Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 246 (2024) 1466–1478

Procedia Computer Science

www.elsevier.com/locate/procedia

28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024)

Impact of Pavlovian Approach Bias on Bidirectional Planning in Spatial Navigation Tasks

Reza Kakooee[a], Mohammad TH Beheshti[a,*], Mehdi Keramati[b]

[a]*Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran*
[b]*Department of Psychology, City University of London, London, United Kingdom*

Abstract

Bidirectional planning refers to a form of goal-directed decision-making process that combines forward and backward planning. Forward planning expands decision trees from the current state towards simulated futures, while backward planning starts the tree expansion from specific goal points in the opposite direction. Previous research has highlighted the impact of Pavlovian approach bias on behavior, showing that animals move towards appetitive outcomes regardless of the appropriateness of such behavior for achieving those outcomes. However, it remains unexplored whether the Pavlovian approach influences behavior by biasing backward planning. This research introduces a spatial navigation task to investigate the involvement of backward planning in humans' action-selection process and to determine whether the Pavlovian approach biases behavior through backward planning. The results reveal the behavioral signature of backward planning in humans and show that Pavlovian approach bias can influence both forward and backward planning, leading to decisions that are not necessarily instrumentally more efficient. Additionally, we developed a bidirectional planning algorithm based on reinforcement learning to simulate the participants' decisions. The simulation results suggest that the observed behavioral patterns can be parsimoniously explained by assuming that the Pavlovian approach bias acts as a pruning mechanism when expanding decision trees in both forward and backward directions.

Keywords: Reinforcement Learning; Computational Modeling; Bidirectional Planning; Decision-Making; Pavlovian Bias

1. Introduction

Goal-directed behavior in humans is fundamentally underpinned by decision-making processes and reinforcement learning mechanisms [20]. In the realm of decision-making, individuals are confronted with a myriad of choices and are tasked with selecting the option that maximizes expected utility. Although the outcomes of these choices

* Corresponding author.
E-mail address: mbehesht@modares.ac.ir

1877-0509
10.1016/j.procs.2024.09.593

can manifest as either rewards or penalties, reinforcement learning algorithms serve to update these utility estimates, thereby optimizing subsequent decision-making efforts.

One primary strategy employed for this purpose is forward planning, characterized by the extension of decision trees from an initial state to multiple potential future states [25]. This form of planning involves a chronological ordering of steps, effectively serving as a temporal framework for strategic planning [2]. However, this unidirectional approach may not encapsulate the full spectrum of human planning strategies. An alternative methodology, termed backward planning, has been examined in artificial intelligence literature [23]. This approach initiates the decision-making sequence from the goal state and works backward to identify the sequence of actions required to reach the intended outcome [29]. Furthermore, certain empirically observed behavioral phenomena in both animal models and human subjects, such as outcome-specific Pavlovian-to-instrumental transfer and differential outcome effects, may be accounted for by the incorporation of backward planning [1].

Computational demands inherent in the accurate evaluation of action sequences necessitate significant cognitive resources, which are inherently limited in biological agents. Consequently, exhaustive evaluation of decision trees—be it through forward-only or backward-only planning—is computationally infeasible. This limitation necessitates a control mechanism to regulate the depth and directionality of tree expansions for optimized planning outcomes. Prior research has indicated that artificial agents employing bidirectional planning mitigate this limitation by pruning the depths of both forward and backward decision trees. This balanced allocation of computational resources across both planning directions has been shown to enhance planning efficiency, as compared to solely employing deep forward or backward planning strategies [1].

Human decision-making is a complex process governed by multiple learning mechanisms. One such mechanism, instrumental planning, plays a pivotal role in shaping goal-directed behavior [22]. This form of learning involves establishing associations between potential actions and their consequent outcomes to identify the most advantageous sequence of actions. In contrast, Pavlovian learning represents a more instinctual form of behavioral guidance. It influences instrumental planning but operates independently of the outcome. These Pavlovian influences manifest as evolutionarily hardwired responses, typically characterized by an approach toward rewarding stimuli or an aversion to actions that predict negative consequences [28, 3, 13, 10].

While the Pavlovian valuation system can confer adaptive advantages in scenarios requiring rapid action or inhibition—such as abruptly stopping to avoid an oncoming vehicle—these heuristic responses generally yield quicker decisions than the deliberative instrumental processes, which necessitate data accumulation through environmental interaction [17]. However, the Pavlovian approach bias is not without its pitfalls. It can precipitate maladaptive decisions with detrimental long-term consequences, such as relapsing into smoking or drug use after prolonged abstinence [12], exhibiting alcohol-seeking behavior in the presence of associated cues [18, 27], or indulging in unhealthy foods despite dietary restrictions [4]. Thus, an empirical investigation into the interplay between Pavlovian and instrumental mechanisms across varied contexts holds the promise of enriching our understanding of human behavior.

While existing studies have elucidated the influence of the Pavlovian approach bias on forward planning [14], the impact of this bias on backward planning remains largely unexplored. Understanding the role of Pavlovian processes in decision-making is pivotal for dissecting how individuals engage in sequential choices during spatial navigation tasks [11]. Such insights also have the potential to augment our comprehension of action control mechanisms. Therefore, the primary objective of this research is to scrutinize the interplay between goal-directed planning and Pavlovian biases within the context of navigation tasks, with a particular focus on ascertaining how the Pavlovian approach bias modulates decision-making in backward planning scenarios.

This paper pursues the following hypotheses: (1) human decision-making in navigation tasks incorporates a forward planning strategy that is subject to modulation by the Pavlovian approach tendency; (2) a backward planning strategy is also employed by humans and is similarly influenced by the Pavlovian approach tendency; and (3) although both planning strategies are subject to Pavlovian biases, the impact is more pronounced in forward planning. To empirically test these hypotheses, we have devised a spatial navigation task incorporating Pavlovian approach cues. The task utilizes maps from simplified car racing games, into which Pavlovian approach cues have been integrated.

2. Related Works

Numerous studies have explored the multifaceted learning mechanisms in the brain that govern decision-making, categorizing them into Pavlovian and instrumental systems. Pavlovian learning, often studied through conditioning, occurs when an unconditioned stimulus becomes associated with a conditioned stimulus. Pavlovian learning can also refer to Pavlovian biases that are innate, pre-encoded responses in the brain that have been shaped through evolution. These biases can influence behavior automatically. Instrumental learning, consisting of habitual and goal-directed controllers, involves learning from action consequences to maximize rewards and minimize punishments. While these systems can operate independently, they also interact intricately, with the Pavlovian system exerting a directional influence on instrumental decision-making, regardless of whether this impact is positive or negative.

[6] investigates the interaction between habitual and goal-directed systems, proposing they compete based on uncertainty, with the brain arbitrating between them according to their situational accuracy. The study suggests a Bayesian approach to understand how the brain balances these systems to maximize rewards while minimizing computational costs. [8] explore how Pavlovian responses can disrupt instrumental actions aimed at obtaining rewards, underscoring the complexity of behavioral control and the need for nuanced models accounting for these interactions.

Instrumental learning often involves forward planning, expanding decision trees to determine optimal action sequences. [14] examine how Pavlovian processes influence instrumental behavior, showing Pavlovian responses shape actions by promoting or inhibiting behaviors based on predicted rewards or punishments. This highlights Pavlovian modulation of forward planning. [15] demonstrate how Pavlovian systems prune decision trees, simplifying decision-making, even when counterproductive, typically in response to losses. This illustrates a fundamental interaction between Pavlovian and goal-directed systems affecting decision-making efficiency.

[1] studied bidirectional planning, a model where forward and backward planning operate simultaneously to optimize decision-making. This AI-based approach enhances planning efficiency by expanding decision trees from current and goal states. [24] present a model optimizing decision-tree expansion, balancing speed and accuracy. The model predicts behavioral outcomes like time-pressure impacts on planning depth and reward effects on planning direction. [16] investigate hippocampal reactivations in reward-based learning, presenting a bidirectional search model incorporating forward and backward trajectory sampling to explain reactivation patterns. The model shows how reactivations contribute to updating and stabilizing state-action values for adaptive behavior and learning.

In summary, Pavlovian mechanisms encourage reward-aligned actions while discouraging those with anticipated negative consequences [21]. Pavlovian biases significantly influence forward planning [14], often leading to suboptimal decisions. However, their impact on backward planning is largely unexplored. We aimed to fill this gap by examining Pavlovian biases in bidirectional planning, crucial for understanding behavioral phenomena like Pavlovian-to-instrumental transfer and differential outcome effects [1].

3. Materials and Methods

3.1. Experiment

To examine the influence of Pavlovian approach cues on human decision-making in both forward and backward planning, we designed a spatial navigation task utilizing various maps from simplified car racing games. Approach cues were strategically embedded near the starting and goal positions on these maps to assess their impact on both forms of planning. Figure 1 showcases sample maps, categorized into five distinct classes: symmetric, nonsymmetric, forward, backward, and bidirectional maps. It should be noted that the varying path colors and yellow arrows are included solely for visualization purposes; in the actual experiment, all paths were uniform in color, and no arrows were present. For the sake of simplification in this paper, paths featuring Pavlovian approach cues are referred to as *green* paths, while those devoid of such cues are termed *red* paths. Furthermore, in each map, the goal point is situated at the northern end, the starting point at the southern end, and all paths are of identical length across all map classes, with the exception of the nonsymmetric maps.

Figure 1A-B present examples of forward and backward maps, respectively, with approach cues strategically positioned near the start and goal points. In the forward map, as indicated by the arrows, expansion of the decision tree from the starting point towards the goal state suggests that the green path brings the decision-maker geographically

closer to the goal. Conversely, the red path appears to increase the geographical distance from the goal. This configuration may elicit a sense of *approaching* or *withdrawing* for the decision-maker, depending on the chosen path. In contrast, in the backward maps, the decision-maker, when thinking from the goal state towards the starting point, may find that the red path increases the geographical distance between the start and goal points. This could potentially bias the decision-maker towards selecting the green path, which conveys a sense of approaching.

Additionally, we constructed maps featuring approach cues positioned proximally to both the start and end goal points, as illustrated in Figure 1C. These maps serve a critical function in elucidating the dynamics of bidirectional planning. A fundamental objective of this experiment is to assess whether the Pavlovian influences exert a more pronounced impact on decision-making in the context of bidirectional planning compared to unidirectional forward or backward planning.

Since the two available paths are identical in length in forward, backward, and bidirectional maps, there should be no inherent goal-directed preference favoring one path over the other based solely on distance. The distinguishing factor lies in the positioning of the Pavlovian approach cues. The primary objective of this investigation is to determine whether the presence of these cues can influence participants' goal-directed decision-making.

To evaluate participants' choices in the absence of Pavlovian cues, we created a set of *neutral* maps, devoid of such cues. These neutral maps are divided into two categories: symmetric and nonsymmetric maps, as illustrated in Figure 1D and Figure 1E, respectively. In the symmetric maps, both available paths are identical in length, offering no inherent directional preference. Conversely, in the nonsymmetric maps, one path is noticeably longer than the other, thereby introducing a variable that could influence goal-directed decision-making.

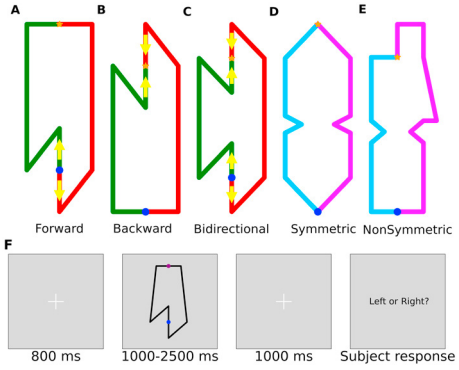

Fig. 1: The designed maps and the experiment. **(A, B, C)**: Three samples maps in which Pavlovian approach cues are embedded close to the start point (blue point) and or the goal point (gold star). **(D, E)**: Neutral maps with no Pavlovian cues. **(F)**: The experiment architecture

3.2. Tasks Description

To address our three primary hypotheses, we implemented an experiment as depicted in Figure 1F. A total of 30 participants were recruited for online participation. Participants were instructed to evaluate two paths on each map and identify which one would lead them to the goal state more quickly. Responses were registered via the right or left keys on the participants' keyboards. Beyond the principal hypotheses, this study also explores the influence of three secondary factors: map appearance time, map length, and the intensity of the Pavlovian approach cues. The experiment was partitioned into three distinct tasks, each undertaken by a separate subset of 10 participants. Each task assessed the influence of Pavlovian approach bias across all three planning paradigm. Additionally, within each task, the effect of one of the aforementioned secondary factors was examined. To provide a baseline, 24 neutral maps devoid of either forward or backward cues were also included in the experimental design. The study's procedures received ethical approval from the Ethical Committee of Tarbiat Modares University.

Task 1: To probe the influence of map appearance time on participants' decisions, two categories of maps were devised: *slow* and *fast* maps, with the appearance time of 2.5 seconds and 1 second, respectively. Within each category, participants were exposed to 8 purely forward, 8 purely backward, and 8 mixed forward-backward maps, along with their mirrored versions for counterbalancing purposes. This yielded a total of 96 maps ($2 * 8 * 3 * 2$), in addition to 24

neutral maps. **Task 2**: The aim of this task was to examine the influence of path length on the Pavlovian approach bias across different planning strategies. Accordingly, two categories of maps were introduced: *short* and *long* maps. These were created by equally altering the length of the left vertical street and its corresponding right street in each map. The number of maps used in this task was identical to that in Task 1. **Task 3**: This task sought to assess whether the length of the approaching or withdrawing segments within the maps affected participants' path selection. To this end, two categories of maps were generated: *compact* and *sharp*. These categories were created by modifying the length of the streets under the yellow arrows, as illustrated in Figure 1A-C. The number of maps was consistent with Tasks 1 and 2. In summary, the study utilized 6 distinct categories of maps: Fast, Slow, Short, Long, Compact, and Sharp. Each category comprised 5 different classes of maps: forward, backward, bidirectional, symmetric, and nonsymmetric.

4. Results

4.1. Statistical Analysis

Figure 2 summarizes the data collected from the three experimental tasks. Probabilities of path selection in symmetric and nonsymmetric maps across all subjects and tasks are depicted in Figure 2A and Figure 2B, respectively. In symmetric maps, the probability of path selection aligns with chance levels ($p = 0.17172$), as participants were instructed to make random selections if they perceived no difference between paths. In nonsymmetric maps, the probability of selecting the shorter path is significantly higher than choosing the longer path ($p < 0.001$), suggesting that participants adhere to goal-directed strategies, opting for shorter paths in the absence of Pavlovian cues.

Figure 2C presents the probabilities of choosing the green paths (those with Pavlovian approach cues) for each type of planning across all tasks. The data reveals that the likelihood of choosing green paths exceeds 0.5 in all classes (forward: $p < 0.001$, backward: $p = 0.006$, and bidirectional: $p < 0.001$), suggesting the presence of an additional influencing factor beyond goal-directed mechanisms. Given that no other elements in the maps could affect choice, it can be concluded that Pavlovian approach cues biased subjects' goal-directed decisions, leading to a higher frequency of green path selection. Figure 2C also shows the likelihood of choosing green paths is higher in the forward class than in the backward class ($p = 0.033$), indicating a stronger influence of Pavlovian approach bias in forward planning. Similarly, the probability of opting for the green path in bidirectional planning exceeds that in forward planning ($p = 0.002$).

Figure 2D-F present the behavioral analyses of subjects in Tasks 1-3, respectively. In Task 1 (Figure 2D), the probability of choosing green paths over red ones is significantly higher across all planning classes ($p < 0.001$), and this probability is further elevated in the slow category compared to the fast category ($p = 0.021$), suggesting that the influence of Pavlovian approach bias on goal-directed decisions is more pronounced when subjects have more time for planning. Figure 2E, the probabilities of selecting green paths exceed 0.5 in all cases ($p < 0.001$), indicating that Pavlovian approach bias influences goal-directed decision-making by skewing subject choices toward green paths. Moreover, the probabilities of opting for green paths in all three planning classes are higher in sharp maps compared to compact maps ($p = 0.019$), suggesting that more intense approach bias leads to greater influence on subject decisions. In Task 3 (Figure 2F), which examines the influence of map length on the interplay between Pavlovian approach and goal-directed decision mechanisms, the probabilities of selecting green paths consistently exceed 0.5 across all planning classes ($p < 0.001$). Additionally, the likelihood of choosing green paths is higher in short maps compared to long maps for all three classes ($p = 0.041$), suggesting that the impact of Pavlovian approach cues on decision-making is more pronounced when maps are shorter.

In summary, the statistical analysis confirms our initial hypotheses. The data indicates that subjects employ forward, backward, and bidirectional planning strategies in our navigation task, and each form of planning is subject to influence from the Pavlovian approach bias, with subjects demonstrating a preference for paths with approach cues. The extent of Pavlovian influence varies, being more potent in forward planning than in backward planning, and even stronger in bidirectional planning.

4.2. Simulation

Reinforcement Learning (RL) serves as a computational framework for the theory of adaptive optimal control, which aims to identify the most effective sequence of actions to achieve a specific goal [26]. Within the RL paradigm, there are two primary learning approaches for understanding reward mechanisms: Model-Based Reinforcement Learning (MBRL) and Model-Free Reinforcement Learning (MFRL). These approaches are particularly relevant for exploring goal-directed and habitual learning, respectively [6, 9].

This study concentrates exclusively on MBRL as the decision-makers have access to a complete map of the environment. In MBRL might also assume that the agent encodes the task's structure by storing associations between the current state, chosen action, subsequent state, and outcomes. This information is then utilized to form a cognitive map, which serves as a tool for prospective planning [7]. However, navigation tasks present unique challenges for achieving consistent stimulus control. Spatial behavior may not solely depend on goal-directed mechanisms but could also be influenced by the Pavlovian approach or specialized geometric cognition systems [9].

In the context of MBRL, an agent leverages its understanding of environmental dynamics and engages in prospective mental simulations to extend its decision tree from the current state into a simulated future. Conversely, during backward planning, the agent backpropagates the anticipated rewards from the goal state toward the initial state [1]. Under typical conditions—absent of Pavlovian biases or other pruning strategies, one would expect the agent to uniformly expand the decision tree across all available branches. However, as observed in the previous subsection, the presence of Pavlovian approach bias disrupts this symmetry. To account for this in our simulation, we propose that Pavlovian tendencies prompt the agent to unevenly extend its decision tree. Specifically, branches containing Pavlovian approach cues are expanded more deeply compared to those without such cues. This notion aligns with previous research, suggesting that agents, under the influence of Pavlovian biases, develop trees with varying depths across available branches during forward planning [24]. Similarly, we propose that in backward planning, the decision tree will not be uniformly expanded across all branches. Instead, the tree's depths will vary depending on the presence or absence of Pavlovian approach cues. Specifically, we hypothesize that the Pavlovian tendency will exert a selective influence during the reward backpropagation process. In branches containing Pavlovian approach cues, the agent is hypothesized to backpropagate the rewards more deeply compared to branches devoid of such cues.

Accurate path evaluation often demands deep planning, a process that can be computationally expensive given the constraints of time and cognitive resources. However, the planning horizon can be effectively extended with minimal cognitive expenditure through the use of bidirectional planning. Bidirectional planning offers a resource-efficient approach by allowing forward and backward planning trees to intersect at a point somewhere between the start and goal states. This reduces the necessity for each tree to be expanded to its full depth, thereby conserving cognitive resources. The overlapping of forward and backward planning trees also permits the reciprocal sharing of evaluations. That is, insights gained through backward planning can be utilized to inform and enhance the forward estimation of action values [1]. In essence, bidirectional planning serves as a mechanism that enables more efficient decision-making by facilitating the exchange of information between forward and backward planning processes.

5. Algorithm Development

RL theory has frequently been employed in neuroscience to elucidate forward planning [19, 5, 25]. To elucidate the observed behavior from our experimental findings, we employ a simulation based on the framework of a Markov Decision Process (MDP) in order to develop our RL algorithm. An MDP is characterized by the tuple (S, A, T, R, γ); where S is a finite set of states the agent can occupy, A signifies the set of actions available to the agent in each state, $T(s, a, s')$ represents the transition probability, describing the likelihood of moving from state s to s' when action a is taken in state s. $R(s, a, s')$ is the reward function, which quantifies the reward obtained by the agent for executing action a in state s and transitioning to state s'. γ is the discount factor, which balances the weight between immediate and future rewards. The agent aims to discover an optimal policy π, which maximizes its long-term expected reward.

This objective is encapsulated by the state-action value function, represented as $Q^{\pi}(s, a)$. This function calculates the expected sum of future rewards, starting from a given state s and taking action a, as defined in Equation 1.

This function can be determined using its respective Bellman equation as defined in Equation 2. Where $V^{\pi}(s') = \max_{a'} Q(s', a')$. The policy at any given state selects the action that maximizes the value of $Q^{\pi}(s, a)$.

$$Q^{\pi}(s,a) = E_{\pi}[\sum_{k=0}^{\infty} \gamma^{k} R_{t+k} | S_t = s, A_t = a] \tag{1}$$

$$Q^{\pi}(s,a) = R(s,a) + \sum_{s'} T(s'|s,a)\gamma V^{\pi}(s') \tag{2}$$

As previously mentioned, our model assumes that the agent has full knowledge of the maps. Therefore, MBRL serves as a suitable framework for the decision-making process. Value iteration is a specific MBRL algorithm that estimates the state value function, or alternatively, uses Equation 1 to directly compute the state-action value function [26]. Conventional MBRL algorithms typically use one-step prediction. To accommodate forward, backward, and bidirectional planning, we will need to adapt this standard update rule.

In the context of forward planning, we posit that the agent expands its decision tree into the future up to a finite number of steps, reflecting the agent's computational limitations or cognitive capacity for tree expansion. Moreover, the depth to which the decision tree is expanded can vary across different branches or paths. To illustrate this, consider Figure 3A, which depicts a simple map with two paths leading from a start state to a goal state. Each circle represents a discrete state. At each state, the agent has two potential actions: move one step forward or move one step backward. The reward associated with each action is zero, except for actions that lead to the goal state, which carry a specific positive reward. In our model, we suggest that the subject expands the decision tree along the green path to a depth denoted by d_g^f (superscript f and subscript g refer to forward planning and green path, respectively) and along the red path to a depth of d_r^f (subscript r refers to red path), such that the sum of both $d_g^f + d_r^f$ does not exceed the agent's capacity for forward planning. The greater the depth to which the tree is expanded in either direction, the higher the likelihood that the forward planning will either reach the goal state or intersect with the evaluations from backward planning somewhere in the middle of the map.

To incorporate the influence of Pavlovian approach bias into our model, we posit that the agent is more likely to expand the decision tree more deeply along the path that contains Pavlovian approach cues compared to the alternative path. In other words, the probability of extending the decision tree along the green path is higher than that for the red path. To account for the effects of the Pavlovian approach bias, we hypothesize that this difference in the expansion depths is modulated by this bias. Under these assumptions, we can modify Equation 2 for both the green and red paths as follows in which the subscript s in s_s denotes the *start* point, and the subscript g and r denote green and red, respectively.

$$Q^{f}(s_s, a_g) = r_0 + \gamma r_1 + ... + \gamma^{d_g^f - 1} r_{d_g^f - 1} + \gamma^{d_g^f} V^{\pi}(s_{d_g^f}) \tag{3}$$

$$Q^{f}(s_s, a_r) = r_0 + \gamma r_1 + ... + \gamma^{d_r^f - 1} r_{d_r^f - 1} + \gamma^{d_r^f} V^{\pi}(s_{d_r^f}) \tag{4}$$

In the case of backward planning, the agent also has computational constraints, limiting the depth to which it can backpropagate the goal state's rewards towards the starting state. Similar to forward planning, we assume that these backpropagation depths can vary between paths. For instance, in the map depicted in Figure 3A, the agent might backpropagate to a depth of d_g^b in the green path and d_r^b in the red path (superscript b refers to *backward* planning). Here, the constraint is that the sum $d_g^b + d_r^b$ should not exceed the agent's total capacity for backpropagation.

Deeper backpropagation increases the likelihood that the agent's backward planning will either reach the start state or intersect with the forward planning tree. In the context of Pavlovian approach bias, this tendency influences the agent's backward planning by promoting deeper backpropagation of rewards in the path with the approach cue, compared to paths without it. In other words, the agent is more likely to backpropagate rewards in the green path than in the red path. Similar to forward planning, where future rewards are discounted, rewards are also discounted in backward planning. The updated value functions for backward planning can then be represented by the following equations, where m and n represent states that are m and n steps away from the goal state along the green and red paths, respectively; and the term r_{goal} stands for the reward associated with reaching the goal state.

$$V_g^b(s_m) = \gamma^m r_{goal}\ ,\ m = 1, 2, ..., d_g^b \tag{5}$$

$$V_r^b(s_n) = \gamma^n r_{goal}\ ,\ n = 1, 2, ..., d_r^b \tag{6}$$

These equations take into account the Pavlovian bias by allowing for different depths of backpropagation in the green and red paths. This enables the model to capture the observed behavior wherein agents are more likely to choose the green path due to the influence of Pavlovian approach cues.

In bidirectional planning, we propose an approach where the agent expands its forward decision trees subject to the constraint $d_g^f + d_r^f \leq c^f$, and backpropagates its reward in the backward planning subject to $d_g^f + d_r^b \leq c^b$. Here, c^f and c^b are the respective capacities for forward and backward tree expansion. Additionally, the total capacity constraint is given by $c^f + c^b \leq c^{total}$. In scenarios where the forward and backward trees meet at some point in the middle of the map, the information gathered via backward planning is used to inform the forward planning process. Specifically, the agent expands the forward tree to a depth d_g^{fm}, which represents the depth of the branch in the green path from the start point to the meeting point. Similarly, the agent backpropagates the reward to a depth d_g^{bm}, representing the depth of the branch in the green path from the goal to the meeting point (superscript m refers to meeting point). The updating rule for state-value function can be defined as Equation 7, where $Q^{bid}(s_s, a_g)$ represents the value of taking action a_g (the action that leads to the green path) at the starting state s_s, when the agent employs a bidirectional planning strategy. Here, V_{max} can be computed as Equation 8.

$$Q^{bid}(s_s, a_g) = r_0 + \gamma r_1 + ... + \gamma^{d_g^{fm}-1} r_{d_g^{fm}-1} + V_{max} \tag{7}$$

$$V_{max} = max[V^b(s_{d_g^{bm}}),\ V^f(s_{d_g^{fm}})] \tag{8}$$

Here, $V^b(s_{d_g^{bm}})$ and $V^f(s_{d_g^{fm}})$ represent the value functions for the states that are d_g^{bm} and d_g^{fm} steps away from the goal state along the green path in backward and forward planning, respectively. $V^b(s_{d_g^{bm}})$ can be formulated as follows, where r_{goal} is the reward associated with reaching the goal state, and $\gamma^{d_g^{bm}}$ is the discount factor raised to the power of the depth of backpropagation in the green path.

$$V^b(s_{d_g^{bm}}) = \gamma^{d_g^{bm}} r_{goal} \tag{9}$$

Equation 7 illustrates a key aspect of bidirectional planning. In this framework, rather than extending the forward decision tree to a fixed depth d_g^f, the agent expands the tree only up to a depth $d_g^{fm} < d_g^f$. At this point, the agent can do one of two things: either bootstrap the value of the state $s_{d_g^{fm}}$ using its own forward-estimated value or leverage the value estimated through backward planning for the state $s_{d_g^{bm}}$.

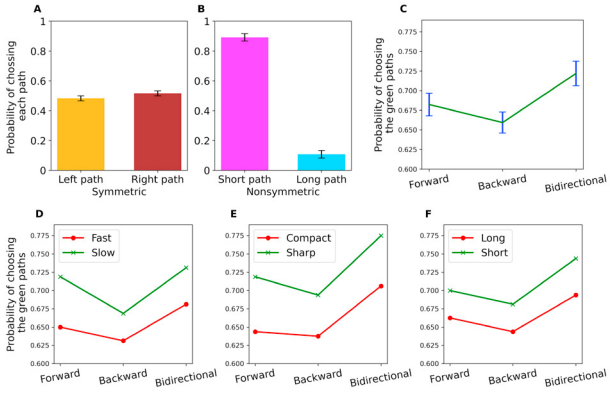

Fig. 2: Experimental results. **(A, B)**: The mean probability of choosing left and right paths in neutral maps. **(C)**: The mean probability of choosing the green path across non-neutral maps. **(D, E, F)**: The probability of choosing the green paths in non-neutral maps in each task

6. Simulation results

Simulated task 1: This task aims to replicate the conditions of experimental Task 1 by varying the appearance time of the maps. The hypothesis is that having more time to make a decision allows for deeper tree expansions within the given computational capacity of the agent. To model this, we manipulated the tree expansion capacity such that $c^{Fast} < c^{Slow}$. To isolate the effect of this capacity change from the inherent Pavlovian bias, we set the probability of expanding trees in the green paths to be the same for both fast and slow maps. Mathematically, this is represented as follows where $\pi^{Fast}(a_g \mid s_s)$ and $\pi^{Slow}(a_g \mid s_s)$ are the probabilities of choosing the green action when the map appears for a fast and slow map, respectively.

$$\pi^{Fast}(a_g|s_s) = \pi^{Slow}(a_g|s_s) > 0.5 \tag{10}$$

To capture the influence of Pavlovian cues on agent behavior, we assigned the probability of selecting the green paths at greater than 0.5, reflecting the embedded approach cues in these paths. Additionally, the likelihood of reaching the goal state in forward planning (or the starting point in backward planning) increases with the depth of the decision trees expanded by the agents. Therefore, longer trees allow the Pavlovian approach cues to exert a stronger bias on the agents' goal-directed decisions.

Figure 3B presents simulation outcomes that align closely with the experimental data displayed in Figure 2D. First, the probability of opting for the green paths across backward, forward, and bidirectional planning is consistently above 0.5, corroborating the idea that Pavlovian approach cues exert a bias on agent decision-making. Second, the data reveals that the tendency to choose the green path is higher in forward planning as compared to backward planning.

Simulated task 2: In this task, we aim to capture the more pronounced Pavlovian bias observed in the second experimental task. To do this, we assigned a higher probability for tree expansion in the sharp maps compared to the compact maps, as indicated by Equation 11.

$$\pi^{Sharp}(a_g|s_a) > \pi^{Compact}(a_g|s_s) > 0.5 \tag{11}$$

However, given that both map types have the same length and allow for the same decision-making time, the expansion capacities were kept identical: $c^{Compact} = c^{Sharp}$. All other simulation parameters were consistent with those outlined in the previous sections. The findings from this simulation, displayed in Figure 3C, closely align with the experimental outcomes presented in Figure 2E. Note that to achieve the conditions specified in Equation 11, the weights for the *SoftMax* policy can be manipulated in the following way, where w_g and w_r show the weighs assigned to green and red paths, respectively.

$$\pi(s_s, a_g) = \frac{e^{w_g Q(s_s,a_g)}}{e^{w_g Q(s_s,a_g)} + e^{w_r Q(s_s,a_r)}} \qquad \ni w_g > w_r \tag{12}$$

Simulated task 3: In Simulated Task 3, we focused on examining how the map lengths could differently influence the agent's decisions under constraints of limited expansion capacities. To isolate the effect of map length, we assigned equal capacities for both short and long maps: $c^{Long} = c^{Short}$. Additionally, to account for the influence of Pavlovian cues embedded in the green paths, we set the probability for tree expansion along the green paths to be higher than for the red paths. The outcomes of this simulation, displayed in Figure 3D, closely align with the experimental results presented in Figure 2F.

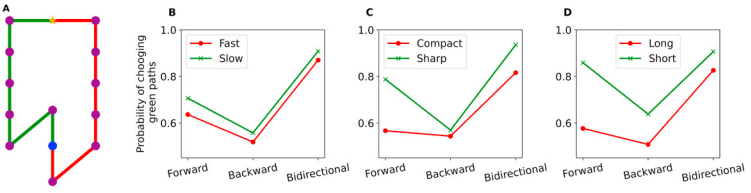

Fig. 3: Simulation results. **(A)**: A sample map used in the simulation. **(B, C, D)**: The mean probability of choosing the green path in the three simulated tasks

In summary, the simulation results provide multiple insights into the role of Pavlovian cues in decision-making under different planning conditions. First, the Pavlovian approach bias has a stronger influence on forward planning compared to backward planning and has an even greater effect in bidirectional planning than in either of the unidirectional methods. Second, the more capacity the agent has for tree expansion in both forward and backward directions, the greater the influence of the Pavlovian approach bias on decision-making. This is due to the increased likelihood of the decision trees reaching the goal state in forward planning and the starting point in backward planning, thus amplifying the Pavlovian bias. Third, the greater impact of Pavlovian cues on forward planning relative to backward planning is attributed to the larger expansion capacity we allocated for forward planning. Furthermore, the possibility that forward and backward planning trees might intersect somewhere in the middle of the map amplifies the Pavlovian influence in bidirectional planning compared to either forward or backward planning alone. Finally, the simulation results validate our research hypothesis and demonstrate the effectiveness of our proposed MBRL algorithm for capturing the complexities of bidirectional planning under the influence of Pavlovian approach bias.

7. Conclusion and Future Directions

We designed a navigation experiment with Pavlovian approach cues near the start and end points of the maps to explore how these cues influence forward, backward, and bidirectional planning in a navigation task where participants selected between two paths. The experimental data confirmed that humans engage in backward planning and that all three forms of planning are susceptible to Pavlovian approach cues, often leading to choices that may not necessarily be better choices.. The data indicates varying degrees of Pavlovian influence across planning types, with the least impact on backward planning, moderate influence on forward planning, and the most significant impact on bidirectional planning. We examined the influence of map length, decision-making time, and approach cue size on different forms of planning. Our analysis revealed that reducing map length, extending decision-making time, and increasing approach cue size all amplify the impact of Pavlovian approach bias on instrumental planning.

To model the observed behaviors, we proposed an algorithm for bidirectional planning within the framework of MBRL. We incorporated the Pavlovian approach bias into the algorithm as a weighting factor for the RL agent's Soft-Max behavioral policy. Specifically, this allows the agent to be more inclined to choose paths that include Pavlovian approach cues. In other words, the algorithm assigns a higher probability to the expansion of deeper decision trees along paths that include these cues as opposed to paths that do not. Our simulation results indicate that the stronger the approach cue, the greater the weight given to the corresponding action. Importantly, these simulation outcomes align well with our experimental data. Additionally, the influences of the three aforementioned factors were modeled by adjusting the depth of the decision trees and their expansion capacity. This concept of bidirectional planning can also be contextualized within the framework of the information sampling phenomenon as described by [13]. While decision trees can grow exponentially with increased depth, constraints such as working memory limitations and computational resources set boundaries on how deep these trees can be. Instead of solely expanding deep trees in just a forward (or solely backward) direction until reaching the goal (or starting point) for evaluation, bidirectional planning samples shallower trees in both directions. This strategy optimizes working memory utilization: the capacity is allocated towards efficient bidirectional planning rather than deep singular directional planning. This is because backward planning allows for the transfer of learned information to forward planning.

In conclusion, our study highlights the importance of considering Pavlovian influences on bidirectional planning and provides a computational framework to model this effect. This work could be extended by designing tasks to test the hypotheses under sequential decision-making paradigms and incorporating reward and state prediction errors to improve the learning of cognitive maps. Investigating the interplay between Pavlovian cues, habitual learning, and goal-directed control would provide a more comprehensive understanding of the decision-making process. Exploring engineering applications for the proposed algorithm and modeling Pavlovian innate responses as a rule-based system, or a pretrained neural network integrated with instrumental learning could lead to more biologically plausible representations and practical utility in fields such as robotics and autonomous navigation.

Acknowledgements

We acknowledge Amir-Homayoun Javadi, Arsham Afsardeir, Hugo Spiers, and Peter Dayan's help in initial ideation.

References

[1] Afsardeir, A., Keramati, M., 2018. Behavioural signatures of backward planning in animals. European Journal of Neuroscience 47, 479–487. doi:https://doi.org/10.1111/ejn.13851.

[2] Boulton, K., 2019. Started from the future now we're here: The effect of planning direction on motivation. Honours Bachelor of Arts in Psychology, Wilfrid Laurier University .

[3] Cavanagh, J.F., Eisenberg, I., Guitart-Masip, M., Huys, Q., Frank, M.J., 2013. Frontal theta overrides pavlovian learning biases. Journal of Neuroscience 33, 8541–8548. doi:https://doi.org/10.1523/JNEUROSCI.5754-12.2013.

[4] Colagiuri, B., Lovibond, P.F., 2015. How food cues can enhance and inhibit motivation to obtain and consume food. Appetite 84, 79–87. doi:https://doi.org/10.1016/j.appet.2014.09.023.

[5] Corneil, D.S., 2018. Model-based reinforcement learning and navigation in animals and machines. Technical Report. EPFL. doi:http://dx.doi.org/10.5075/epfl-thesis-8950.

[6] Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience 8, 1704–1711. doi:https://doi.org/10.1038/nn1560.

[7] Dayan, P., Berridge, K.C., 2014. Model-based and model-free pavlovian reward learning: revaluation, revision, and revelation. Cognitive, Affective, & Behavioral Neuroscience 14, 473–492.

[8] Dayan, P., Niv, Y., Seymour, B., Daw, N.D., 2006. The misbehavior of value and the discipline of the will. Neural networks 19, 1153–1160. doi:https://doi.org/10.1016/j.neunet.2006.03.002.

[9] Dolan, R.J., Dayan, P., 2013. Goals and habits in the brain. Neuron 80, 312–325.

[10] Dorfman, H.M., Gershman, S.J., 2019. Controllability governs the balance between pavlovian and instrumental action selection. Nature communications 10, 1–8. doi:https://doi.org/10.1038/s41467-019-13737-7.

[11] Epstein, R.A., Patai, E.Z., Julian, J.B., Spiers, H.J., 2017. The cognitive map in humans: spatial navigation and beyond. Nature neuroscience 20, 1504. doi:https://doi.org/10.1038/nn.4656.

[12] Heinz, A., Beck, A., Halil, M.G., Pilhatsch, M., Smolka, M.N., Liu, S., 2019. Addiction as learned behavior patterns. Journal of clinical medicine 8, 1086. doi:https://doi.org/10.3390/jcm8081086.

[13] Hunt, L.T., Rutledge, R.B., Malalasekera, W.N., Kennerley, S.W., Dolan, R.J., 2016. Approach-induced biases in human information sampling. PLoS biology 14, e2000638. doi:https://doi.org/10.1371/journal.pbio.2000638.

[14] Huys, Q.J., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R.J., Dayan, P., 2011. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. PLoS Comput Biol 7, e1002028. doi:https://doi.org/10.1371/journal.pcbi.1002028.

[15] Huys, Q.J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., Roiser, J.P., 2012. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS computational biology 8, e1002410. doi:https://doi.org/10.1371/journal.pcbi.1002410.

[16] Khamassi, M., Girard, B., 2020. Modeling awake hippocampal reactivations with model-based bidirectional search. Biological cybernetics , 1–18doi:https://doi.org/10.1007/s00422-020-00817-x.

[17] Lally, N., Huys, Q.J., Eshel, N., Faulkner, P., Dayan, P., Roiser, J.P., 2017. The neural basis of aversive pavlovian guidance during planning. Journal of Neuroscience 37, 10215–10229.

[18] Mogg, K., Field, M., Bradley, B.P., 2005. Attentional and approach biases for smoking cues in smokers: an investigation of competing theoretical views of addiction. Psychopharmacology 180, 333–341. doi:https://doi.org/10.1007/s00213-005-2158-x.

[19] Na, S., Chung, D., Jung, J., Hula, A., Fiore, V.G., Dayan, P., Gu, X., 2019. Humans use forward thinking to exert social control. Available at SSRN 3443690 .

[20] Palminteri, S., Lefebvre, G., Kilford, E.J., Blakemore, S.J., 2017. Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. PLoS computational biology 13, e1005684. doi:https://doi.org/10.1371/journal.pcbi.1005684.

[21] Raab, H.A., Hartley, C.A., 2020. Adolescents exhibit reduced pavlovian biases on instrumental learning. Scientific reports 10, 1–11. doi:https://doi.org/10.1038/s41598-020-72628-w.

[22] Rescorla, R.A., 1988. Pavlovian conditioning: It's not what you think it is. American psychologist 43, 151. doi:https://psycnet.apa.org/doi/10.1037/0003-066X.43.3.151.

[23] Russell, S.J., Norvig, P., 2010. Artificial intelligence-a modern approach, third int. edition. Prentice Hall Series in Artificial Intelligence. Englewood Cliffs NJ: Prentice Hall .

[24] Sezener, C.E., Dezfouli, A., Keramati, M., 2019. Optimizing the depth and the direction of prospective planning using information values. PLoS computational biology 15, e1006827. doi:https://doi.org/10.1371/journal.pcbi.1006827.

[25] Simon, D.A., Daw, N.D., 2011. Neural correlates of forward planning in a spatial decision task in humans. Journal of Neuroscience 31, 5526–5539. doi:https://doi.org/10.1523/JNEUROSCI.4647-10.2011.

[26] Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.

[27] van Timmeren, T., Quail, S.L., Balleine, B.W., Geurts, D.E., Goudriaan, A.E., van Holst, R.J., 2020. Intact corticostriatal control of goal-directed action in alcohol use disorder: a pavlovian-to-instrumental transfer and outcome-devaluation study. Scientific reports 10, 1–12. doi:https://doi.org/10.1038/s41598-020-61892-5.

[28] Watson, P., De Wit, S., Hommel, B., Wiers, R.W., 2012. Motivational mechanisms and outcome expectancies underlying the approach bias toward addictive substances. Frontiers in psychology 3, 440. doi:https://doi.org/10.3389/fpsyg.2012.00440.

[29] Wiese, J., Buehler, R., Griffin, D., 2016. Backward planning: Effects of planning direction on predictions of task completion time. Judgment & Decision Making 11.

Appendix A. Proposed Bidirectional Planning Algorithm under Pavlovian Approach Bias

Algorithm 1 Initialization and Setup for Bidirectional Planning under Pavlovian Approach Bias

Initialize V to a zero vector of size $env.n\_states$
Initialize Q to a zero matrix of size $env.n\_states \times env.n\_actions$
Set γ ▷ Discount factor
Set $capacity_{total}$ ▷ Total working memory capacity of the agent
Set $capacity_{forward}$ to be larger than $capacity_{backward}$ ▷ Forward capacity is prioritized
Compute $capacity_{backward} \leftarrow capacity_{total} - capacity_{forward}$ ▷ Calculate remaining backward capacity
Set $P_{PavlovianBias} > 0.5$ ▷ Probability of Pavlovian bias
$depth_{PBFS}, depth_{NPBFS} \leftarrow DepthDesign(capacity_{forward}, P_{PavlovianBias})$
$depth_{PBFG}, depth_{NPFG} \leftarrow DepthDesign(capacity_{backward}, P_{PavlovianBias})$
▷ PBFS: Pavlovian branches from start state
▷ NPBFS: Non-Pavlovian branches from start state
▷ PBFG: Pavlovian branches from goal state
▷ NPFG: Non-Pavlovian branches from goal state
if $agent.task =' slow\_fast'$ **then**
    Set $agent.capacity['slow']$ to be greater than $agent.capacity['fast']$
    Ensure $Q_{slow}[s_{start}][a_{green}]$ is greater than $Q_{fast}[s_{start}][a_{green}] > 0.5$
else if $agent.task =' sharp\_compact'$ **then**
    Set $agent.capacity['sharp']$ equal to $agent.capacity['compact']$
    Ensure $Q_{sharp}[s_{start}][a_{green}]$ is greater than $Q_{compact}[s_{start}][a_{green}] > 0.5$
else if $agent.task =' short\_long'$ **then**
    Set $agent.capacity['long']$ equal to $agent.capacity['short']$
    Ensure $Q_{short}[s_{start}][a_{green}]$ is greater than $Q_{long}[s_{start}][a_{green}] > 0.5$
end if
▷ Green refers to branches containing Pavlovian cues
call BidirectionalPlanning(V, Q)

Algorithm 2 Bidirectional Planning Algorithm

function BidirectionalPlanning(V, Q)
    $s \leftarrow env.reset()$
    $V, Q \leftarrow RewardBackpropagation(V, Q)$
    while True **do**
        $V, Q \leftarrow ForwardPlanning(s, V, Q)$
        $a \leftarrow SoftmaxActionSelection(s)$
        $s', r, done \leftarrow env.step(a)$
        if done **then**
            break
        end if
    end while
end function

Algorithm 3 Helper Functions for Bidirectional Planning

function DepthDesign($capacity$, $P_{PavlovianBias}$)
  $depth_{PavlovianBranch} \leftarrow 0$
  for $c \leftarrow 0$ **to** $capacity - 1$ **do**
    if $random.random() < P_{PavlovianBias}$ **then**
      $depth_{PavlovianBranch} \leftarrow depth_{PavlovianBranch} + 1$
    end if
  end for
  $depth_{NonPavlovianBranch} \leftarrow capacity - depth_{PavlovianBranch}$
  return $depth_{PavlovianBranch}, depth_{NonPavlovianBranch}$
end function
function RewardBackpropagation(V, Q)
  $V[s_{goal}] \leftarrow r_{goal}$
  for $branch$ **in** $env.branches_{s_{goal}}$ **do**
    for $i \leftarrow 0$ **to** len(branch) **do**
      $s \leftarrow branch[i]$
      $V[s] \leftarrow \gamma^i * V[s_{goal}]$
    end for
  end for
  for s **in** env.states **do**
    for a **in** env.actions[s] **do**
      $s', r, done \leftarrow env.step(a)$
      $Q[s, a] \leftarrow r + \gamma * V[s']$
    end for
  end for
  return V, Q
end function
function ForwardPlanning(s, V, Q)
  for $a \in env.actions[s]$ **do**
    $r_{branch} \leftarrow 0$
    for $d \leftarrow 0$ **to** $depth_{branch}$ **do**
      $s', r, done \leftarrow env.step(a)$
      $r_{branch} \leftarrow r_{branch} + \gamma^d * r$
      $s \leftarrow s'$
      $Q \leftarrow UpdateQValues(Q, s, a, r_{branch}, s')$
    end for
  end for
  $V \leftarrow UpdateVValues(V, Q)$
  return V, Q
end function
function UpdateQValues(Q, s, a, r_{branch}, s')
  $Q[s, a] \leftarrow \max(Q[s, a], r_{branch} + \gamma * V[s'])$
  return Q
end function
function UpdateVValues(V, Q)
  $V[s] \leftarrow \max(Q[s, :])$
  return V
end function
function SoftmaxActionSelection($Q, s, P_{PavlovianBias}$)
  $w_{\text{green}}[s] = a \cdot P_{\text{Pavlovian}} + b$ ▷ $a, b > 0$ are hyperparameters
  $w_{\text{red}}[s] = c \cdot w_{\text{green}}[s]$ ▷ $0 < c \leq 1$ is a hyperparameter
  $P_{a_{\text{green}}} \leftarrow \frac{e^{w_{\text{green}}[s] \cdot Q[s, a_{\text{green}}]}}{e^{w_{\text{green}}[s] \cdot Q[s, a_{\text{green}}]} + e^{w_{\text{red}}[s] \cdot Q[s, a_{\text{red}}]}}$
  $P_{a_{\text{red}}} \leftarrow 1 - P_{a_{\text{green}}}$
  $a \leftarrow$ choose a_{green} with probability $P_{a_{\text{green}}}$, otherwise a_{red}
  return a
end function