UNIVERSIDADE DE LISBOA

Faculdade de Medicina

Episodic and semantic memories in cognitive planning: evidence from a two-stage decision task.

Inês Alexandra Ribeiro Lopes Alves

Orientador: Doutor Bruno André e Silva Miranda

Dissertação especialmente elaborada para obtenção do grau de mestre em Neurociências

2021

UNIVERSIDADE DE LISBOA

Faculdade de Medicina

Episodic and semantic memories in cognitive planning: evidence from a two-stage decision task.

Inês Alexandra Ribeiro Lopes Alves

Orientador: Doutor Bruno André e Silva Miranda

Dissertação especialmente elaborada para obtenção do grau de mestre em Neurociências

2021

A impressão desta dissertação foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de

24 de Novembro de 2020

Contents

Figure index

I. Methods

II. Results

III. Appendices

Table index

I. Appendices

Acknowledgements

I would like to pay my special regards to Dr. Bruno Miranda who coordinated my work (in co-coordination only with his patience and wisdom), during this challenging period in which I learned so much about so many things, me included. It seems almost too obvious to state that this thesis would not be possible without him. With abundant knowledge, means and support, I sincerely doubt there's anyone who would do what he did for me on this project.

My family, all my family… Biological and non-biological, the ones I chose, and those who chose me. For those who have been a source of data, a source of motivation, a source of inspiration: obrigada! I say it in Portuguese because I cannot just say ‘thank you’, it’s too little, and I want to express a higher level of gratitude. I am in debt to you in an immeasurable way.

Abstract

Before a planned decision, we imagine what might happen. Such anticipation of actions' consequences is cognitively demanding and its theoretical and neural substrates remain largely unknown. The dominant view is that when we plan "where am I going to eat tomorrow?" the decision is based on episodic memories, which are autobiographical and reliant on a specific time and place. Therefore, future simulation uses the past personal experience of "last month I've been to Setúbal and ate an amazing cuttlefish" to evaluate whether that option is a good possibility or not. However, neuropsychological evidence has been showing that patients with medial temporal lobe lesions, the region responsible for episodic memory, are still able to imagine certain non-personal future experiences despite their autobiographical impairments both on recollection and anticipation. On the other hand, recent theories emphasize a parallel contribution of semantic memory (reliant on the anterior temporal lobe) to prospective mental activities. Even without any personal experience, by knowing that "Setúbal is best known for its cuttlefish" we could plan to have lunch there tomorrow. However on the extant research data does not specify the particular relationship between semantic memory and planning.

Through a reinforcement-learning framework, the present study tried to explore the role of semantic memory in cognitive planning in a quantitative way. We compared strategies employed by healthy adults, in two versions — episodic and semantic — of a multi trial two-stage decision task, considered the goal-standard in differentiating cognitive planning (model-based or goal-directed behavior) from habitual (or model-free) responses. On our episodic version, the task structure relies on the personal experience of a specific relationship, whereas on the semantic version choice behavior could benefit from a 'semantic clue'. Our initial hypothesis was that such semantic knowledge could potentiate goal-directed behavior. However, contrary to our hypothesis, the balance between cognitive planning and habitual responses did not seem to be influenced by the presence of a 'semantic clue'. The vigour of response (studied through reaction time analysis) also did not show significant differences between the episodic and semantic versions. Various limitations are discussed as possible explanations of such non-significant results, but the main issue might reside in the low sensitivity of the task.

We hypothesized that the preserved episodic ability of our sample of healthy adults might be masking the eventual beneficial role of semantic content in planning. Future work may consider addressing this sensitivity issue by including in this study early-stage Alzheimer's disease (AD) patients. The specific episodic impairment with relatively preserved semantic abilities typically observed in these patients will probably make them more reliant on the semantic association when planning their actions. If our assumption is right, this will demonstrate that semantic memory can be used as an alternative support to cognitive planning.

Key-words: cognitive planning; episodic memory; semantic memory; reinforcement learning; two-stage decision task.

Resumo

É hoje bastante evidente que o cérebro possui vias distintas para o processamento de informação e regulação do comportamento. Umas mais automáticas, dedicadas a coisas que fazemos de forma habitual e rotineira, e outras vias mais controladas que exigem atenção, deliberação, e suportam processos cognitivos superiores como a tomada de decisão planeada. O planeamento cognitivo surge através do funcionamento coordenado de inúmeros sistemas, o que o torna um 'processo' extremamente complexo e uma área cuja investigação ainda tem muito por explorar. Planear geralmente implica um exercício de previsão, uma antecipação de consequências de possíveis ações que é feita com base na informação que cada um dispõe. Relatos de pacientes amnésicos com défices paralelos na imaginação de cenários futuros, e estudos de neuroimagem que mostram uma grande sobreposição da atividade cerebral durante a recordação de experiências passadas e a simulação de experiências futuras, demonstram bem a importância da memória em atividades mentais prospetivas como o planeamento cognitivo.

A visão dominante é a de que, quando planeamos "onde irei almoçar amanhã?", recorremos a informação relacionada com acontecimentos pessoais e associada a um contexto — memória episódica. Assim, a simulação do futuro usa a experiência de "eu estive no mês passado em Setúbal e comi uns chocos deliciosos" para incluir essa hipótese na sua pesquisa, e então avaliar se é ou não uma opção adequada. Contudo, o estudo desta relação memória/prospeção tem-se expandido a outros tipos de memória. Doentes com lesões no lobo temporal médio — região considerada responsável pela memória episódica — apesar dos défices em recordar e projetar conteúdos episódicos, apresentam por vezes capacidades preservadas no domínio semântico. A memória semântica corresponde ao conhecimento de conceitos e significados, informação culturalmente partilhada. Está associada à atividade do lobo temporal anterior e por sua vez, lesões nesta região resultam em défices específicos no domínio semântico mas com preservação de informação do tipo episódica. Estas dissociações no perfil neuropsicológico têm levado alguns autores a sugerir a memória semântica como um suporte alternativo e independente para a simulação. Assim, ainda que sem qualquer referência pessoal ou contexto, utilizar o conhecimento de que "Setúbal é conhecido por ter bons chocos" pode ser uma forma alternativa de planear o almoço do dia seguinte.

Utilizando uma abordagem computacional dirigida à aprendizagem por reforço, o presente trabalho estudou a influência do conteúdo semântico no planeamento cognitivo, através de um método quantitativo. Comparámos a performance de adultos saudáveis em duas versões, episódica e semântica, de uma tarefa de dupla escolha, desenvolvida para distinguir decisões planeadas (*model-based,* MB) de outras mais automáticas e menos deliberadas (*model-free,* MF). Esta tarefa composta por diversos *trials* (ensaios) permite avaliar múltiplas decisões numa só experiência. Através da análise de que variáveis de um *trial* influenciam mais a escolha no seguinte, podemos perceber até que ponto essa ação foi ou não planeada. Cada trial é composto por duas etapas e em cada uma das quais uma de duas imagens deve ser

escolhida. Entre as duas etapas existe uma estrutura de transição estocástica, i.e uma relação que é predominante mas que só acontece em 70% das vezes.

Nesta tarefa, compreender qual a imagem da primeira etapa que maximiza a probabilidade de alcançar uma imagem pretendida na segunda, é fundamental para planear devidamente as escolhas e não agir apenas de forma automática. Quando as escolhas dos participantes se baseiam exclusivamente nas recompensas passadas, considera-se que existe uma predominância de comportamento MF. Por outro lado, se o que se verifica é um efeito significativo da interação entre a recompensa e a estrutura de transição considera-se que há uma atividade prospetiva deliberada: comportamento MB. O que distingue as nossas duas versões da tarefa é que, ao contrário da tarefa episódica, compreender a relação entre primeira e segunda escolhas na versão semântica é mais fácil, uma vez que se trata de uma associação comum e muito provavelmente previamente conhecida. Acreditamos que ter esta 'pista' semântica, ao invés de depender exclusivamente de informação adquirida durante a execução da tarefa, facilite compreensão da dinâmica da tarefa e promova escolhas mais deliberadas.

Uma análise descritiva das escolhas realizadas pelos participantes revelou que, apesar da elevada variabilidade entre sujeitos, tanto a estrutura de transição (relevante para comportamento MB) como o histórico de recompensas (relevante para comportamento MF e MB) influenciaram significativamente as escolhas em ambas as versões da tarefa. Os participantes apresentaram, em média, uma maior probabilidade de repetir a escolha do *trial* anterior quando uma recompensa máxima era obtida através de uma transição comum, do que se fosse através de uma transição rara. O padrão oposto (maior probabilidade de repetir a escolha depois de uma transição rara) verificou-se para os dois níveis de recompensa média e mínima. A nossa análise estatística confirmou que em ambas as tarefas as escolhas basearam-se numa combinação das duas estratégias. Os resultados de uma regressão logística mostraram que tanto a recompensa como a interação da recompensa com a transição, foram bons preditores da escolha, contudo uma predominância de comportamento MB (planeado) também foi observada. Quando analisados os efeitos destas variáveis dos cinco *trials* antecedentes sobre a escolha, confirmámos resultados de estudos anteriores que demonstraram que os efeitos sobre as estratégias MF e MB permanecem ao longo do tempo, e decaem de uma forma exponencial.

Ao contrário da nossa hipótese, de que a associação semântica iria facilitar a utilização da estrutura de transição nas escolhas, os nossos resultados não revelaram qualquer relação estatisticamente significativa entre a versão da tarefa (episódica ou semântica) e o balanço entre estratégias automáticas e planeadas. São discutidas várias limitações que poderão estar a influenciar a análise, mas uma possível explicação para estes resultados não significativos é a de uma baixa sensibilidade da tarefa para detetar o contributo do conteúdo semântico. O facto de a aquisição da estrutura de transição não ser muito exigente (associar duas imagens a dois conjuntos imagens), e a nossa amostra não possuir qualquer tipo de défice cognitivo diagnosticado, poderá tornar irrelevante a presença de uma relação conhecida. Ainda que a associação semântica possa facilitar numa fase inicial a aquisição do modelo,

ao longo de quase duas dezenas de *trials* esse efeito acaba por ser mascarado e a nível do comportamento nas escolhas efetuadas, as nossas duas versões acabam por parecer uma só tarefa apenas com um conjunto de estímulos diferentes.

A hipótese de que as capacidades de armazenamento episódico (preservadas nos nossos participantes) 'camuflaram' o possível efeito do conteúdo semântico para o planeamento poderá ser estudada aumentando a dificuldade da tarefa (como por exemplo diminuindo os tempos de reação limites) ou aplicando este estudo a populações clínicas com problemas mnésicos. Perspetivas futuras passam por realizar este estudo com pacientes em fases iniciais da doença de Alzheimer. O perfil neuropsicológico desta patologia, caracterizado por um defeito episódico específico e uma memória semântica preservada, irá eventualmente fazer com que os participantes dependam mais da associação semântica para corretamente estabelecerem a relação entre escolhas. Se estivermos corretos, ao contrário dos controlos, no caso de doentes de Alzheimer um aumento no comportamento MB será observado na tarefa semântica por comparação à episódica, o que irá suportar a nossa hipótese de que a memória semântica pode ser um suporte alternativo à memória episódica, no planeamento cognitivo.

Além da análise das escolhas dos participantes, também os tempos de reação foram avaliados. O *trade-off* entre velocidade e precisão do comportamento MF e MB (sendo o último mais lento) previamente descrito não foi confirmado pelo nosso trabalho. Por outro lado, o facto de que, à medida que a tarefa progredia nos *trials* os participantes foram ficando mais rápidos, sem detrimento da sua performance (em relação à capacidade de agirem de forma planeada), suporta teorias de automatização que defendem que um comportamento planeado e estruturado com base num modelo, com a prática pode tornar-se automático (ou MF) mas que no entanto continua a ter em conta um modelo. Neste sentido, perspetivas futuras passam também por treinar os participantes nesta tarefa ao longo de várias sessões. Avaliar a evolução da performance permitirá estudar esta área e perceber melhor o paralelismo entre as dicotomias de comportamento MF/MB e automático/planeado.

Palavras-chave: Planeamento cognitivo; memória episódica; memória semântica; aprendizagem por reforço; tarefa de dupla escolha

1 Introduction

"Planning is the crowning achievement of human cognition", this is what Grafman, Sepctor & Ratterman conclude, while providing a description of the neural correlates of the distinct cognitive components of planning (Morris & Ward, 2005). If we scrutinize planning behavior we realize that it involves a myriad of cognitive components, orchestrated to "integrate events from past, present, and future into a single plan-level memory" (Morris & Ward, 2005). This prospective activity represents a very helpful strategy to prepare the individual for future actions. In fact, the idea that we can imagine future scenarios and how our actions may impact upon imagined scenarios, provide humans with a huge evolutionary advantage (Benson, 1993). Plans range from short-term and motoric to long-term and cognitive, and they are a constant in our existence as individuals that plan actions every day and several times a day. Planning also integrates a societal context, much of our cultural products (e.g. symbolic representations, science, religion, etc) derive in part from our capacity to plan with respect to an uncertain (but potentially controllable) future (Klein et al., 2010).

Generating plans is such a basic condition in our life experience that most people do not realize how important and valuable this skill is until it becomes compromised. Planning deficits have been described in numerous conditions such as aging (Craik, 2006), brain damage (Goel & Grafman, 1995; Harlow, 1868; Owen, 1997), neurodevelopment disorders (Demetriou et al., 2019), neurodegenerative diseases (Grafman et al., 1992; Irish et al., 2016; Kliegel et al., 2005; Weintraub et al., 2012), and other mental illnesses (Zhu et al., 2010). According to several developmental theories, teenage pregnancy, drunk driving, and risk-taking behavior in adolescence are consequences of the failure or resistance to plan (Furby & Beyth-Marom, 1990). A deep understanding on how we construct, execute and monitor our plans will make easier to find therapeutic solution to compensate or rehabilitate from specific cognitive dysfunctions (Mikadze, 2014). As noted by Redish (2013), planning is the least exploited mode of decision-making. While neural correlates of simple, fast, automatic action-selection systems (like reflexes and pavlovian/non-pavlovian habits) are well documented, the neurobiology of planning is still on its infancy (Redish, 2013).

Formerly the term "planning" was used in a very loose way (Goel & Grafman, 1995). Different types of planning were not taken into account and a sharp distinction between the construction, execution and evaluation phases of planning was missing (along with the steps and cognitive demands that each of these phases require). Moving from subjective anecdotal reports of disorganized behavior, to continuously improving methods of neuropsychological assessment, deep investigations on the cognitive architecture underlying planning behavior, and neuroimaging studies revealing the neural correlates of such activity, today we have a much more detailed

and complex description of the construct of planning. As a multidisciplinary area of research, extant knowledge about the different aspects of planning is the cumulative result of contributions from different fields such as neuropsychology, cognitive science/neuroscience and artificial intelligence (Morris & Ward, 2005).

Classically, planned behavior has been associated with the function of frontal lobe (FL), in particular the prefrontal cortex (PFC). Such correlate goes back to the mid 19th century, when the remarkable survival of Phineas Gage and the detailed evaluation of behavioral changes and his inability to make decisions, made Harlow to be the first to argue that human frontal lobe lesions resulted in loss of 'planning skills' (Harlow, 1868; Owen, 1997). Numerous limitations can stem from using brain damage to infer correlations between function and location e.g. lesions are not circumscribed to a particular 'functional module', possible reconfiguration after lesion, among other possibilities (Rorden & Karnath, 2004). Nonetheless, large collections of data from studies in patients with FL lesions (like those provided by Luria in the mid-20th century) (Stuss & Benson, 1984) reinforced the idea that PFC activity is responsible for 'higher cortical functions' such as planning. Luria noted that lesions in what he called 'the third unit' (on the PFC), would cause malfunction when planning actions, consciously controlling impulses, and other complex competencies (Languis & Miller, 1992; Mikadze et al., 2019; Stuss & Benson, 1984).

Thereafter numerous theories were developed around this idea of a superior mechanism of control and attention carried out by the PFC and responsible for 'higher-level' functions (Benson, 1993; Lezak, 1982; E. K. Miller & Cohen, 2001; Shallice, 1982; Stuss, 2011). In the 1970's the term 'executive functions' emerged (Naglieri et al., 2014), as an umbrella term used for a diversity of hypothesized top-down mental processes needed when you have to concentrate and pay attention (or when going on automatic or relying on instinct or intuition would be ill-advised, insufficient, or impossible) (Diamond, 2017). Among these diverse processes, planning is recognized as a prototypical executive function (Benson, 1993; Lezak, 1982; Shallice, 1982; Zelazo et al., 2006) as well as a good candidate to prove the existence of such construct. In the early 1980's a theory-driven task — the Tower of London test (ToL) — was adapted from an older puzzle by Shallice and McCarthy (described in Shallice, 1982) and became a widely used test in the assessment of the 'look-ahead' ability, a construct said to underpin planning behavior (Anderson et al., 2008; Baker et al., 1996; Goel & Grafman, 1995; Morris & Ward, 2005; Owen, 1997). However, such task might not be a good model for studying planning in more realistic situations "because it is too structured"[1](Goel & Grafman, 1995; Morris & Ward, 2005). Although the term executive function emerged as a list of processes that seemed to be impaired after FL lesions, as noted by Stuss, planning deficits (found in ToL task) are also seen in several conditions where there is no involvement of frontal structures (Owen, 1997; Stuss, 2011).

[1] According to information processing theories, planning is seen as a problem-solving method. For Newell and Simon (1972), problem-space consists of four kinds of elements: *initial state, goal state, operators and path constraints.* 'Ill-defined' problems are those where information about these four components are incomplete, like the two-stage decision task; the ToL task is an example of well-defined problem.

In the decision-making literature, planning behavior is regarded as the ability to select actions in a deliberative way (Redish & Mizumori, 2016). In contrast to automatic systems (that rely on reflexes, Pavlovian or habitual processes), planning is a conscious and computationally expensive process (Meer et al., 2015). It entails: a sequential and serial search through possibilities; prediction ("what if?"); evaluation; and finally action-selection (organized in time and space). To make those steps possible, one requires: knowledge of the structure of the world (particularly cause-effect structures that allows prediction); imagination, to mentally construct possible future scenarios (and recognize potential futures as imagination, and not real); an active memory process, capable of remembering what paths you've searched through (and those that are still left to explore); an evaluation process, where you compare options attending to specific pros and cons; and a working memory ability to keep multiple options, their evaluations, and the paths to them (Meer et al., 2015; Redish, 2013). Despite not being explicitly mentioned long-term memory is implicitly associated with this process since the initial search is only possible due to the knowledge that is stored in our memory systems and the ability to access it.

This dichotomy between reflexive (automatic) versus reflective (conscious) decision-making and choice parallels the long-lasting idea in psychology and cognitive neuroscience that different learning strategies can be used to control action (Dolan & Dayan, 2013). Theories on instrumental behavior[2] stated that it can be controlled by two mechanisms: a stimulus-response or habit mechanism; and a goal-directed process (Balleine & Dickinson, 1998). Thorndike's Law of Effect (1911), a pioneer on stimulus-response theories, states that "responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation, and responses that produce a discomforting effect become less likely to occur again in that situation" (Nevin, 1999). Few years later, based on his rodent studies in maze navigation, Tolman (1948) argued that animals are not merely "condemned" to this simple and habitual associative structure. He proposed that animals could learn a maze task by forming "a field map of the environment": an internal representation (or image) of external environmental features or landmarks that could be use to guide behavior in a flexible, adaptable and planned fashion (i.e. through the use of a cognitive map) (Tolman, 1948). From the perspective of neuroscience, habits and goal-directed action systems appear to coexist in different corticostriatal circuits (Owen, 1997).

Computational approaches interpret these two action systems, habitual and goal-directed, as complementary mechanisms for reinforcement learning (RL) (Daw et al., 2011; Redish, 2013; Sutton & Barto, 2015). The habitual mechanism is referred to as "model-free", in the sense of eschewing the representation of task structure, working directly by reinforcing successful actions and operating on a trial-and-error

[2] Instrumental learning or *operant conditioning* is a learning process distinct from what Pavlov previously described as *classical conditioning*. During the latter, conditioning a neutral stimulus becomes a conditioned stimulus by pairing its occurrence with an unconditioned stimulus that naturally elicits some response; whereas, during instrumental conditioning the agent is trained to make a response by delivering an attractive outcome (Skinner, 1984).

basis (Sutton & Barto, 2015). On the other hand, goal-directed planned behavior works by using a world model (or 'cognitive map'), being known in RL as "model-based" (MB) (Daw et al., 2005). Each of these methods have different advantages and disadvantages, and the brain seems to make use of each method in circumstances for which it is best (Daw et al., 2005). The MF action selection is fast and not computationally demanding, however much training is needed to obtain accurate performance. On the other hand, the MB method computes through forward search/simulation, the consequences of actions. Despite being computationally expensive, MB predictions are more accurate in instable environments and flexible to changes (Sutton & Barto, 2015).

Although it has been neglected for a long time, it is becoming increasingly clear that memory plays a major role in decision-making. Some authors even propose the thesis that memory is an adaptive function that evolved specifically to support future decisions and judgments, enabling its possessor to anticipate and respond to future contingencies that cannot be known with certainty (Klein et al., 2010). In classical neuropsychology the classification into multiple memory systems, seems to present a curious and relevant parallel with the distinction of multiple decision-making systems (Doll et al., 2014). As seen above, there can be multiple paths to a decision: habitual or model-free mechanisms and goal-directed or model-based processes. In the memory literature we find a similar and related distinction for long-term memory 'sub-systems': procedural or implicit and declarative or explicit (Squire, 2004). Procedural memory is mainly related to skill acquisition and is thought to be inflexible, implicit, incremental (and reliant on subcortical structures such as striatum). By contrast, declarative memories are used for learning facts and events, being more flexible and relational (and reliant on hippocampus) (Squire, 2004). Research has shown that model-free strategies could explain classic procedural memory tasks (Shohamy et al., 2004), cognitive declarative load relates to MB but not MF RL (Otto et al., 2013), and MB interference disrupts declarative, but not procedural task learning (Foerde et al., 2006). Thus, the distinction between procedural and declarative memories appears to be closely related to that between MF and MB RL, with MF RL corresponding to procedural memory and MB corresponding to declarative memory (Doll et al., 2014).

The importance of declarative memory in prospective activities, such as planning, has been increasingly highlighted (Gershman, 2017; Klein et al., 2010; Schacter et al., 2012; Suddendorf & Corballis, 2017; Szpunar et al., 2014). Despite recent interest, the role of 'memory for the future' is a relatively old idea. Testimonies of patients with Korsakoff Amnesia revealing planning issues (Talland, 1965) and anecdotal reports of severely impaired amnesic patients lost 'ability to imagine the future' (Harlow, 1868; Tulving et al., 1991) are good examples of early reports on how memory serves future-oriented processes (like planning). Planning implies making choices based on simulations that allow us to evaluate actions not yet performed (Geffner, 2013). This 'simulation' crucial for planning (Szpunar et al., 2014) requires access and manipulation of information stored in our memory systems. Accordingly, numerous studies support the hypothesis that imagining the future depends greatly on the same neural machinery necessary to remembering the past

(Gershman, 2017; Irish et al., 2012; Klein et al., 2002, 2010; Schacter et al., 2012; Suddendorf & Corballis, 2017).

The dominant view is that planning depends on the episodic memory and the well functioning of the medial temporal lobe (MTL), in particular of the hippocampus (Gershman & Daw, 2017; Hassabis et al., 2007; Klein et al., 2002; Schacter et al., 2007; Vikbladh et al., 2016). Episodic (along with semantic) memory, is commonly described as a subtype of declarative memory (Squire, 2004). Episodic memory is related to personal/autobiographic events, associated to a specific time and space and reliant on hippocampal formation (Tulving, 1985; Tulving & Markowitsch, 1998). A simple example of how we use this type of memory in planning is, for instance, remembering that: 'when I was in Setúbal I had this amazing cuttlefish' and the next time I want to eat good cuttlefish, I will recall that episode and consider it as a candidate option. According to the predominant view these autobiographical memories form the basis of the knowledge about action consequences (Klein et al., 2002).

Besides being empirically evident, the postulate of episodic memory dependency for planning relies on robust evidence of classical and clinical neuropsychology. On one hand, patients with hippocampal amnesia reveal very poor descriptions when asked to imagine future experiences (Hassabis et al., 2007). Likewise, patients with vascular or tumoral lesions in the hippocampus show lower chances of choosing options that requires imagination of future outcomes, even when those options lead to better rewards (Palombo et al., 2015). Moreover, neuroimaging studies in healthy subjects show that the hippocampal formation is significantly active during planned decision-making (Doll et al., 2015). Nonetheless, several authors question the reductive idea that planning only relies on episodic memories, and alert for other essential aspects. The relationship between memory and future-oriented processes should involve other types of information besides episodic, in particular semantic knowledge (Irish et al., 2016; Klein et al., 2002; Szpunar et al., 2014).

A typical example of episodic memory deficit is the impairment observed in Alzheimer's disease (AD), which is the most important cause of dementia in the western population (Eratne et al., 2018). This neurodegenerative disease is characterized by a progressive deposition of amyloid plaques and neurofibrillary tangles in the MTL, leading to consequent MTL atrophy (Eratne et al., 2018). Early stage patients commonly reveal specific neuropsychological deficits in the episodic domain, both on the anticipation and recollection of personal events. However, these patients show a relatively preserved ability to access and imagine semantic content (Weintraub et al., 2012). Evidence shows that patients with hippocampal lesions are able to predict non-personal events (e.g. providing adequate answers when questioned about important medical breakthroughs likely to take place in the next few years) (Irish et al., 2016; Klein et al., 2002). Such findings highlight the relevance of studying the role of semantic memory in cognitive planning.

Semantic memory refers to non-personal facts, meanings, and culturally shared knowledge, without autobiographical references (Squire, 2004). According to some authors (Irish et al., 2016; Szpunar et al., 2014), semantic memory supports an independent and alternative architecture for cognitive planning. While remembering

the episode "of eating cuttlefish in Setúbal" is part of one's episodic memory, knowing that "Setúbal is best known for its cuttlefish" belongs to the semantic repertoire[3]. In this regard, in case of a specific memory type deficit (e.g. episodic deficit) one could always rely on preserved counterpart (e.g. semantic alternative). Going back to our previous example, if we ever want to eat cuttlefish but we are missing episodic (autobiographical) information, we could plan a trip to Setúbal based on our semantic knowledge (and go there as if it was the first time). Furthermore, evidence from patients with semantic dementia (with semantic deficits but episodic memory preservation in early stages) show a specific deficit on imaging future 'semantic' scenarios relying only on autobiographical (episodic) projections (e.g. when questioned about technological developments, they tend to relate with personal aspects (Irish et al., 2016)).

Semantic dementia (SD) is a rare variant of fronto-temporal dementia (Olney et al., 2017). It is characterized by a progressive loss of knowledge about words, known objects, meanings and concepts. Anatomically, it has been associated with bilateral atrophy of both ventral and lateral portions of the anterior temporal lobe (ATL), predominantly affecting the left hemisphere (Hodges & Patterson, 2007). Usually, one of the first manifestations is the difficulty of 'finding the words', people and place's names, which is commonly associated with a general memory deficit. However detailed neuropshychological evaluation of these patients reveals preserved episodic memory function (at least in early stages). Contrasting with AD patients, those with SD do not have any episodic defect either on the remembering autobiographical past events or imagining themselves in the future. Their deficits reside on impersonal knowledge, forgetting words and meanings, and struggling to prospectively think about non-personal facts (Irish et al., 2016).

These above-mentioned findings clearly favor the relevance of semantic information to prospective activities; however these studies do not inform directly on the cognitive planning process itself. The ability to simulate the future was only assessed through subjective questionnaires and the use of tasks requiring episodic or semantic knowledge for planned decision-making has been very limited. Therefore, exploring the role of both types of declarative memories in cognitive planning seems relevant, innovative and promising. Through the RL computational framework, considered the goal-standard in differentiating habitual and goal-directed (planned) actions, we can compare planning behavior in the presence (and absence) of semantic information.

[3] This example will only make sense to those that, in some circumstance, acquired this information, and that's what culturally shared knowledge means.

Main objective and specific aims

The main goal of the present work was to investigate the cognitive processes underlying the simulation phase of the planning process. In particular, we aimed to ascertain the role of semantic (versus episodic) memories in cognitive planning, through a RL computational framework. To achieve this, we quantified and compared performance of healthy adults in two slightly distinct - 'episodic' and 'semantic' - versions of a two-stage decision-making task (Daw et al., 2011). The original version is widely considered as the gold-standard measure for the contribution of model-based (planned) and model-free (habitual) systems to human choice (Shahar et al., 2019), being a reference for the study of cognitive planning (Daw et al., 2014). Furthermore, such decision-making task also allows the analysis of several decisions per session, along with a behavioral readout of whether those decisions were planned or not (K. J. Miller et al., 2016). These are relevant advantages that allowed us to overcome major limitations of previous studies on planning.

The task consists on multiple trials; on each trial two moments of choice occur and ultimately a reward is received. The main goal of the participants is to get, on each trial, the maximum reward possible. To achieve this purpose, they should understand the relationship between first and second stages, i.e. realize that each option at the first stage choice leads more likely to a particular second stage scenario. If participants successfully create this transition model, and use it to guide decision, when assessing which events in a trial better predict choice behavior in the next one, what we find is that both reward and transition affect their behavior. This expected pattern of model-based behavior is distinct from model-free behavior once the latter is agnostic to the transition model and for which only reward history influences choice. What distinguishes the two task versions is the link from first to second stages that is dependent, either on personal experience and knowledge acquired during task execution — 'episodic task', or on culturally shared and previously acquired knowledge — 'semantic task'.

In regard to more specific aims, we want to compare choice behavior in both versions of the task (through multiple logistic regressions), looking for which past trial events can best predict subsequent choice behavior. The rational is that by adding a 'semantic clue' relating both stages, the transition structure could be better perceived by participants and, hence, model-based behavior will be facilitated (identified by an interaction of task version with the MB indicator). In addition, analysis of participants' reaction time will allow the study of the previously described speed/accuracy trade-off between MF and MB computations (with the former typically being faster and the latter relatively slower) (Keramati et al., 2011).

2 Methods

Participants

A total of 22 healthy adults (12 females, mean age 36 $\pm$ 12.42 SD years, range from 20 to 62), with mean education duration of 12.95 $\pm$ 2.17 SD years participated in this study. Subjects with neurologic and psychiatric pathologies were excluded from the study. Data from one subject had to be removed from analysis due to a persistent behavior in first stage choice (always choosing the same option); therefore, the size of our sample was 21 for all the analyses. Subjects were recruited based on proximity circles of the researchers (7 relatives, 10 friends and 5 coworkers). Before participation in the study, all participants read and signed an informed consent form. The study was approved by a local Ethics Committee (Faculdade de Medicina, Universidade de Lisboa) and participants did not receive any kind of remuneration for participation in the study.

All participants were native European Portuguese speakers and the instructions for the task were exactly the same among the subjects (instructions shown in appendices section). Subjects were instructed about the main task design and the overall structure of the transition matrix, specifically that each first stage option was predominantly (i.e. 'common transition') associated with either second stage states, but not which one (no more specificities were provided regarding the options). They were also instructed that the probability of common transitions remain fixed throughout the task but the specific percentage (70%) of common versus rare transitions was not mentioned. Finally we also instructed the participants that the reward levels of each second stage option would remain fixed during some trials, but would then eventually change.

In order to guarantee that the instructions were clear and sufficient (and no difficulties with equipment interactions existed), we asked participants to perform few trials of an 'induction task', similar to the actual episodic and semantic tasks, except in the stimuli set presented (see detailed description below).

Behavioral tasks

Two behavioral tasks were developed to study the episodic and semantic components of cognitive planning. The design of both tasks was based on the two-stage decision task, previously used in humans (Daw et al., 2011) and animals (Miranda et al., 2019), and considered as a reference in the study of goal-directed behavior (Daw et

al., 2014). The importance of the episodic content for the original task has been previously mentioned (Daw et al., 2011; Doll et al., 2015). On the other hand, the version of the task that includes a semantic component is novel and designed specifically for this study.

Following the original task (Daw et al., 2011), both episodic and semantic tasks consisted on several trials where the main objective of the participants was to accumulate the highest possible rewards. Each trial was composed by two stages (two moments of choice), followed by a reward associated with the option chosen in the second stage (Fig. 1A). The first stage choice took the participant predominantly (0.7 probability) to one of the two possible states in the second stage: state 2a (orange background) and state 2b (blue background) (Fig. 1A). Following nomenclature of figure 1A, if picture A is chosen at first stage, 70% of the times the second stage options will be state 2a (picture C and D). This is referred to as a 'common' transition; while the 30% of the situations where the opposite occurs (i.e. choosing picture A and encountering state 2b), is referred to as a 'rare' transition. First stage picture B has the inverse relationship for common and rare transitions.

Each second stage option (pictures C, D, E, and F) had an associated outcome (referred to as 'Reward') that could assume one of three categorical levels: high (3 coins), medium (1 coin), and no reward (red cross). In order to encourage ongoing learning and assess planning behavior, these rewards evolved over time according to a random Gaussian independent walk (Figure 1B). Importantly, for each second stage picture the outcome level remained the same for a minimum number of trials (a uniformly distributed pseudorandom integer between 5 and 9) and then, either stayed in the same level (with one third probability) or changed randomly to one of the other two possible outcome levels. The reward matrix (example in Fig. 1B) was the same for both tasks, so that the results could be compared (episodic versus semantic) in terms of global performance, choice behavior, and reaction time.

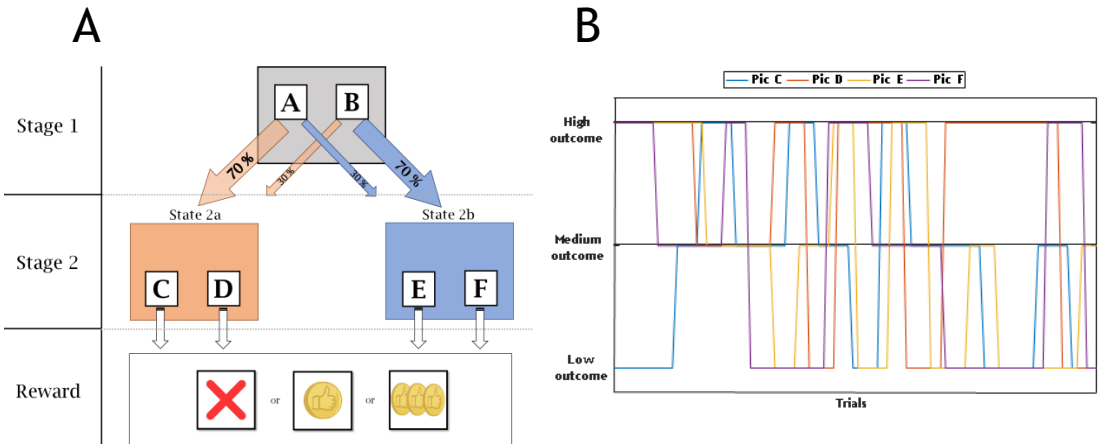


Figure 1: Two-stage decision task. (A) Main task design. Each first-stage choice is predominantly associated (70% probability) with one or the other of the second-stage states. Second-stage choice (picture C, D, E, or F) has an associated reward that can be one of three possible levels. **(B) Example of a reward Matrix.** Each second-stage stimuli has an independent outcome level that remains the same for a minimum number of trials, and changes independently, according to Gaussian random walks. Plot depicts the actual reward structure used for the first 100 trials.

Episodic and semantic versions of the two-stage decision task

In the episodic task, the first stage choice was between two fractal pictures and the second stage choice was between either two 'measuring instruments' or between two 'musical instruments'. As shown in figure 2A, the state-transition structure is semantically irrelevant: choosing the fractal with pink patterns leads 70% of the time to state 2a (ruler and watch), while for the other fractal the common transition is to state 2b (tuba and cello).

On the other hand for the semantic task, pairs of pictures with semantic relationship had been chosen (Figure 2B). First stage options were a 'tree' and a 'globe'; the common transition of choosing 'tree' was a picture set of 'fruits' (orange and banana); whereas for the 'globe', the common transition was a picture set of 'means of transport' (train and airplane).

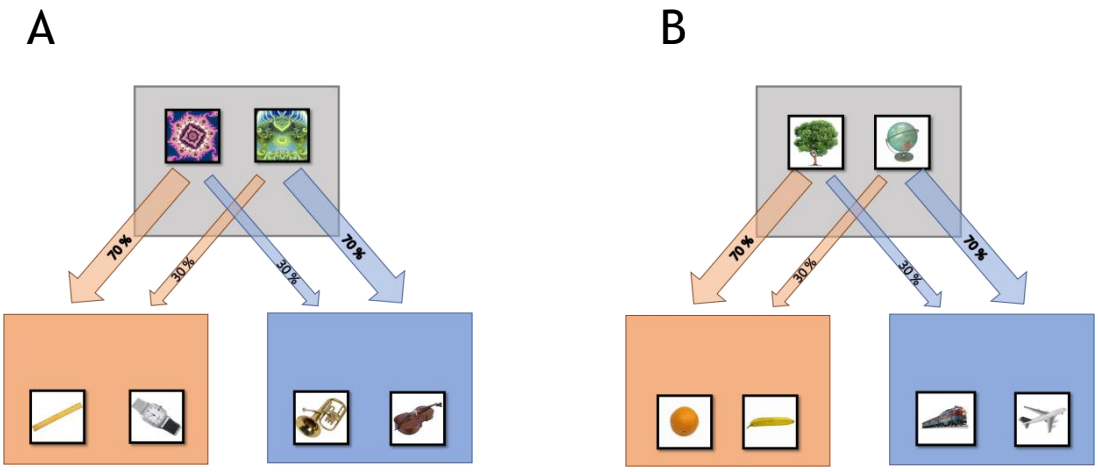


Figure 2: State-transition structures. (A) Episodic task: the transition structure from first to second stages is semantically irrelevant; learning the relation between the first and second stages depends exclusively on episodic information acquired during task execution**; (B) Semantic task:** the common state transition of each first-stage stimuli is semantically related; the tree picture takes predominantly the participants to a set of fruits that grow on trees, whereas the globe picture more often transitions to the means of transportation picture set.

Procedure

The tasks were programmed in Matlab® R2019a environment, using for the presentation of stimuli and response monitoring the Monkeylogic software (http://www.brown.edu/Research/monkeylogic/). Participants performed the task sitting in a quiet room at a comfortable distance from a laptop computer with touch screen. After receiving the task instructions, participants performed the 'induction task' (in all the same as the test task except in the set of stimuli presented) for approximately 2-3 minutes, followed by one of either versions of the task (episodic or semantic). The order of administration of the episodic and semantic tasks was randomly assigned to each patient in order to avoid a learning bias (11 started with the episodic task). Between tasks there was a time interval for participants' rest that ranged from 5 to 20 minutes according to each subject's needs. Each task was performed during a period of 20 minutes. At the end of each task, subjects were questioned about the state-transition structure, to see whether or not it was correctly perceived.

For each stage participants had 5 seconds to choose between two available options (if longer, the trial was aborted). Only the choice made at the second stage had an associated reward. After the reward was presented, it disappeared and a delay of 3 seconds was given until the next trial started (the events in the trial are presented in figure 3). Participants made their choices by touching with a finger in the desired picture on the screen. The on-screen display side of each picture, for the first and second stages options, was randomized (to avoid any type of motoric bias).

To limit potential bias in task performance related to specific picture salience, the pictures were selected from *The Bank of Standardized Stimuli (BOSS)* (Brodeur et al., 2010). In here stimulus are normalized for name, category, familiarity, visual complexity, object agreement, viewpoint agreement, and manipulability.

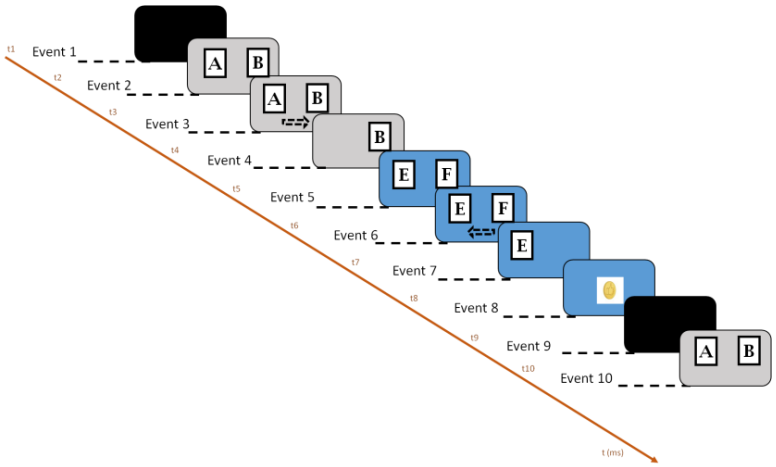


Figure 3: Example sequence of events in the behavioral task. A first moment of choice between two options (picture A and B). Here, choosing B takes through a common transition to state 2b options (picture E and F). Second choice is made and the reward is revealed (picture E rewarded with one coin (medium outcome)).

Data acquisition and processing

Behavioral data was automatically collected during task execution by the Monkeylogic software. At the end of each task a file with all 'raw' behavioral data was saved for further processing. This file included a set of variables saved by default and a set of variables that we programmed to be saved (e.g. picture options, picture choice, side choice, reward, reaction times, among others). In the beginning of the processing, all aborted trials were removed, and a basic analysis of bias for each stage was conducted and included *picture chosen* and *side response* biases. For first and second stages reaction times, we looked for significant differences between episodic and semantic versions, and according to the order of task administration. As mentioned before, one participant had to be excluded due to a major bias on picture chosen at first stage (the exact same picture was chosen for the entire session).

Analysis strategy

The analyses mentioned bellow were performed in Matlab® R2019a environment. For inferential statistics, significance was assessed at $\alpha=0.05$, unless otherwise stated. Each variable was mean centered, and continuous variables were also scaled by dividing them by twice their standard deviations so that the magnitudes of regression coefficients could be directly compared (Gelman, 2008).

Although our main goal was to compare MF and MB RL strategies in episodic and semantic tasks, as we instructed the participants that the main goal of the task was, on each trial, to collect the maximum reward, we were curious to see if there was any significant difference in global reward collected between task versions. For that, we took the minimum number of trials performed among all participants, and used that number as a reference. We calculated for each subject, and each task, the cumulative reward (up to the minimum number of trials found) using 0 for the low outcome, 1 for the medium, and 3 for the high (as shown to participants during task execution); the results were compared and for statistic evaluation we used a paired t-test.

Choice behavior analysis

There are two key variables (Daw et al., 2011) that can condition subjects' choice behavior (and make them more likely or not to keep previous choices). Those variables are: the reward previously received (relevant for both MF and MB behaviors); and whether it occurred after a common or rare transition (given the first stage choice) (relevant only for MB behavior).

First, we studied both episodic and semantic tasks data independently. To describe the basic features of data, we plotted the probability of repeating a first choice as functions of previous reward and previous transition.

To assess further the extent to which a participant relies on a MF or MB strategy, a multiple logistic regression was used to quantify the previous trial (t-1) factors predicting first stage choice at trial t, C_t. Behavioral variables were defined as: C is first stage choice (1=pic A, 0=pic B); R is outcome level (referred to as "Reward"; assumed as continuous, with low=1, medium=2, high=3); T is transition (rare=1, common=0). Our predictors of interest were: $R_{t-1}\times C_{t-1}$ which quantified the main effect of reward on repeating first stage choice (indicator of MF behavior) and, $R_{t-1}\times T_{t-1}\times C_{t-1}$ that captured the interaction effect of both reward and transition type (indicator of MB behavior). Although those were the predictors of interest, we estimate a full factorial model with all interactions to ensure that our results are not confounded by other possibilities that we did not hypothesized: C_{t-1}, modeling a potential independent tendency to stick with the same option (perseveration); R_{t-1}, T_{t-1}, $R_{t-1}\times T_{t-1}$, measuring any potential preference in first stage picture choice given the previous reward, the previous transitions and the interaction effect of both, respectively; and $T_{t-1}\times C_{t-1}$, which quantified the main effect of transition in repeating the previous choice.

According to reinforcement learning theory, in both MB and MF strategies, the effects of trials in the further past tail off, typically exponentially (Sutton & Barto, 2015). Therefore, we included also information from the last 5 trials $i \in \{1,2,3,4,5\}$ into our model the evaluate such decay, using the exact same predictors (C_{t-i}, R_{t-i}, T_{t-i}, $R_{t-i}\times T_{t-l}$, $R_{t-i}\times C_{t-i}$, $T_{t-i}\times C_{t-i}$, and $R_{t-i}\times T_{t-i}\times C_{t-i}$).

To evaluate whether MB and MF learning effects covaried with task version (episodic versus semantic), we put together both episodic and semantic tasks data, into a single model, and performed another multiple logistic regression. The explanatory variables were each interacted, across subjects, with task version characterizing to what extent each of the baseline model parameters changed, across subjects, as function of whether it was the episodic or semantic version. We defined task version as V (episodic=1, semantic=0). Here, our predictors of interest were: the three-way interaction $R_{t-1}\times C_{t-1}\times V_{t-1}$ (influence of task version in MF), and the four-way interaction $R_{t-1}\times T_{t-1}\times C_{t-1}\times V_{t-1}$ (influence of task version in MB). As mentioned before, to avoid learning bias, the order of administration of the two task versions was randomly assigned to each patient. To examine whether possible relations between task version and RL strategies were affected or influenced by this variable, we also include as a predictor in our logistic regression a variable O (first task=1, second task=0), measuring any potential effect of learning. For this analysis only information from previous trial (t-1), was included.

Reaction times analysis

First stage reaction time, RT1 was defined as the time from first stage stimuli presentation to participants' choice. For each participant and task version, RT1 was independently log transformed and z-scored for the two possible side responses. The analysis strategy was very similar to the above mentioned for choice behavior: first a descriptive analysis; followed by inferential statistics done independently for episodic and semantic tasks; and finally a single model where all data was included and differences between tasks could be evaluated.

To assess the effect of behavioral variables on RT1, we performed a multiple linear regression analysis on log transformed and z-scored RT1, independently for episodic and semantic tasks. Despite some similarities with the approach used for choice behavior, in our predictive model for RT1 the effect of previous reward, transition and reward × transition do not include the interaction with previous first stage choice information. That said, we used as predictors: *Linear*, to model (linearly-increasing) fatigue or evolution in the test by counting the trials in the session; and the predictors of interest R_{t-1}, T_{t-1} and $R_{t-1} \times T_{t-1}$ which quantified the main effect of reward, the main effect of transition and the reward × transition interaction effect, respectively.

To further investigate the influence of task version in reaction times, we then combined both datasets (from episodic and semantic tasks results) into a multiple linear regression, like in choice behavior, adding task version and task order (and their respective interactions with previously mentioned predictors) also as explanatory variable to the model. Here, we removed the linear variable used when regressed RT1 independently for both tasks, to avoid disturbance with task order.

3 Results

A total of 21 subjects were included in this study and completed both episodic and semantic tasks of a two-stage decision task (Daw et al., 2011). Participants completed, on average, 166.86 / 166.28 trials (±16.68 / ±18.23 SD) in episodic / semantic tasks, respectively. Only a very small fraction (average across task versions: 2.14 trials ± 1.47 SD) had to be excluded from the analysis due to prolonged response times. The minimum number of valid trials performed was 131 (subject 20, in episodic task), and was used as a reference to compare the amount of points won during the whole task. While trying to maximize reward, participants showed a small but not statistically significant difference ($t(20)=2.04$, $p= 0.0546$) in the amount of obtained rewards between the episodic (mean total reward= 142.76 ± 20.59 SD) and semantic (mean total reward= 134.29 ± 16.42 SD) tasks.

Choice behavior

When analyzing possible biases in first stage choice, we found a significant preference on both episodic and semantic tasks for picture A ($t(20/20)=3.5688$, 4.3947, $p=0.0019/2.7963e-04$ for episodic/semantic respectively) (in episodic version it corresponds to the 'fractal with pink patterns', and in semantic, the 'tree'). These preferences are probably explained by a somewhat better reward matrix for picture D that belongs to state 2a (the common transition of picture A). For the first 150 trials, the cumulative reward of each second stage picture was: C=161; D=283; E=221; F=209. Regarding the side preferences on first stage, no significant results were found (all $p > 0.05$).

To assess MF and MB RL behavior we explored how previous trial's reward level (high, medium, and low) and previous transition type (common or rare) affected current first stage choice. Figure 4 depicts the probability of repeating first stage choice across participants as a function of previous trials' reward and transition, averaged across the sample. The adaptive switch in face of rare transitions, characteristic of MB behavior, was present in both episodic and semantic tasks. Subjects were indeed much more likely to repeat the same first stage choice if a high reward was obtained through a common transition than when obtained following a rare transition. The opposite choice pattern (i.e. more likely to repeat first stage choice after rare than common transition) was seen following either a medium or a low reward.

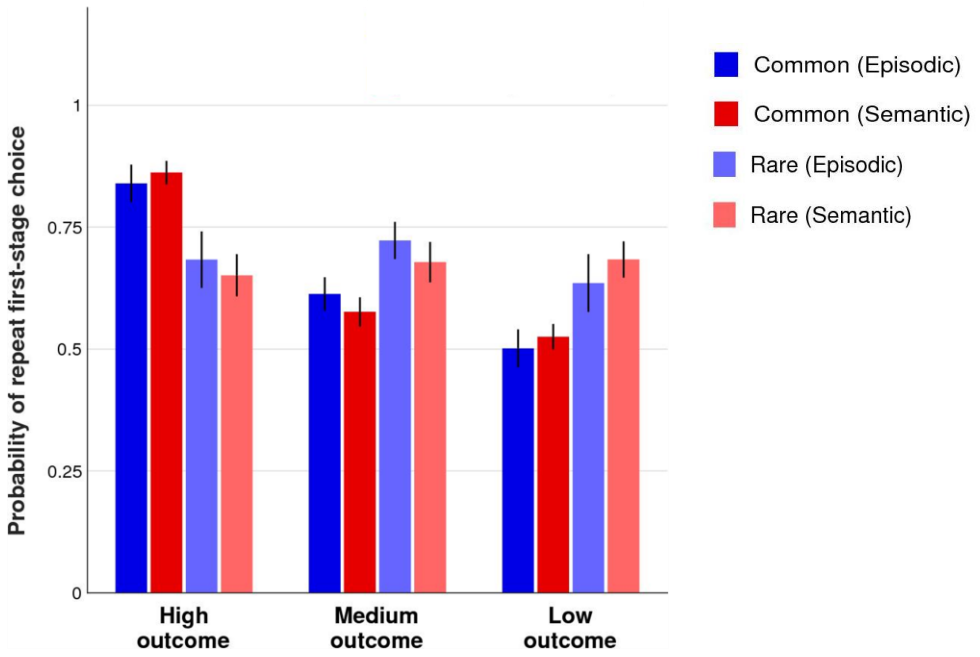


Figure 4: Probability of repeat first stage choice (or "stay probability") as a function of previous trial reward level and previous transition type (common vs rare). Error bars depict SEM. In both tasks we can see the adaptive switch in first-stage choice following a high reward obtained through a rare transition, and the opposite for medium and low outcomes

To quantify further the influence of MF and MB RL on choice behavior, we first performed the logistic analysis independently for episodic and semantic task (aiming to predict first stage choice) using as explanatory variables the events from previous trial (t-1). Instead of creating a variable that codifies choice as 'stay' or 'switch' (according to whether or not a repetition of first stage choice occurs), we use as dependent variable the first stage choice at trial t, and use as predictors not only previous reward and previous transition type, but also previous first stage choice. This in an equivalent approach, however less 'hypothesis-driven' that allows capturing eventual unexpected interactions.

As can be seen in figure 5 (and table S1), we confirmed that at the group level a combination of MF and MB strategies was present on both versions of the task. There was a significant main effect of previous reward on choice (see R × C, MF indicator) (beta (SE) = 2.025 (0.62) / 1.980 (0.34); CI's 95% [0.8015 3.2484] / [1.3228 2.6369]) for episodic/semantic respectively), as well as the weight of previous reward and transition interaction (see R × C × T, MB indicator). We also observed a previously reported (Daw et al., 2011; Feher da Silva & Hare, 2019b) perseveration effect showing the propensity for repeating previous trial's first stage action, regardless the outcome or the transition type (see C) (beta= 2.056 (0.09) / 1.790 (0.25); CI's 95% [1.2926 2.2868] / [1.2926 2.2868]) for episodic/semantic respectively) (see C), and also associated with the preference for picture A that participants showed.

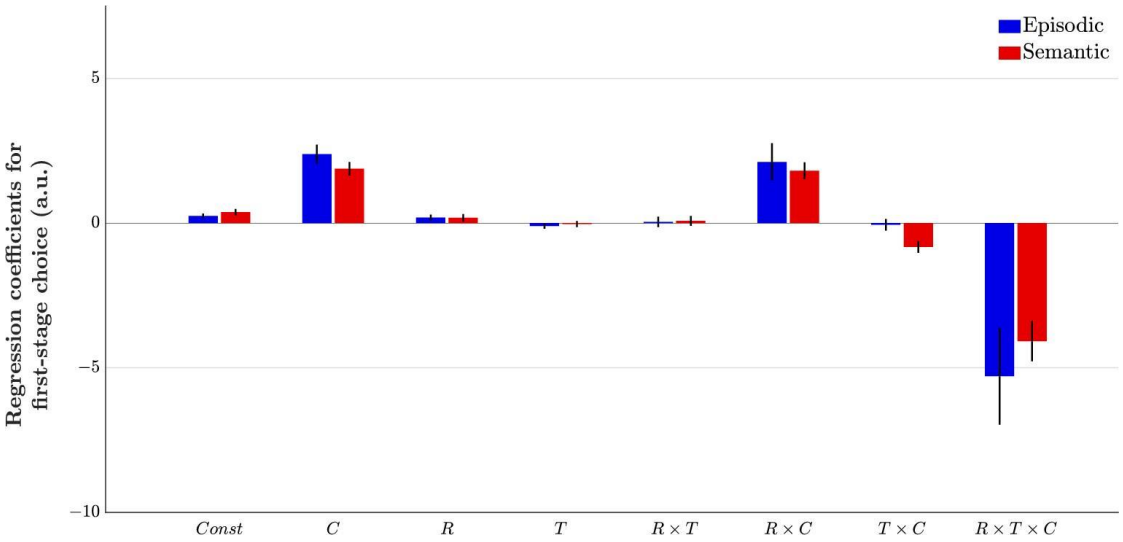
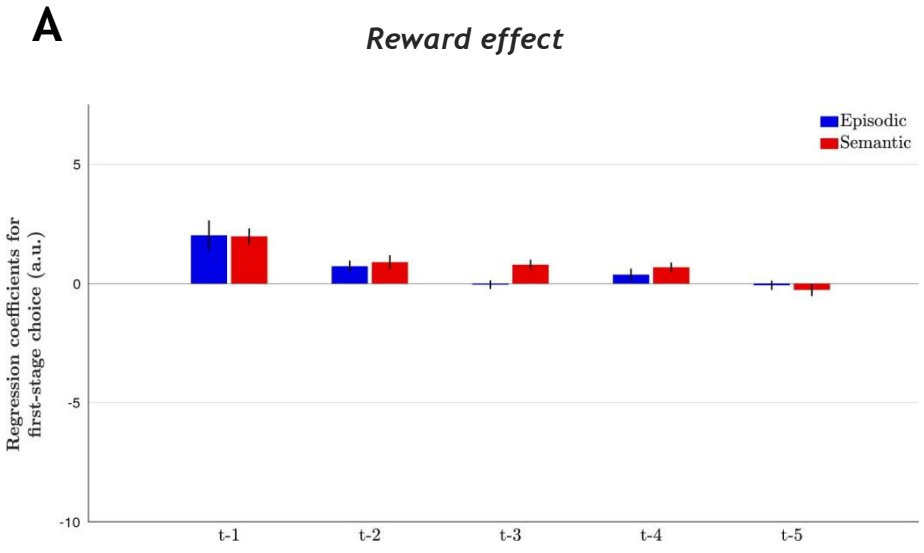


Figure 5: Logistic regression results on first stage choice with the predictors: previous choice (C), reward (R), transition (T), as well as their double and triple interactions, for both episodic and semantic versions. Error bars depict SEM. As expected, evidence for both model-free (R x C) and model-based (R x T x C) influences on choice behavior were observed at group level, for both episodic and semantic tasks. A significant effect of C was also found corresponding to a perseveration effect.

Comparing task versions there seem to be no significant differences between the regression coefficients. Actually, contrary to our hypothesis the tendency seems to be a greater MF and MB influences on behavior in the episodic version vs. semantic (this is explored below). Both MF and MB markers had individual variability in their degree but this variability was bigger in the episodic task than in semantic and the difference was even greater to the reward x transition effect.

According to both MB and MF RL, the effects of trials in the further past tail off, typically exponentially (Sutton & Barto, 2015). So, we included into our model information up to five trials back to evaluate such evolution and figure 6 shows those results for MF and MB indicators (see also table S2).



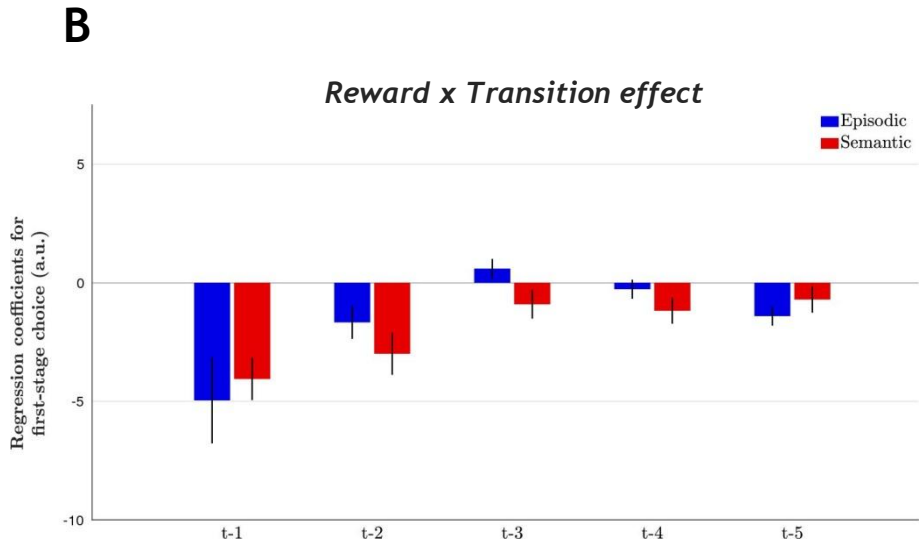


Figure 6: Logistic regression results on first stage choice for the main effect of reward (A) and for the interaction of reward and transition (B) from the five previous trials. Error bars depict SEM. Exponentially decay of both MF and MB effects of trials in the further past.

We found that the contributions to first stage choice from both reward (Fig. 6A) and combined reward × transition (Fig. 6B) reduced across five trials into the past in a way consistent with an exponential decay fit. Despite this decay, these MB and MF RL effects on current choice were present in each of the five trials into the past (Table S1).

We further examined the relationship between task versions and RL task performance merging data from both episodic and semantic task version into a single model, and adding as a predictors to the regression, task version and the order by which each participant performed the two versions of the task (to capture an eventual expertise effect). Here, task version positive interactions with R x C or R x T x C and no interactions with task order would indicate greater MF or MB behavior (respectively) for subjects performing the episodic task, independently of the order of task administration. When adding these two predictors to our regression, the MF (beta (SE) = 1.945 (0.28); CI's 95% [1.3928 2.4977]) effect was still significant, but the weight of the MB was significantly greater (beta (SE) = -3.862 (0.88); CI's 95% [-5.5937 -2.1314]). We confirm our descriptive analysis, no interactions between task version and MF or MB behavioral signatures were found (all p> 0.05) (see R x C x V and R x T x C x V Table S2).

Overall, this logistic analysis indicated that 1) both MF and MB RL strategies coexist, 2) participants showed greater reliance on MB RL and, 3) task version didn't influenced significantly choice behavior in terms of MF and MB signatures.

Reaction times

RTs were consistently, but not significantly (t(40)=-0.68, p =0.5024) faster in the episodic task than in semantic task (episodic: mean ± SD = 1058.4 ± 218.87 ms for first stage and 1163.1 ± 211.33 ms for second stage; semantic: 1104.7 ± 225.24 for first stage and 1196.9 ± 227.3 ms for second –stage). RTs in the second task performed were faster (first: mean ± SD = 1123.9 ± 238.88, second: mean ± SD = 1039.2 ± 197.32), however task order wasn't a significant predictor of first stage RTs (beta (SE) = -0.002 (0.01); CI's 95% [-0.026 0.022]). Faster reaction times in the second task are expected in case of habituation and fatigue absence. In regard to the order by which the tasks were performed (if first episodic or first semantic), no significant differences were found between doing episodic first vs. second, or semantic first vs. second.

Figure 7 depicts the impact of both previous reward and transition on first stage reaction times. In figure 7A no significant differences can be seen, however if we look to the same data in a different way (plotting the difference in RT1 between common and rare transitions) (Fig. 7B), we can see a distinct and opposite pattern on how transition type influences reaction times between tasks. For the episodic version, if a low reward was received after common transition participants' next first choices were faster than if it was through rare, on the other hand for the semantic the same reward level led to much faster reaction times if received through rare transitions compared to common ones. The opposite pattern can be seen for the high outcome: in the episodic, much faster if received through a rare transition, in the semantic, faster if common.

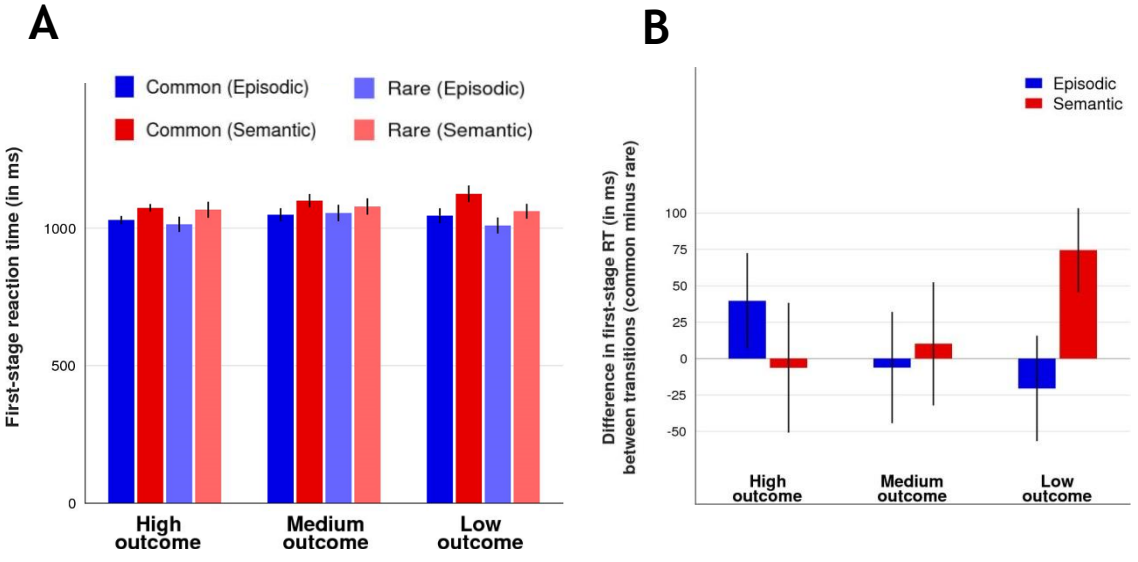


Figure 7: (A) Observed averaged across participants' first stage reaction times (in ms) as a function of previous reward and transition, for both episodic and semantic versions. Error bars depict SEM. (B) Difference between previous common and previous rare trials as a function of reward on the previous trial. Error bars depict SEM. A positive value indicates faster RT1 in common versus rare transitions. As the reward decreases, the RT1 difference between common and rare trials in the episodic version goes down, whereas in the semantic is increasing.

To further assess the impact of our explanatory variables on RT1, we performed a multiple linear regression on log transformed and z-scored RT1, independently for episodic and semantic versions (Table S3). As predictors we used: a linear variable counting trial number (Linear), previous trial reward (R) and transition (T) and, the reward x transition interaction. In contrast to some literature and studies demonstrating that when compared to habitual (MF), goal-directed behavior (MB) is slower (Dezfouli & Balleine, 2013; Keramati et al., 2011; Miranda et al., 2019), we did not find such association. The only significant predictor for RT1 (to both tasks) was the linear variable counting trial numbers (beta (SE)= -0.161 (0.03) /-0.193 (0.03); CI's 95%= [-0.10709 -0.002768]/[-0.22834 -0.09449] for episodic/semantic respectively). In accordance with other studies (Keramati et al., 2011) these results suggest that as task progresses in trials, participants become faster in their first stage choice (despite not statistically significant, the faster RT1 found in the second task performed supports this idea).

We then put together both episodic and semantic tasks results into a single regression model and added task version and task order also as predictors (Table S4). We removed the linear variable to avoid confusion with task order execution. As before, no correlation with MF or MB signatures was found. Neither task version nor order of execution influenced significantly RT1. Only previous transition type was associated significantly with RT1 (beta (SE) = -0.055 (0.03); CI's 95% [-0.10709 -0.002768]); subjects were faster after a rare transition.

4 Discussion

The importance of memory to cognitive planning is becoming increasingly clear (Gershman, 2017; Irish et al., 2016; Klein et al., 2002, 2010; Schacter et al., 2007; Suddendorf & Corballis, 2017). The dominant view defends that we plan based on autobiographical (episodic) information (Daw et al., 2014; Doll et al., 2014; Schacter et al., 2012; Vikbladh et al., 2016). Comprehensive visions extended the study of future-oriented mental activities to other types of memory and the importance of semantic content in prospective activities, like imagining future non-personal events, has already been demonstrated (Irish et al., 2016; Klein et al., 2002). However, these studies did not inform us about the relationship between memory and planning behavior. In the current study we aimed to further ascertain the influence of semantic memory when planned decisions are made. To this end, we used a reinforcement learning (RL) task, considered a gold standard measure for the contribution of goal-directed (model-based) and habitual (model-free) systems to choice. We compared choice behavior on an 'episodic task' against a 'semantic task' in order to see if there are any significant differences on planning when a semantic 'clue' is added.

We successfully reproduced the task and prior reported results (Daw et al., 2011; Eppinger et al., 2013; Feher da Silva & Hare, 2019a; Otto et al., 2013) showing that human behavior is driven by both MF and MB processes operating in parallel. Our logistic analysis on choice, for both episodic and semantic tasks revealed that the transition structure of the task (of MB importance) and the reward history (of MF and MB importance) significantly influenced the decision-making process (see R x C and R x T x C in Fig. 4 and Table S1).

Previous studies have found a speed/accuracy trade-off between MF and MB computations, with the former being faster (Keramati et al., 2011; Miranda et al., 2019). In our study, we did not find clear evidence to support an influence of MB choice on response time. Our linear regression analysis in first stage reaction time (RT1) did not find an influence of reward nor transition (although a stronger, though not significant effect, was observed for reward only, reflecting the influence of MF behavior). However, by showing that as task progresses in trials, subjects time response decreased without detriment of MB performance, our data supports the view that with training MB behavior can become automatic (Keramati et al., 2011) (Table S3). This tendency of faster RT1 throughout the task, independent of the order of execution, was also a good indicator of fatigue absence, demonstrating that performing each task during twenty minutes, in the same session (only with a brief interval between them) was feasible. Future work could consider extending the task' length, given that previous studies calculations point to ~350 trials for when MB behavior is expected to attain reasonable internal reliability (Shahar et al., 2019).

Contrary to the hypothesis that a semantic clue could potentiate the knowledge of the state-transition function of the task, we did not find evidence of enhanced MB

behavior in the semantic task (when compared to the episodic version). In choice analysis, task version (V) did not interact significantly with either MF (see R x C x V regressor in Table S2) or MB (see R x T x C x V regressor in Table S2) signatures. In fact, one could view episodic and semantic versions as being just one simple task with different set of pictures. As evidence in support of an independent role of semantic content for cognitive planning is vast, we believe that factors related to our task design may well explain our negative findings. We will then focus our discussion on potential reasons and limitations.

Regarding task structure, it was presumed at the start that the semantic associations chosen were good examples of easy and widely known semantic relationships. In fact, the brief post-task questionnaire revealed that all participants (n=21) correctly indentified the specific relationships between first and second stages pictures for episodic and semantic tasks. However the level of uncertainty about the state-transition structure was much bigger for the episodic task. While not ensuring that the chosen set reflects clear semantic associations, it is a good indicator. Some participants when referring to the semantic task directly stated: 'it was easy because I just thought - fruits come from trees and we use means of transport to discover the globe'. Although work can be done to improve the behavioral tasks (and the semantic association), this is unlikely to explain the results.

During observation of participants' task execution we realized that, in some sequences of trials, rare transitions were distributed consecutively. In such occasions, participants seemed confused and sometimes questioned if it was possible that the state-transition structure had changed. This confusion, caused by randomly assigned transition types (stochastically defined as 0.7 probability of common and 0.3 rare), revealed a subtle difference between tasks. In semantic versions this perturbation had less effect on subsequent trials than in episodic version, suggesting that the 'semantic clue' helped participants to solve this mismatch with expectations. It was also mentioned by some participants that, when for some reason they became distracted, it was easier to 'return into the game' on the semantic task. Such anecdotal report could be viewed as 'semantic help' (it might be a sign of a crucial role of semantic knowledge in planning) and this facilitation after consecutive rare transitions should be addressed in future work.

Other possible explanation is that a source of confusion might reside in the way we classify and infer about the existence or not of MB behavior (Akam et al., 2015; Feher da Silva & Hare, 2019b). According to the described analysis strategy, a MB learner is identified when a significant prediction effect on choice exists in the interaction of previous reward and transition type. In particular, the indicators of MB behavior are:

1. Switch first stage choice when previous reward was low and obtained through a common transition;
2. Switch first stage choice when previous reward was high and obtain through a rare transition.

With respect to expected behavior 1, one can ask: why keeping a first stage choice when previous reward was low and obtained through a common transition, cannot be

considered model-based? By its definition (deliberative and map-based) there are some circumstances when it fits. If we think that each second stage state is composed by two available options, after a low reward a subject that repeats first stage choice but not the second, might just be exploring. Here the participant takes into consideration transition structure but in a deliberative way, repeats the non-rewarded action in order to explore reward in the other second stage option. Creating a variable that represents these 'exploratory' trials and add it to behavior analysis might be interesting and useful in terms of studying planning behavior.

In turn, indicator 2 can certainly be expected from a MB agent when previous trials rewards were minor that current high reward, however this condition (rare transition & high reward) can occur after a common trial also with high reward. Sometimes participants showed what I like to call "loyalty to hunch": whenever subjects found a second stage picture with high reward they persist with that choice until a reward drift occurs. For instance, imagine performing the semantic version (see Fig. 2B to follow the reasoning): in a certain moment of the task you are betting in 'globe' as a first stage choice because you want to reach the 'train', that is being reward with 3 coins, and you know that that's the option more likely to take you there. However, through a rare transition, you are surprisingly taken to the state where the options are 'orange' and 'banana'; despite your original hunch was 'train', you have to choose one of those. Luckily you decide for the 'orange', also being highly rewarded. What would you do next? Be 'loyal' to your original hunch 'train' or changed to 'orange' because you just realize that is also being highly rewarded? Both decisions can reflect cognitive planning: on one hand you change your choice because you discovered other stimuli with equally good reward; or as both options are equal in terms of outcome, you decide to keep your previous choice. Expected behavior 2 does not take into consideration such situations. In future work with the task, reward matrix design must ensure that there are no two stimuli both associated with high reward simultaneously.

The significant prediction effect of transition found in RT1 multiple regression (faster RT1 when previous trial transition was rare) might also be related to the above mentioned "loyalty to hunch" factor. When we put both task results in the same regression (adding as independent variables task version and order), the only significant predictor was transition i.e. participants were faster following rare transitions, independently of the amount of reward received. This result seemed a bit counter intuitive, since a rare transition is expected to surprise participants and increase subsequent reaction time. However, in cases when participants are already exploiting a particular option, it's not strange that they 'ignore' this rare transition, and quickly decide to repeat first stage option to get to the option desired. In such cases, we believe that the decision is made at the moment of the rare transition occurrence, explaining faster RT1. The transition effect was not present when we regressed RT1 for episodic and semantic tasks independently (Table S3) but a trend was observed. After joining both tasks data into a single model, the effect became significant, suggesting a statistical power issue (Table S4).

Although the issues previously raised deserve further attention, it might be the case that our results derive from a more elementary question related with the sensitivity

of the task itself: “Could it be possible that the potential role of our semantic clue is being masked by the preserved abilities of our healthy sample of adults?” Thinking in both episodic and semantic tasks, one can realize that what’s episodic in one is also in the other. A subject to ideally perform the task must remember: stimuli set being presented; state-transition structure; choices made in first and second stages; and previous reward (given the second stage choice). In our semantic task, the difference is that we add a ‘semantic helper’ to the state-transition structure, but in case of preserved cognitive abilities, this help is probably unnecessary.

Remembering a well established semantic relationship is probably cognitively less demanding than keeping in mind a new-formed one. However, establishing an association between two sets of stimuli is not a very hard task to perform and so, we suppose that for our healthy subjects, the existence of a semantic clue is unnecessary. Nonetheless, the uncertainty revealed by participants in post-task questionnaire when asked about state-transition structure, supports the idea that in the semantic task the association between first and second stages was stronger and more reliable than in the episodic one. The greater variability found for episodic tasks in both MF and MB signatures, also points to the same direction. Despite we did not found any differences in MF and MB strategies, the semantic task was perceived by most (n=18) participants as easier (independently of the order of administration). In addition, the different reaction time patterns, found for the different rewards and transitions between episodic and semantic tasks suggest that distinct processes may be being used.

To investigate our previous supposition of low sensibility of the task (and possibly demonstrate our hypothesis that semantic content is also part of cognitive planning), future work may consider to increase task difficulty (by decreasing maximal time response limit and/or when instructing, giving less emphasis to the importance of the transition structure), and apply this protocol to clinical populations particularly to early-stage Alzheimer’s disease (AD) patients. As mentioned before (in the Introduction), the neuropsychological profile of such patients commonly reveals specific deficits in the episodic domain, but a relatively preserved ability to access semantic content. ‘Executive’ deficits, such as working memory[4], are also commonly reported (Weintraub et al., 2012), and studies evaluating AD patients in the ToL task, also suggest planning ability impairments. Comparing AD patients’ performance in both episodic and semantic tasks seems promising to our main goal. If semantic memory is really helpful in goal directed behavior as we hypothesize, a subject with impairments in both episodic and working memory (and also possibly[5] deficits in planning), but preserved semantic knowledge, would probably rely on such information to inform his decision. With this, we expect that if we were right and our results were being masked by other preserved abilities, in this group with scarce

[4] “Working memory” refers to a processing system whereby information that is the immediate focus of attention is temporarily held in a limited capacity, language- or visually-based, immediate memory buffer while being manipulated by a “central executive” (Baddeley 2003).

[5] Besides the issues raised (in the introduction) about inferring planning impairments through ToL task performance, there is also evidence that ToL task execution depends not only but also on working memory (Morris & Ward, 2005(Morris & Ward, 2005) and so reported ‘planning impairments’ might reflect that.

cognitive resources, an evident distinction between MF and MB strategies will be present. In particular, we anticipate that AD patients will present an enhanced MB behavior in the semantic task (comparative to episodic), but worse than healthy matched controls

Future work might further explore the effect of training in this task. Prior work showing that training can enable model-based reasoning even when executive resources are devoted to another task (Economides et al., 2015), and the evidence we found of faster reaction times as tasks progresses in trials, favors automation theories that defend a some kind of 'habitual model-based behavior', where goal directed reasoning can occur with less cognitive demands. Evidence from non-human primates with extensive training on this task shows a 'boost' on MB behavior likely due to such overtraining (Miranda et al., 2019). Studying how participants' performance evolve through several sessions of training this task, in terms of reaction time analysis and choice behavior, seems relevant to the study of decision-making and to illuminate the contrast between MF and MB predictions.

5 Concluding comments

Our results did not prove our hypothesis of enhanced model-based behavior in the semantic task. We considered and discussed several modifications that deserve further attention and may improve behavioral task analysis. However, a possible explanation for such non significant results is related to an eventual low sensitivity of the task. We believe that episodic preserved abilities of our healthy sample of adults can probably be masking the eventual role that semantic memory can have on planning behavior.

Future work may consider addressing this question by applying these two versions of a two-stage decision task to early-stages patients of Alzheimer's disease and study if, in case of episodic impairment (and preserved semantic memory) the semantic clue works as an alternative to episodic content in cognitive planning or not.

To better understand subjects' strategy evolution, and the shift from goal-directed (model-based) to habitual (model-free) control, future work might profitably study the effect of training in this task. Investigating choice behavior and time responses modifications, across several sessions, seems relevant to comprehend strategies and clarify this apparent conflict between classical descriptions in decision making literature of goal-directed and deliberative as being synonyms and automation theories which argue that model-based behavior can become habitual.

The two-stage decision task is a powerful tool for studying behavior and cognition (particularly decision making) through a quantitative and computational approach. However, careful considerations are required to all features and slight modifications since they will probably influenced the balance between MF/MB strategies applied by subjects.

6 Appendices

Episodic or not?

This work was developed with the intent of better understand cognitive processes underlying planning ability, and based on growing evidence that past information storage is intimately related to prospective activities, such as planning. Since prior studies on the topic, mostly infer conclusions about the relationship between planning and previous lived experiences, our purpose was (and still is) demonstrating that non-personal and culturally shared knowledge also has a role in this goal-directed behavior. One possible explanation of our results was that, maybe the importance of semantic information was being masked by subjects preserved memory abilities. To note that, I used the words 'preserved memory abilities' instead of 'episodic memory abilities' deliberately. We called 'episodic task' due to the autobiographical reference required to perform the task: participants must recall previous personal experience (played trials before) to inform decision; previous studies using the same task, referred to as an episodic content.

The memory taxonomy used (Squire, 2004) is based on time persistence, and so a primary distinction is made between short (STM) and long (LTM) term memory, where the later includes episodic and semantic subtypes. Since it is considered that episodic and semantic subtypes belong to LTM, an important set of questions arise: Chronologically speaking, when does it start LTM? Is it in a scale of minutes, hours, days? Can it be considered that our 'episodic' version, in this format of one single evaluation session, includes some kind of long-term memory? Maybe after 10, 50, 100 trials? Or does it fit well with the concept of working memory? Despite empirical evidence showing that STM and LTM are not as distinct as it was initially thought, and some recent models on memory ('knowledge') systems transcend this dichotomy (Nadel & Hardt, 2011), we believed that, in terms of conceptual framework, might be relevant.

If such a dichotomy between STM and LTM exists, in a first moment of evaluation, the episodic task (version) might better be considered as a working memory (WM) performance, whereas the semantic one besides that also includes a long term component. As suggested in future perspectives, if we administrate these tasks in several sessions to the same participants, to study the effect of expertise, then we can clearly assume that an episodic LTM memory would be present. The fact that our 'episodic task', might not be episodic (according to the dominant taxonomy) does not nullify the potential of performing this work in early-stages AD patients (impairments in WM are also common). Meanwhile, maybe it's more precise if we assume that both tasks are associated with a present *lived* experience, and only semantic possesses a past *known* content. This dichotomy between lived and known experience is intrinsically related to episodic and semantic information respectively (Klein et al., 2002).

Supplementary tables

1. Table S1: Multiple logistic regression results for predictors of first stage choice up to five trials back, independently for the episodic and semantic task versions.

Predictors ‡	Mixed-effects *							
	Episodic				Semantic			
	Estimate	SE	CI's 95%		Estimate	SE	CI's 95%	
Constant	0.154	0.09	-0.030	0.339	0.328	0.12	0.084	0.572
C_{t-1}	2.056	0.33	1.408	2.704	1.790	0.25	1.293	2.287
C_{t-2}	0.280	0.10	0.094	0.467	-0.124	0.16	-0.429	0.181
C_{t-3}	0.116	0.15	-0.179	0.411	0.090	0.12	-0.140	0.320
C_{t-4}	-0.026	0.12	-0.261	0.208	0.038	0.13	-0.215	0.292
C_{t-5}	0.031	0.13	-0.220	0.282	-0.288	0.13	-0.538	-0.039
R_{t-1}	0.158	0.10	-0.033	0.348	0.192	0.17	-0.143	0.526
R_{t-2}	0.126	0.10	-0.064	0.317	-0.003	0.11	-0.220	0.215
R_{t-3}	0.069	0.09	-0.108	0.246	0.038	0.16	-0.275	0.351
R_{t-4}	-0.095	0.09	-0.277	0.088	0.226	0.11	0.003	0.450
R_{t-5}	0.007	0.12	-0.219	0.232	-0.151	0.13	-0.412	0.109
T_{t-1}	-0.117	0.10	-0.304	0.071	0.015	0.15	-0.277	0.307
T_{t-2}	-0.089	0.11	-0.305	0.127	0.222	0.15	-0.062	0.507
T_{t-3}	-0.027	0.12	-0.268	0.214	0.023	0.10	-0.170	0.216
T_{t-4}	-0.142	0.09	-0.321	0.036	0.016	0.10	-0.178	0.209
T_{t-5}	-0.138	0.10	-0.325	0.048	0.251	0.14	-0.021	0.523
$R_{t-1} \times T_{t-1}$	0.177	0.19	-0.201	0.554	0.053	0.23	-0.398	0.505
$R_{t-2} \times T_{t-2}$	-0.161	0.18	-0.523	0.201	-0.126	0.25	-0.617	0.364
$R_{t-3} \times T_{t-3}$	-0.063	0.19	-0.441	0.315	-0.195	0.24	-0.660	0.269
$R_{t-4} \times T_{t-4}$	-0.102	0.18	-0.463	0.258	-0.175	0.25	-0.656	0.305
$R_{t-5} \times T_{t-5}$	0.443	0.19	0.076	0.810	-0.073	0.22	-0.511	0.364
$R_{t-1} \times C_{t-1}$	2.025	0.62	0.802	3.248	1.980	0.34	1.323	2.637
$R_{t-2} \times C_{t-2}$	0.724	0.24	0.252	1.196	0.898	0.29	0.336	1.461
$R_{t-3} \times C_{t-3}$	-0.046	0.18	-0.402	0.310	0.788	0.21	0.384	1.192
$R_{t-4} \times C_{t-4}$	0.373	0.25	-0.119	0.864	0.680	0.20	0.282	1.078
$R_{t-5} \times C_{t-5}$	-0.070	0.19	-0.450	0.309	-0.265	0.26	-0.767	0.237
$T_{t-1} \times C_{t-1}$	-0.266	0.21	-0.679	0.147	-0.841	0.32	-1.474	-0.208
$T_{t-2} \times C_{t-2}$	-0.015	0.19	-0.383	0.352	0.181	0.24	-0.289	0.651
$T_{t-3} \times C_{t-3}$	0.419	0.30	-0.175	1.013	1.070	0.21	0.6621	1.479
$T_{t-4} \times C_{t-4}$	0.383	0.18	0.023	0.743	0.365	0.25	-0.119	0.849
$T_{t-5} \times C_{t-5}$	0.157	0.30	-0.423	0.736	-0.223	0.26	-0.727	0.282
$R_{t-1} \times T_{t-1} \times C_{t-1}$	-4.957	1.81	-8.513	-1.400	-4.056	0.89	-5.796	-2.316
$R_{t-2} \times T_{t-2} \times C_{t-2}$	-1.668	0.69	-3.026	-0.310	-2.992	0.89	-4.739	-1.244
$R_{t-3} \times T_{t-3} \times C_{t-3}$	0.593	0.41	-0.209	1.396	-0.908	0.61	-2.095	0.279
$R_{t-4} \times T_{t-4} \times C_{t-4}$	-0.273	0.40	-1.058	0.511	-1.178	0.55	-2.251	-0.105
$R_{t-5} \times T_{t-5} \times C_{t-5}$	-1.405	0.41	-2.200	-0.610	-0.705	0.56	-1.794	0.384

* Values of mixed-effects results are the regression coefficient.

‡ For the given trial t, the variables used were: The predictors used were: Constant captured any potential first stage picture bias; C (first stage choice; 1=pic A, 0=pic B) modeled a potential independent tendency to stick with the same option from trial to trial; R (outcome level; assumed as continuous and with low=1, medium=2, high=3); T (transition; rare=1, common=0); and their interactions. All predictors were mean centered and continuous variables were also scaled by dividing them by two SD (adjustments made before the computation of the interaction terms).

2. Table S2: Multiple logistic regression results for predictors of first stage choice, including task version and order of administration.

Predictors ‡	Mixed-effects *			
	Estimate	SE	CI's 95%	
Constant	0.277	0.06	0.1554	0.3977
C_{t-1}	1.967	0.24	1.4877	2.4457
R_{t-1}	0.154	0.09	-0.0142	0.3224
T_{t-1}	-0.018	0.06	-0.1366	0.1009
V_{t-1}	-0.054	0.09	-0.2354	0.1266
O_{t-1}	-0.078	0.09	-0.2595	0.1025
$R_{t-1} \times C_{t-1}$	1.945	0.28	1.3928	2.4977
$T_{t-1} \times C_{t-1}$	-0.388	0.13	-0.6446	-0.1317
$C_{t-1} \times V_{t-1}$	0.206	0.19	-0.1694	0.5821
$C_{t-1} \times O_{t-1}$	-0.182	0.19	-0.5581	0.1934
$R_{t-1} \times T_{t-1}$	0.108	0.12	-0.1260	0.3425
$R_{t-1} \times V_{t-1}$	0.164	0.13	-0.0941	0.4225
$R_{t-1} \times O_{t-1}$	-0.0488	0.13	-0.3071	0.2096
$T_{t-1} \times V_{t-1}$	-0.246	0.13	-0.4965	0.0038
$T_{t-1} \times O_{t-1}$	-0.204	0.13	-0.4537	0.0465
$V_{t-1} \times O_{t-1}$	-0.137	0.25	-0.6215	0.3477
$R_{t-1} \times T_{t-1} \times C_{t-1}$	-3.862	0.88	-5.5937	-2.1314
$R_{t-1} \times C_{t-1} \times V_{t-1}$	0.088	0.43	-0.7626	0.9378
$R_{t-1} \times C_{t-1} \times O_{t-1}$	0.511	0.43	-0.3393	1.3611
$R_{t-1} \times T_{t-1} \times V_{t-1}$	-0.202	0.24	-0.6788	0.2739
$R_{t-1} \times T_{t-1} \times O_{t-1}$	-0.078	0.24	-0.5539	0.3988
$T_{t-1} \times V_{t-1} \times O_{t-1}$	0.586	0.24	0.1110	1.0610
$T_{t-1} \times C_{t-1} \times V_{t-1}$	0.359	0.33	-0.2957	1.0135
$T_{t-1} \times C_{t-1} \times O_{t-1}$	0.378	0.33	-0.2761	1.0331
$R_{t-1} \times V_{t-1} \times O_{t-1}$	-0.139	0.34	-0.8120	0.5342
$C_{t-1} \times V_{t-1} \times O_{t-1}$	-0.579	0.98	-2.4945	1.3373
$R_{t-1} \times T_{t-1} \times V_{t-1} \times O_{t-1}$	0.526	0.48	-0.4106	1.4632
$T_{t-1} \times C_{t-1} \times V_{t-1} \times O_{t-1}$	-0.157	0.52	-1.1830	0.8684
$R_{t-1} \times C_{t-1} \times V_{t-1} \times O_{t-1}$	-0.436	1.13	-2.6460	1.7739
$R_{t-1} \times T_{t-1} \times C_{t-1} \times O_{t-1}$	1.530	1.10	-0.6065	3.6661
$R_{t-1} \times T_{t-1} \times C_{t-1} \times V_{t-1}$	0.482	1.09	-1.6548	2.6178
$R_{t-1} \times T_{t-1} \times C_{t-1} \times V_{t-1} \times O_{t-1}$	1.074	3.53	-5.8503	7.9991

* Values of mixed-effects results are the regression coefficient.
‡ For the given trial t, the predictors used were: Constant captured any potential first stage picture bias; C (first stage choice; 1=pic A, 0=pic B) modeled a potential independent tendency to stick with the same option from trial to trial; R (outcome level; assumed as continuous and with low=1, medium=2, high=3); T (transition; rare=1, common=0); V (task version; episodic=1, semantic=0); O (task order; first task=1, second task=0); and their interactions. All predictors were mean centered and continuous variables were also scaled by dividing them by two SD (adjustments made before the computation of the interaction terms).

3. Table S3: Multiple linear regression results for predictors of first stage reaction time, independently for episodic and semantic versions.

	Mixed-effects *							
	Episodic				Semantic			
Predictors ‡	Estimate	SE	CI's 95%		Estimate	SE	CI's 95%	
Constant	0.001	0.02	-0.033	0.035	0.003	0.02	-0.031	0.037
Linear	-0.161	0.03	-0.228	-0.095	-0.193	0.03	-0.262	-0.130
R_{t-1}	0.028	0.03	-0.040	0.095	-0.014	0.03	-0.083	0.054
T_{t-1}	-0.01	0.04	-0.124	0.022	-0.046	0.04	-0.120	0.028
$R_{t-1} \times T_{t-1}$	-0.001	0.07	-0.144	0.141	0.024	0.07	-0.119	0.167

* Values of mixed-effects results are the regression coefficient
‡ For the given trial t, the predictors used were: Constant term; Linear was used to model (linearly-increasing) fatigue or evolution by counting the trials in the session; R (outcome level; assumed as continuous and with low=1, medium=2, high=3); T (transition; rare=1, common=0); and R x T interaction. All predictors were mean centered and continuous variables were also scaled by dividing them by two SD (adjustments made before the computation of the interaction terms).

4. Table S4: Multiple linear regression results for predictors of first stage reaction time, including task version and order of administration.

Predictors	Mixed-effects*			
	Estimate	SE	CI's 95%	
Constant	0.001	0.01	-0.023	0.026
R_{t-1}	0.014	0.02	-0.035	0.062
T_{t-1}	-0.055	0.03	-0.107	-0.003
V_{t-1}	-0.001	0.01	-0.025	0.023
O_{t-1}	-0.002	0.01	-0.026	0.022
$T_{t-1} \times V_{t-1}$	-0.002	0.03	-0.054	0.005
$T_{t-1} \times O_{t-1}$	0.018	0.03	-0.034	0.070
$V_{t-1} \times O_{t-1}$	0.001	0.01	-0.023	0.026
$T_{t-1} \times R_{t-1}$	0.022	0.05	-0.079	0.124
$V_{t-1} \times R_{t-1}$	0.019	0.02	-0.029	0.067
$O_{t-1} \times R_{t-1}$	0.003	0.02	-0.045	0.051
$T_{t-1} \times V_{t-1} \times O_{t-1}$	0.022	0.03	-0.030	0.074
$T_{t-1} \times V_{t-1} \times R_{t-1}$	-0.014	0.05	-0.115	0.081
$T_{t-1} \times O_{t-1} \times R_{t-1}$	-0.031	0.05	-0.133	0.070
$V_{t-1} \times O_{t-1} \times R_{t-1}$	0.010	0.02	-0.038	0.059
$T_{t-1} \times V_{t-1} \times O_{t-1} \times R_{t-1}$	0.023	0.05	-0.076	0.127

* Values of mixed-effects results are the regression coefficient
‡ For the given trial t, the predictors used were: Constant term; R (outcome level; assumed as continuous and with low=1, medium=2, high=3); T (transition; rare=1, common=0); V (task version; episodic=1, semantic=0); O (task order; first task=1, second task=0); and their interactions. All predictors were mean centered and continuous variables were also scaled by dividing them by two SD (adjustments made before the computation of the interaction terms).

Task instructions

The two-stage task instructions stated that the purpose was to study participants' ability to plan their actions. They would have to play two versions of a sequential decision task where the main goal was to collect the biggest reward possible. Each task would be performed during 20 minutes with an interval between, after a period of practice playing the exact same task but with a different set of stimuli.

The task consisted in multiple trials composed by two stages (two moments of decision) in each of which an image (from two possible) had to be chosen by touching it on the screen where the images were displayed. After making the second stage choice the associated reward was presented (at this moment of instruction figure S1A was presented). The reward of each second stage image could assume one of three levels: low —red cross, medium — one coin, and high — three coins, that were independent of each other. Participants were instructed that in order to be successful and maximize their rewards there were two main aspects to keep in mind:

1) The value associated to each second stage image vary throughout the task and must be updated i.e. for a certain period the reward remain constant, but then change to any of the three possible levels (here, figure S1B was presented and compared to figure S1A, the reasoning was repeated according to the example).
2) There is a relationship between the first stage choice and the set of images presented at the second stage being that each one image is predominantly associated with a particular set however, more rarely there may be a transition to the other (here, figure S1C was presented and the reasoning was repeated).

Other extra points were also beaded, in particular, the lack of any other relationship beyond the transition structure previously mentioned (to try to lessen the natural trend of creating patterns):

- No relationship between pictures values. Here, some examples reinforcing the idea of independence of rewards were given: "There might be occasions where three images have a high reward, and others where only one is receiving it" or "The fact that one picture from a particular set in being highly rewarded doesn't mean that the other doesn't, or vice-versa.
- Stimuli set presentation was random in terms of position (here figure S1A and S1B were compared).

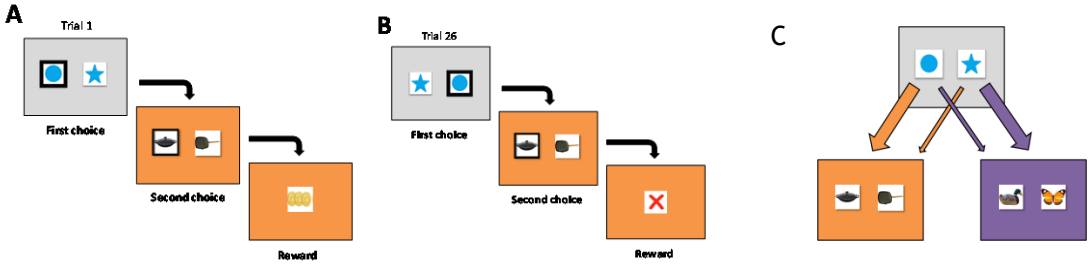


Figure S1: Task instructions images. (A) Example of trial showing the two moments of choice followed by the reward presentation. (B) Example of another trial demonstrating that the order of stimuli presentation is random and that rewards change over time. (C) Transition structure of the induction task, used as example for guiding instructions. It shows the predominant relations between images, without explicitly mention de probability.

7 References

Akam, T., Costa, R., & Dayan, P. (2015). Simple Plans or Sophisticated Habits? State, Transition and Learning Interactions in the Two-Step Task. *PLoS Computational Biology*, *11*(12), 1–25. https://doi.org/10.1371/journal.pcbi.1004648

Anderson, V., Jacobs, R., & Anderson, P. J. (2008). *Executive Functions and the Frontal Lobes*.

Baker, C. S., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S. J., & Robbins, T. W. (1996). *Neural systems engaged by planning : a PET study of the Tower of London task*. *34*(6), 515–526.

Balleine, B. W., & Dickinson, A. (1998). *Goal-directed instrumental action : contingency and incentive learning and their cortical substrates*. *37*, 407–419.

Benson, D. F. (1993). *Prefrontal abilities*. *6*, 75–81.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, *5*(5). https://doi.org/10.1371/journal.pone.0010773

Craik, F. I. M. (2006). *Planning and task management in older adults: Cooking breakfast*. *6*, 1236–1249.

Daw, N. D., Dayan, P., & Dayan, P. (2014). *The algorithmic anatomy of model-based evaluation*.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., Raymond, J., Computational, G., & Unit, N. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1007/978-1-4419-1428-6_2344

Daw, N. D., Niv, Y., & Dayan, P. (2005). *Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control*. *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Demetriou, E. A., DeMayo, M. M., & Guastella, A. J. (2019). Executive Function in Autism Spectrum Disorder: History, Theoretical Models, Empirical Findings, and Potential as an Endophenotype. *Frontiers in Psychiatry*, *10*(November), 1–17. https://doi.org/10.3389/fpsyt.2019.00753

Dezfouli, A., & Balleine, B. W. (2013). *Actions , Action Sequences and Habits : Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized*. *9*(12). https://doi.org/10.1371/journal.pcbi.1003364

Diamond, A. (2017). Executive Functions. *ANAE - Approche Neuropsychologique Des Apprentissages Chez l'Enfant*, *29*(146), 13–15. https://doi.org/10.1146/annurev-psych-113011-143750

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–

325. https://doi.org/10.1016/j.neuron.2013.09.007

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, *18*(5), 767–772. https://doi.org/10.1038/nn.3981

Doll, B. B., Shohamy, D., & Daw, N. D. (2014). Multiple memory systems as substrates for multiple decision systems. *NEUROBIOLOGY OF LEARNING AND MEMORY*, *May*. https://doi.org/10.1016/j.nlm.2014.04.014

Economides, M., Kurth-nelson, Z., & Lübbert, A. (2015). *Model-Based Reasoning in Humans Becomes Automatic with Training*. 1–19. https://doi.org/10.1371/journal.pcbi.1004463

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, *7*(7 DEC), 1–14. https://doi.org/10.3389/fnins.2013.00253

Eratne, D., Loi, S. M., Farrand, S., Kelso, W., Velakoulis, D., & Looi, J. C. L. (2018). Alzheimer's disease: clinical update on epidemiology, pathophysiology and diagnosis. *Australasian Psychiatry*, *26*(4), 347–357. https://doi.org/10.1177/1039856218762308

Feher da Silva, C., & Hare, T. A. (2019a). Humans are primarily model-based and not model-free learners in the two-stage task. *BioRxiv*. https://doi.org/10.1101/682922

Feher da Silva, C., & Hare, T. A. (2019b). *Model-free or muddled models in the two-stage task ? June*. https://doi.org/10.1101/682922

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(31), 11778–11783. https://doi.org/10.1073/pnas.0602659103

Furby, L., & Beyth-Marom, R. (1990). *Risk taking in adolescence: a decision-making perspective*.

Geffner, H. (2013). *Computational models of planning*. *4*(August). https://doi.org/10.1002/wcs.1233

Gelman, A. (2008). *Scaling regression inputs by dividing by two standard deviations*. *October 2007*, 2865–2873. https://doi.org/10.1002/sim

Gershman, S. J. (2017). Predicting the Past, Remembering the Future. *Curr Opin Behav Sci*, 7–13. https://doi.org/10.1016/j.cobeha.2017.05.025.Predicting

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*(1), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. In *Neuropsychologia* (Vol. 33, Issue 5, pp. 623–642). https://doi.org/10.1016/0028-3932(95)90866-P

Grafman, J., Litvan, I., Massaquoi, S., Stewart, M., Sirigu, A., & Hallett, M. (1992).

Cognitive planning deficit in patients with cerebellar atrophy. *Neurology*, *42*(8), 1493–1496. https://doi.org/10.1212/wnl.42.8.1493

Harlow, J. M. (1868). *Recovery from the passage of an iron bar through the head*. https://doi.org/10.1177/0957154X9300401407

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). *Patients with hippocampal amnesia cannot imagine new experiences*. 1–6.

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *Lancet Neurology*, *6*(11), 1004–1014. https://doi.org/10.1016/S1474-4422(07)70266-1

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: Evidence from semantic dementia. *Brain*, *135*(7), 2178–2191. https://doi.org/10.1093/brain/aws119

Irish, M., Eyre, N., Dermody, N., O'Callaghan, C., Hodges, J. R., Hornberger, M., & Piguet, O. (2016). Neural substrates of semantic prospection - Evidence from the dementias. *Frontiers in Behavioral Neuroscience*, *10*(MAY), 1–12. https://doi.org/10.3389/fnbeh.2016.00096

Keramati, M., Dezfouli, A., & Piray, P. (2011). *Speed / Accuracy Trade-Off between the Habitual and the Goal-Directed Processes*. *7*(5). https://doi.org/10.1371/journal.pcbi.1002055

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition*, *20*(5), 353–379. https://doi.org/10.1521/soco.20.5.353.21125

Klein, S. B., Robertson, T. E., & Delton, A. W. (2010). *Facing the future: Memory as an evolved system for planning future acts*. *38*(1), 13–22. https://doi.org/10.3758/MC.38.1.13

Kliegel, M., Phillips, L. H., Lemke, U., & Kopp, U. A. (2005). *Planning and realisation of complex intentions in patients with Parkinson's disease*. 1501–1505. https://doi.org/10.1136/jnnp.2004.051268

Languis, M. L., & Miller, D. C. (1992). *Luria ' s Theory of Brain Functioning : A Model for Research in Cognitive Psychophysiology Luria ' s Theory of Brain Functioning : A Model for Research in Cognitive Psychophysiology*. *February 2015*, 37–41. https://doi.org/10.1207/s15326985ep2704

Lezak, M. D. (1982). *The problem of assessing executive functios*. *17*, 281–297.

Meer, M. Van Der, Kurth-nelson, Z., & Redish, A. D. (2015). *Information Processing in Decision-Making Systems*. *18*(4), 342–359. https://doi.org/10.1177/1073858411435128.Information

Mikadze, Y. V. (2014). *The principles of plasticity in Lurian neuropsychology*. 435–441. https://doi.org/10.3922/j.psns.2014.4.02

Mikadze, Y. V, Ardila, A., & Akhutina, T. V. (2019). *A . R . Luria ' s Approach to Neuropsychological Assessment and Rehabilitation*. *August*. https://doi.org/10.1093/arclin/acy095

Miller, E. K., & Cohen, J. D. (2001). *An integrative theory of Prefrontal Cortex*

function. *May*. https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, K. J., Brody, C. D., & Botvinick, M. M. (2016). *Identifying Model-Based and Model-Free Patterns in Behavior on Multi-Step Tasks*.

Miranda, B., Malalasekera, W. M. N., Behrens, T. E., Dayan, P., & Kennerley, S. W. (2019). *Combined model-free and model-sensitive reinforcement learning in non-human primates*. 1–24.

Morris, R., & Ward, G. (2005). The cognitive psychology of planning. In *The Cognitive Psychology of Planning*. https://doi.org/10.4324/9780203493564

Nadel, L., & Hardt, O. (2011). Update on Memory Systems and Processes. *Neuropsychopharmacology*, 251–273. https://doi.org/10.1038/npp.2010.169

Naglieri, J. A., Princiotta, D., & Otero, T. (2014). *Introduction : A History of Executive Functioning as a Theoretical and Clinical Construct*. *October 2016*. https://doi.org/10.1007/978-1-4614-8106-5

Nevin, J. A. (1999). *Analyzing Thorndike's Law of Effect: the question of stimulus-response bonds*. *3*(3), 447–450.

Olney, N. T., Spina, S., & Miller, B. L. (2017). Frontotemporal Dementia. *Neurologic Clinics*, *35*(2), 339–374. https://doi.org/10.1016/j.ncl.2017.01.008

Otto, A. R., Gershman, S. J., & Daw, N. D. (2013). *The Curse of Planning: Dissecting multiple reinforcement learning systems by taxing the central executive*. *24*(5). https://doi.org/10.1177/0956797612463080.The

Owen, A. M. (1997). Cognitive planning in humans: Neuropsychological, neuroanatomical and neuropharmacological perspectives. *Progress in Neurobiology*, *53*(4), 431–450. https://doi.org/10.1016/S0301-0082(97)00042-7

Palombo, D. J., Keane, M. M., & Verfaellie, M. (2015). The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking. *Hippocampus*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Redish, A. D. (2013). *The mind within the brain*.

Redish, A. D., & Mizumori, S. J. Y. (2016). *Memory and decision making*. 1–3. https://doi.org/10.1016/j.nlm.2014.08.014.Memory

Rorden, C., & Karnath, H. (2004). *Using human brain lesions to infer function : a relic from a past era in the fMRI age ?* *5*(October), 3–9.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, *October 2007*. https://doi.org/10.1038/nrn2213

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*. https://doi.org/10.1161/CIRCULATIONAHA.110.956839

Shahar, N., Hauser, T. U., Moutoussis, M., Moran, R., Keramati, M., Consortium, N. S. P. N., & Dolan, R. J. (2019). Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Computational Biology*, *15*(2), 1–25.

https://doi.org/10.1371/journal.pcbi.1006803

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *298*(1089), 199–209. https://doi.org/10.1098/rstb.1982.0082

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain*, *127*(4), 851–859. https://doi.org/10.1093/brain/awh100

Skinner, B. F. (1984). The evolution of behavior. *Scientific American*, *239*(3), 176–178, 180, 182 passim. https://doi.org/10.1901/jeab.1984.41-217

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Stuss, D. T. (2011). *Functions of the Frontal Lobes : Relation to Executive Functions*. 759–765. https://doi.org/10.1017/S1355617711000695

Stuss, D. T., & Benson, D. F. (1984). *Neuropsychological Studies of the Frontal Lobes*. *95*(1), 3–28.

Suddendorf, T., & Corballis, M. C. (2017). *The evolution of foresight : What is mental time travel , and is it unique to humans ? 299*, 299–351.

Sutton, R. S., & Barto, A. G. (2015). *Reinforcement Learning : An Introduction*.

Szpunar, K. K., Spreng, R. N., & Schacter, D. L. (2014). *A taxonomy of prospection : Introducing an organizational framework for future- oriented cognition*. *111*(52). https://doi.org/10.1073/pnas.1417144111

Tolman, E. C. (1948). COGNITIVE MAPS IN RATS AND MEN. *The Psychological Review*, *55*, 189–208.

Tulving, E. (1985). *How Many Memory Systems Are There?* 385–398.

Tulving, E., Gordon Haymann, C. A., & Macdonal, C. A. (1991). Long-lasting perceptual priming and semantic learning in amnesia_ a case experiment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*(4), 595–617.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and Declarative Memory: Role of the Hippocampus. *HIPPOCAMPUS*, *204*, 198–204.

Vikbladh, O., Shohamy, D., & Daw, N. D. (2016). *Episodic Contributions to Model - Based Reinforcement Learning*. 2016–2017.

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(4). https://doi.org/10.1101/cshperspect.a006171

Zelazo, D., A., C., & S., R. J. (2006). The Early Development of Executive Function. *Lifespan Cognition Mechanisms of Change*, *L*(2), 70–95. https://doi.org/10.1016/j.ijcard.2015.08.165

Zhu, Y., Liu, X., Wang, H., Jiang, T., Fang, Y., Hu, H., Wang, G., Wang, X., Liu, Z., &

Zhang, K. (2010). Reduced prefrontal activation during Tower of London in first-episode schizophrenia : A multi-channel near-infrared spectroscopy study. *Neuroscience Letters*, *478*(3), 136–140. https://doi.org/10.1016/j.neulet.2010.05.003