitive to statistical learning, suggests that there is in fact some statistical structure of the presented stimuli learned implicitly. Our framework agrees with the conclusion that implicit and explicit statistical learning occur in parallel, attention deciding the driver process. A similar account has been suggested by Dale et al. (2012) who state that the system initially learns a readiness response to the task in an associative fashion mapping the stimulus to a response and then undergoes a rapid transition into a "predictive mode" as task regularities are decoded. Reber (2013) suggests a role for the hippocampal-medial temporal lobe (MTL) loop in explicit memory whereas implicit memory is said to be distributed in the cortical areas. However, evidence from studies with Parkinsons patients suggests an important role for the basal ganglia in acquiring such implicit knowledge. We posit a similar role for the basal ganglia and corticostriatal loops in implicit learning; the knowledge that follows this learning may be stored throughout the cortex while keeping the role of MTL and hippocampus intact.

6. CONCLUSION AND FUTURE WORK

Sequencing is a fundamental ability that underlies a host of human capacities, especially those related to higher cognitive functions. In this perspective, we suggest a theoretical framework for acquisition and control of hierarchical sequences. We bring together two hitherto unconnected streams of thought in this domain into one framework—the goal-directed and habitual axis on the one hand and the explicit and implicit sequencing paradigms on the other, with the help of model-based

and model-free computational paradigms. We suggest that attentional engagement and disengagement allow the switching between these dichotomies. While goal-directed and habitual behaviors are related to performance of sequences, explicit and implicit paradigms relate to learning and acquisition of sequences. The unified computational framework proposes how the bidirectional flow in this hierarchy implements these two dichotomies. We discuss the neural correlates in light of this synthesis.

One aspect of applicability of our proposed framework could be skill learning. It is well known that skill learning proceeds from initially being slow, attentive and error-prone to finally being fast, automatic and error-free (Fitts and Posner, 1967). Thus, it appears that sequential skill learning starts being explicit and proceeds to be implicit from the point of view of attentional demands. At first sight, this seems to be at odds with the proposed unified framework here where the hierarchy seems to have been set up to proceed from implicit to explicit learning. However, the phase-wise progression of skill learning is consistent with the framework as per the following discussion.

It is pointed out that different aspects of skill are learned in parallel in different systems—while improvements in reaction time are mediated by implicit system, increasing knowledge of the sequential regularities accrues in the explicit system (Hikosaka et al., 1999; Bapi et al., 2000). The proposed unitary network is consistent with these parallel processes, the implicit processes operating from bottom-up and the explicit system in a top-down fashion. Key factor is the engagement and disengagement of attentional system as demarcated in **Figure 1**. One might wonder how this approach can be applied to research in non-human animals, where explicit mechanisms are difficult to be realized. Historically, while SRT research identifying implicit vs. explicit learning systems are largely based on human experiments, that of goal-directed and habitual research is based on animal experiments. The proposed framework is equally applicable for human and non-human participants. What is proposed here is that the lower-level system operates based on associative processes that allow the system to learn implicitly, respond reactively and the computations at this level are compatible with a model-free framework. On the other hand, the upper-level system is based on predictive processes that allow the system to prepare anticipatory responses that sometime cause

errors. Error-evaluation while learning and error-monitoring during control are part of this system that learns using explicit processes, enables goal-directed control of actions and the computations at this level are compatible with a model-based framework. Level of attentional engagement distinguishes these two levels as shown in **Figure 2**. Of course, non-human animals can not give verbal reports of their knowledge. The explicit system in the case of pre-verbal infants and non-human animals needs to be understood in the lines of predictive systems that can elicit anticipatory, predictive responses and learn rules and transfer them to novel tasks (Marcus et al., 1999; Murphy et al., 2008).

Finally based on this theoretical proposal, we make predictions as to how implicit-to-explicit transition might happen in serial reaction time tasks when response-to-stimulus interval (RSI) is systematically manipulated. The mathematical formulation of such a unified mechanism is yet to be established, along with a formalization of the attentional window and its relation to RSI.

AUTHOR CONTRIBUTIONS

TS, AS, and RB conceptualized the framework. TS and AS did literature review and wrote the introduction and review sections. RB added the modeling linkages. TS, AS, and RB all contributed to preparing and finalizing the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Adams, C. D., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.* 33, 109–121. doi: 10.1080/14640748108400816

Aizenstein, H. J., Stenger, V. A., Cochran, J., Clark, K., Johnson, M., Nebes, R. D., et al. (2004). Regional brain activation during concurrent implicit and explicit sequence learning. *Cereb. Cortex* 14, 199–208. doi: 10.1093/cercor/bhg119

Albus, J. S. (1991). Outline for a theory of intelligence. *IEEE Trans. Syst. Man Cybern.* 21, 473–509. doi: 10.1109/21.97471

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/S0028-3908(98)00033-1

Balleine, B. W., and O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69. doi: 10.1038/npp.2009.131

Bapi, R. S., Doya, K., and Harner, A. M. (2000). Evidence for effector independent and dependent representations and their differential time course of acquisition during motor sequence learning. *Exp. Brain Res.* 132, 149–162. doi: 10.1007/s002219900332

Bapi, R. S., Pammi, V. C., Miyapuram, K. P., and Ahmed. (2005). Investigation of sequence processing: a cognitive and computational neuroscience perspective. *Curr. Sci.* 89, 1690–1698.

Batterink, L. J., Reber, P. J., Neville, H. J., and Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *J. Mem. Lang.* 83, 62–78. doi: 10.1016/j.jml.2015.04.004

Botvinick, M. M., Niv, Y., and Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition* 113, 262–280. doi: 10.1016/j.cognition.2008.08.011

Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Curr. Opin. Neurobiol.* 22, 956–962. doi: 10.1016/j.conb.2012.05.008

Cleeremans, A. (2006). "Conscious and unconscious cognition: a graded, dynamic, perspective," in *Progress in Psychological Science Around the World Vol. I Neural, Cognitive and Developmental Issues,* eds Q. Jing, M. R. Rosenzweig, G. d'Ydewalle, H. Zhang, H. C. Chen, and K. Zhang (Hove: Psychology Press), 401–418.

Cleeremans, A. (2014). Connecting conscious and unconscious processing. *Cogn. Sci.* 38, 1286–1315. doi: 10.1111/cogs.12149

Clegg, B. A., DiGirolamo, G. J., and Keele, S. W. (1998). Sequence learning. *Trends Cogn. Sci.* 2, 275–281. doi: 10.1016/S1364-6613(98)01202-9

Cushman, F., and Morris, A. (2015). Habitual control of goal selection in humans. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13817–13822. doi: 10.1073/pnas.1506367112

Dale, R., Duran, N. D., and Morehead, J. R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Adv. Cogn. Psychol.* 8, 196–209. doi: 10.5709/acp-0115-z

Daltrozzo, J., and Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Front. Hum. Neurosci.* 8:437. doi: 10.3389/fnhum.2014.00437

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Del Fiore, G., Aerts, J., et al. (2005). The neural correlates of implicit and explicit sequence learning: interacting networks revealed by the process dissociation procedure. *Learn. Mem.* 12, 480–490. doi: 10.1101/lm.95605

Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x

Dezfouli, A., Lingawi, N. W., and Balleine, B. W. (2014). Habits as action sequences: hierarchical action control and changes in outcome value. *Philos. Trans. R. Soc. B* 369:20130482. doi: 10.1098/rstb.2013.0482

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007

Doya, K., Samejima, K., Katagiri, K.-I., and Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Comput.* 14, 1347–1369. doi: 10.1162/089976602753712972

Fitts, P. M. and Posner, M. I. (1967). *Human Performance.* Brooks/Cole Publishing Company.

Gazzaniga, M. S. (2004). *The Cognitive Neurosciences.* New York, NY; Norton, OH: MIT Press.

Gershman, S. J., and Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* 20, 251–256. doi: 10.1016/j.conb.2010.02.008

Gooding, P., Mayes, A., and Van Eijk, R. (2000). A meta-analysis of indirect memory tests for novel material in organic amnesics. *Neuropsychologia* 38, 666–676. doi: 10.1016/S0028-3932(99)00119-0

Grafton, S. T., Hazeltine, E., and Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *J. Cogn. Neurosci.* 7, 497–510. doi: 10.1162/jocn.1995.7.4.497

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70, 119–136. doi: 10.1006/nlme.1998.3843

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., et al. (1999). Parallel neural networks for learning sequential procedures. *Trends Neurosci.* 22, 464–471. doi: 10.1016/S0166-2236(99)01439-3

Howard, D. V., and Howard, J. H. (1992). Adult age differences in the rate of learning serial patterns: evidence from direct and indirect tests. *Psychol. Aging* 7:232. doi: 10.1037/0882-7974.7.2.232

Jueptner, M., Frith, C., Brooks, D., Frackowiak, R., and Passingham, R. (1997a). Anatomy of motor learning. II. subcortical structures and learning by trial and error. *J. Neurophysiol.* 77, 1325–1337.

Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S., and Passingham, R. E. (1997b). Anatomy of motor learning. I. frontal cortex and attention to action. *J. Neurophysiol.* 77, 1313–1324.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., and Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychol. Rev.* 110:316. doi: 10.1037/0033-295X.110.2.316

Keramati, M., Dezfouli, A., and Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Comput. Biol.* 7:e1002055. doi: 10.1371/journal.pcbi.1002055

Khamassi, M., and Humphries, M. D. (2012). Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. *Front. Behav. Neurosci.* 6:79. doi: 10.3389/fnbeh.2012.00079

Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* 13, 400–408. doi: 10.1093/cercor/13.4.400

Kinder, A., Rolfs, M., and Kliegl, R. (2008). Sequence learning at optimal stimulus–response mapping: evidence from a serial reaction time task. *Q. J. Exp. Psychol.* 61, 203–209. doi: 10.1080/17470210701557555

Knowlton, B. J., Ramus, S. J., and Squire, L. R. (1992). Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances. *Psychol. Sci.* 3, 172–179. doi: 10.1111/j.1467-9280.1992.tb00021.x

Knowlton, B. J., and Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol.* 22:169. doi: 10.1037/0278-7393.22.1.169

Lee, S. W., Shimojo, S., and O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699. doi: 10.1016/j.neuron.2013.11.028

Marcus, G. F., Vijayan, S., Rao, S. B., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science* 283, 77–80. doi: 10.1126/science.283.5398.77

Murphy, R. A., Mondragón, E., and Murphy, V. A. (2008). Rule learning by rats. *Science* 319, 1849–1851. doi: 10.1126/science.1151564

Nissen, M. J., and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cogn. Psychol.* 19, 1–32. doi: 10.1016/0010-0285(87)90002-8

Pammi, V. C., Miyapuram, K. P., Ahmed, Samejima, K., Bapi, R. S., Doya, K., et al. (2012). Changing the structure of complex visuo-motor sequences selectively activates the fronto-parietal network. *Neuroimage* 59, 1180–1189. doi: 10.1016/j.neuroimage.2011.08.006

Perruchet, P., and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* 10, 233–238. doi: 10.1016/j.tics.2006.03.006

Reber, P. J. (2013). The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia* 51, 2026–2042. doi: 10.1016/j.neuropsychologia.2013.06.019

Robertson, E. M. (2007). The serial reaction time task: implicit motor skill learning? *J. Neurosci.* 27, 10073–10075. doi: 10.1523/JNEUROSCI.2747-07.2007

Schendan, H. E., Searl, M. M., Melrose, R. J., and Stern, C. E. (2003). An fmri study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron* 37, 1013–1025. doi: 10.1016/S0896-6273(03)00123-5

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593

Shea, C. H., Wulf, G., Whitacre, C. A., and Park, J.-H. (2001). Surfing the implicit wave. *Q. J. Exp. Psychol. A* 54, 841–862. doi: 10.1080/713755993

Shukla, A. (2012). *Investigation of the Influence of Attention on Oculomotor Implicit Learning.* Unpublished M. Phil. Dissertation. Center for Neural and Cognitive Sciences, University of Hyderabad, Hyderabad.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Mach. Learn.* 3, 9–44. doi: 10.1007/BF00115009

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction,* Vol. 1. Cambridge: MIT Press.

Sutton, R. S., Precup, D., and Singh, S. (1999). Between mdps and semi-mdps: a framework for temporal abstraction in reinforcement learning. *Artif. Intell.* 112, 181–211. doi: 10.1016/S0004-3702(99)00052-1

Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (2013). The extraction and integration framework: a two-process account of statistical learning. *Psychol. Bull.* 139:792. doi: 10.1037/a0030801

Tricomi, E., Balleine, B. W., and ODoherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x

Tricomi, E. M., Delgado, M. R., and Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron* 41, 281–292. doi: 10.1016/S0896-6273(03)00848-1

Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/JNEUROSCI.0564-07.2007

Van Tilborg, I. A., Kessels, R. P., Kruijt, P., Wester, A. J., and Hulstijn, W. (2011). Spatial and nonspatial implicit motor learning in korsakoffs amnesia: evidence for selective deficits. *Exp. Brain Res.* 214, 427–435. doi: 10.1007/s00221-011-2841-6

Walk, A. M., and Conway, C. M. (2011). "Multisensory statistical learning: can associations between perceptual categories be acquired," in *Paper Session Presented at the Meeting of Cognitive Science Society* (Boston, MA).

Willingham, D. B., Salidis, J., and Gabrieli, J. D. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *J. Neurophysiol.* 88, 1451–1460. doi: 10.1152/jn.00461.2001

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.