Carolina Feher da Silva

Decisão binária em seres humanos e agentes de inteligência artificial

Trabalho de Conclusão de Curso apresentado
ao Instituto de Matemática e Estatística, da
Universidade de São Paulo

Curso de Bacharelado em Ciências da
Computação

ORIENTADOR: Profa. Dra. Leliane Nunes de Barros

São Paulo
2014

Sumário

Agradecimentos

Agradeço a todos os professores, funcionários e alunos do IME e do IF que me ajudaram, direta ou indiretamente, a realizar este projeto. Em especial, à Professora Leliane Nunes de Barros (IME) por ter orientado o meu TCC e aos Professores Nestor Caticha (IF) e Marcus Vinícius C. Baldo (ICB) por terem conversado comigo sobre o projeto e me dado sugestões valiosas.

Agradeço também a todos os voluntários que, ganhando um chocolate ou não, tiveram paciência e me ajudaram a testar algumas das dezenas de versões alfa, beta e 1.0 do experimento com seres humanos.

Gostaria também de agradecer à minha família e aos meus amigos pelo apoio.

Principais Constantes e Variáveis

p**:** probabilidade da alternativa em maioria em uma tarefa de escolha binária repetida

α**:** taxa de aprendizado do algoritmo Q-Learning

γ**:** fator de desconto do algoritmo Q-Learning

τ**:** temperatura, um parâmetro da distribuição de Boltzmann, usada no algoritmo Q-Learning

K**:** tamanho da memória dos agentes computacionais

η**:** histórico de resultados durante uma tarefa de escolha binária repetida

R_a**:** recompensa recebida por um agente Q-Learning após uma previsão certa em uma tarefa de escolha binária repetida

R_e**:** recompensa recebida por um agente Q-Learning após uma previsão errada em uma tarefa de escolha binária repetida

Parte 1

Objetiva

CAPíTULO 1

Proposta

Desde os anos 1950, uma longa série de estudos vem sendo publicada sobre o comportamento de seres humanos em uma tarefa simples de tomada de decisão, que envolve escolher repetidamente entre duas alternativas — uma *tarefa de escolha binária repetida* [**1**]. Em cada *apresentação* de um experimento típico, voluntários humanos tentam prever qual dentre duas luzes, a da esquerda ou a da direita, se acenderá a seguir; um experimento consiste de múltiplas apresentações. Os voluntários são instruídos a maximizar o número de previsões corretas e frequentemente recebem uma recompensa em dinheiro por cada acerto.

Qual luz se acende de fato é determinado aleatoriamente com probabilidades fixas e independentes dos resultados anteriores e das respostas do voluntário. Assim, como somente uma das luzes se acende em cada apresentação, se $p \geq 0{,}5$ é a probabilidade de uma das luzes se acender, a probabilidade de a outra luz se acender é $1 - p \leq 0{,}5$. Em geral, os voluntários são informados de que suas respostas não influenciam a determinação da alternativa correta; no mais, não recebem nenhuma outra informação sobre o algoritmo utilizado no experimento.

Após um período de aprendizado, se $p > 0{,}5$, os voluntários são capazes de perceber que a luz do lado com probabilidade p — a alternativa em maioria — se acende mais frequentemente do que a luz do outro lado — a alternativa em minoria — e passam a escolher a alternativa em maioria com maior frequência. Os resultados de um experimento típico para $p = 0{,}7$ podem ser vistos nas Figuras 1.0.1 e 1.0.2. Nas últimas 10 apresentações dentre as 300 que compunham a tarefa, os voluntários tiveram uma resposta média de 0,763 (IC 95% [0,726-0,799]), considerando a alternativa em maioria como 1 e a alternativa em minoria como 0[1]; ou seja, em média eles escolheram a alternativa em maioria 76,3% das vezes. Em geral, quando um número relativamente pequeno de respostas é analisado, os resultados de experimentos com seres humanos indicam que a proporção média com que os voluntários escolhem cada alternativa está próxima da probabilidade com aquela alternativa está associada à recompensa; por exemplo, no experimento citado, a resposta média de 0,763 está próxima de $p = 0{,}7$. Por isso, a estratégia empregada por seres humanos é conhecida como *pareamento de probabilidades* [**1**].

No entanto, para se obter o maior ganho esperado, deve-se escolher *sempre* a alternativa em maioria, uma estratégia conhecida como *perseveração* ou

[1]Quando calcularmos a resposta média neste trabalho, sempre consideraremos a alternativa em maioria como 1 e a alternativa em minoria como 0. Assim, ela será sempre igual à frequência méïa com que os agentes ou voluntários humanos escolheram a alternativa em maioria.

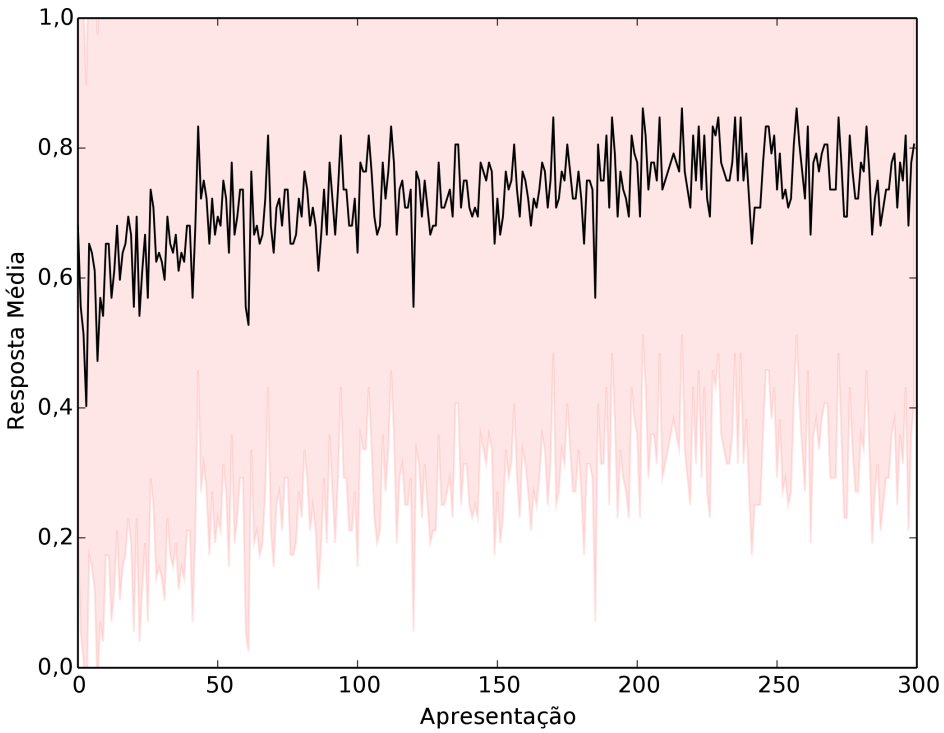


FIGURA 1.0.1. Resultados de um experimento de escolha binária repetida realizado com 72 alunos de odontologia da USP [**2**]. Para cada uma das 300 apresentações da tarefa, é dada a resposta média dos voluntários. A área colorida corresponde ao desvio padrão.

maximização [**1**]. Se $p > 0{,}5$, a probabilidade de acertar o resultado de uma apresentação usando perseveração é $1{\cdot}p{+}0{\cdot}(1{-}p) = p$. No entanto, se a alternativa em maioria é escolhida com probabilidade $\phi < 1$, a probabilidade de acertar o resultado é $\phi p{+}(1{-}\phi)(1{-}p) = \phi(2p{-}1){+}1{-}p < (2p{-}1){+}1{-}p = p$, ou seja, ela é menor do que p. Portanto, os seres humanos em geral não adotam a estratégia ótima.

Devido à simplicidade desta tarefa, é possível adaptá-la para ser realizada com outras espécies de animais. Em 1958, por exemplo, Parducci e Polt realizaram-na com ratos, usando $p = 0{,}85$ [**3**]. Para isso, os ratos eram repetidamente colocados na base de um labirinto em T, como mostrado na Figura 1.0.3. Ao chegar ao topo do labirinto, eles deveriam escolher um dos dois braços e, na extremidade do braço escolhido, poderiam encontrar ou não uma porção de comida. O desempenho dos ratos na tarefa foi superior ao de seres humanos: nas últimas 10 apresentações dentre as 60 totais, os ratos escolheram a alternativa em maioria 98,8% das vezes. Resultados semelhantes foram obtidos ao se realizarem tarefas análogas com pombos [**4**] e peixes [**5**]. Ou seja, no experimento de escolha binária repetida, os seres humanos não só não adotam a estratégia ótima, como também são passados para trás por ratos, pombos e peixes.

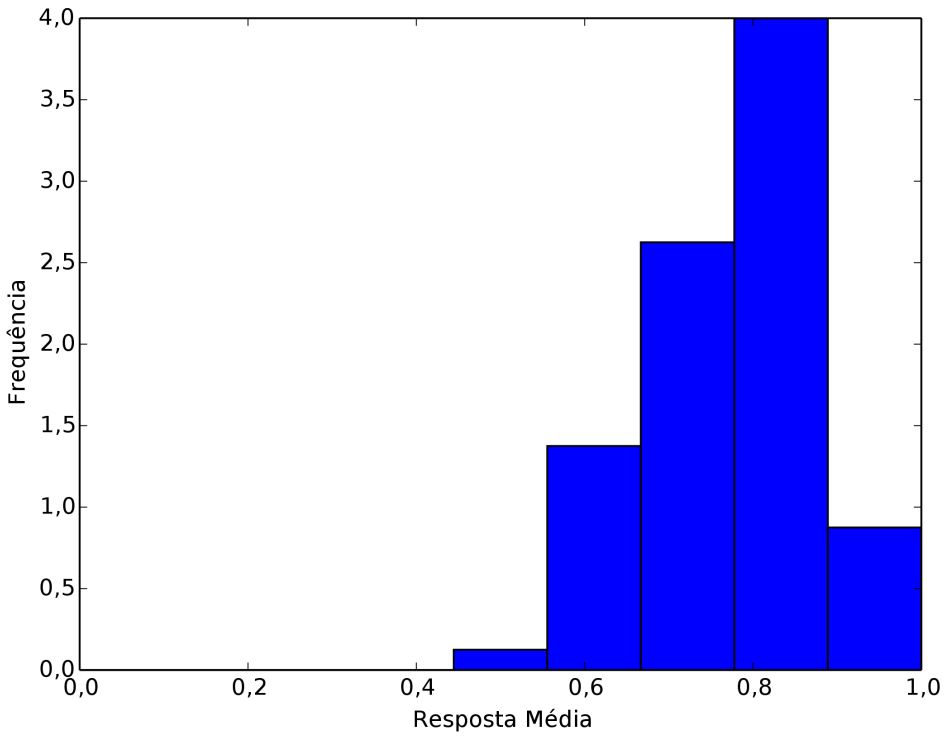


FIGURA 1.0.2. Resultados de um experimento de escolha binária repetida realizado com 72 alunos de odontologia da USP [**2**]. É dado o histograma da resposta média dos voluntários nas 100 últimas apresentações da tarefa.

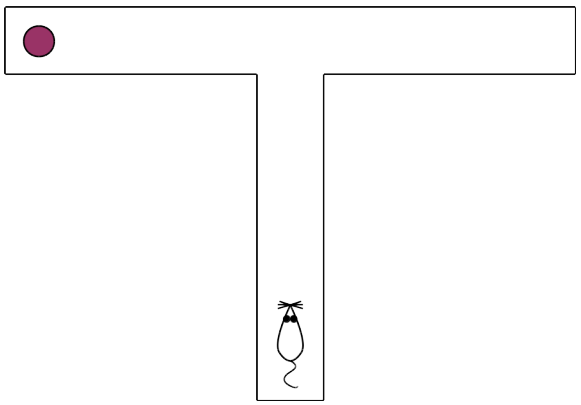


FIGURA 1.0.3. Labirinto em T usado para estudar como os ratos tomam decisões binárias.

Não é claro por que os seres humanos não obtêm um bom desempenho nesta tarefa. A explicação mais aceita é a de que eles não percebem que a recompensa está distribuída aleatoriamente e que cada apresentação é independente das apresentações anteriores; ao invés disso, eles acreditam na existência de um padrão na sequência de resultados [**1**]. Segundo esta explicação, os voluntários tentam reproduzir um padrão com as mesmas

proporções dos resultados observados e assim acabam fazendo pareamento de probabilidades. Outros animais, tais como ratos e pombos, não têm a mesma habilidade de buscar padrões em sequências e adotam a estratégia mais simples de perseverar. As evidências favoráveis a esta explicação são abundantes [**6, 7, 8, 9, 10, 11, 12**]; ainda assim, há muitas explicações alternativas plausíveis, tais como intuição matemática equivocada [**13**], motivação e prática insuficientes [**14**], adaptação à incerteza [**15, 16**], adaptação ao forrageamento em um ambiente competitivo [**17, 18**], uso de atalhos cognitivos [**11**] e pareamento de probabilidades como consequência do aprendizado quase ótimo de estruturas [**19**].

Além disso, como já foi dito, o pareamento de probabilidades só é observado quando um número pequeno de respostas é analisado [**1**]. Já em experimentos mais longos, em média os voluntários escolhem a alternativa em maioria nas últimas apresentações com frequência superior a p [**20**], o que nos levaria a falar em *supra*pareamento de probabilidades. (Note que, mesmo no experimento da Figura 1.0.1, embora a resposta média dos voluntários tenha sido próxima do valor de p, ela está acima dele.) Assim, a descrição "pareamento de probabilidades" não descreve de forma acurada o comportamento assintótico dos seres humanos nesta tarefa [**1**].

Para explicar tais resultados, foi sugerido que, em tarefas de escolha binária repetida, os seres humanos adotam inicialmente o pareamento de probabilidades, mas, se a tarefa for longa o suficiente, eles mudam de estratégia e passam a perseverar [**9**]. No entanto, não há evidências de uma mudança brusca de estratégia durante o experimento. Ao invés disso, é possível que a probabilidade média de um voluntário escolher a alternativa em maioria cresça constantemente até atingir a perseveração e que, portanto, o pareamento de probabilidades seja apenas uma situação transiente e não uma estratégia de decisão propriamente dita. Assim, deste ponto de vista, a pergunta a ser feita não é "Por que os seres humanos não perseveram?" e sim "Por que os seres humanos demoram para aprender a perseverar?"

Para tentar responder essa pergunta, observamos que dois estudos indicam que o desempenho dos seres humanos é maior quanto menor é a sua capacidade de reter informações na memória de curto prazo[2] [**8, 10**]. Tais resultados vêm sendo usados para apoiar a teoria de que os seres humanos fazem pareamento de probabilidades por buscarem padrões, pois, quando a memória de curto prazo tem baixa capacidade, os voluntários não conseguem se lembrar dos resultados das apresentações anteriores e, portanto, não conseguem buscar padrões.

No entanto, se supusermos que os seres humanos apenas demoram para aprender a perseverar, devemos também concluir que estes estudos são evidência de que a capacidade da memória de curto prazo de um voluntário influencia a sua velocidade de aprendizado. Podemos imaginar que, ao tentar encontrar um padrão, um voluntário tenta estabelecer uma relação entre o resultado de cada apresentação e os resultados das apresentações anteriores das quais ele se lembra. Se ele só se lembra de uma apresentação anterior,

[2] A memória de curto prazo contém a informação relevante que mantemos em mente quando estamos executando uma tarefa ou resolvendo um problema. Este tipo de memória é discutido com mais detalhes a partir da página 14.

só existem duas situações que serão por ele consideradas: (1) o resultado da apresentação anterior é a alternativa em maioria e (2) o resultado da apresentação anterior é a alternativa em minoria. Sua tarefa passa a ser, portanto, estimar qual resultado é mais provável em cada uma dessas duas situações. Já quando ele se lembra de duas apresentações anteriores, ele terá que estimar o resultado mais provável em cada uma de quatro situações. De modo geral, se ele é capaz de se lembrar de K resultados anteriores, ele deverá estimar 2^K probabilidades. Cada combinação de resultados anteriores se torna mais rara quanto maior for K e portanto leva mais tempo para que o voluntário tenha uma boa estimativa das probabilidades envolvidas.

Assim, propomos os seguintes pontos para explicar o comportamento de seres humanos em experimentos de escolha binária repetida.[3]

(1) Os seres humanos não empregam o pareamento de probabilidades como estratégia de decisão. Eles apenas demoram para aprender a perseverar.
(2) Isso ocorre porque eles não são em geral capazes de perceber que as alternativas corretas da escolha binária são selecionadas aleatoriamente, sem dependência dos resultados anteriores. Ao invés disso, eles procuram padrões na sequência de resultados.
(3) Quanto maior é o número de resultados armazenados na memória de curto prazo e usados para tomar as próximas decisões, menor é a velocidade de aprendizado.

Para testar essa proposta, desenvolvemos agentes computacionais baseados no Processo de Decisão Markoviano [**21**] e capazes de realizar a tarefa de escolha binária repetida, e realizamos experimentos com voluntários humanos. A abordagem utilizada, bem como os resultados obtidos, serão descritos nos capítulos a seguir.

[3]Ou seja, experimentos em que repetidamente os voluntários fazem uma previsão e são imediatamente informados do resultado. Há estudos que usam tarefas diferentes ([**13**], por exemplo) para abordar as mesmas questões, mas aos quais a nossa proposta não se aplica.

CAPíTULO 2

Hábitos e Memória

Primeiramente, para modelar computacionalmente os mecanismos de tomada de decisão do cérebro humano, devemos entender que as decisões humanas não são tomadas por um único mecanismo. Ao invés disso, no sistema nervoso, parecem existir múltiplos sub-sistemas que são ativados em situações distintas [**22**]. O sistema *direcionado a um objetivo*, por exemplo, toma decisões baseado em considerações sobre os possíveis resultados e suas futuras ações.

No entanto, um mecanismo decisional diferente parece estar envolvido em tarefas que envolvem reforço a cada decisão. Neste caso, as escolhas dos indivíduos parecem ser menos influenciadas por deliberações sobre qual estratégia leva ao maior ganho e mais por um *sistema habitual,* que aprende a tomar boas decisões por tentativa e erro [**23, 22**]. A tarefa de escolha binária repetida se encaixa claramente neste contexto; assim, neste trabalho, nos concentraremos no sistema habitual.

O estudo do sistema habitual teve suas origens no início do século XX. Naquela época, o psicólogo americano Edward Thorndike treinou gatos para escapar de caixas ativando diferentes mecanismos [**24**]. Durante as primeiras tentativas de escapar de uma determinada caixa, o gato testava diferentes comportamentos sem sucesso até que, por acidente, ativava o mecanismo que abria a caixa. A partir deste ponto, o tempo que o gato levava para abrir a caixa diminuia, pois ele deixava gradualmente de usar comportamentos ineficazes. Isso levou Thorndike a propôr a Lei do Efeito, que afirma que ações seguidas por uma recompensa têm maior probabilidade de recorrer em situações semelhantes futuras; ações que são seguidas por punições, por sua vez, têm menor probabilidade de recorrer no futuro [**25**].

O sistema habitual é capaz de aprender a atribuir às ações um valor proporcional à recompensa esperada que estas ações geram. Para isso, é necessário que haja prática suficiente e o ambiente seja suficientemente estável, pois, como o aprendizado se dá por tentativa e erro, ele é relativamente lento [**26**].

Nós nos referimos às ações controladas por este sistema por "hábitos". Um exemplo de hábito é acender a luz automaticamente, sem antecipar as consequências, quando chegamos em casa no escuro [**27**]. É possível tomar a mesma ação usando o sistema direcionado a um objetivo; no entanto, isso requer o desejo consciente de iluminar o ambiente e a crença em que um determinado movimento levará à realização deste desejo. A diferença entre os dois sistemas pode ser ilustrada no caso em que, quando se sabe que há um apagão, o hábito de acender a luz pode persistir, enquanto que a ação direcionada a um objetivo não persiste. As ações direcionadas a um objetivo

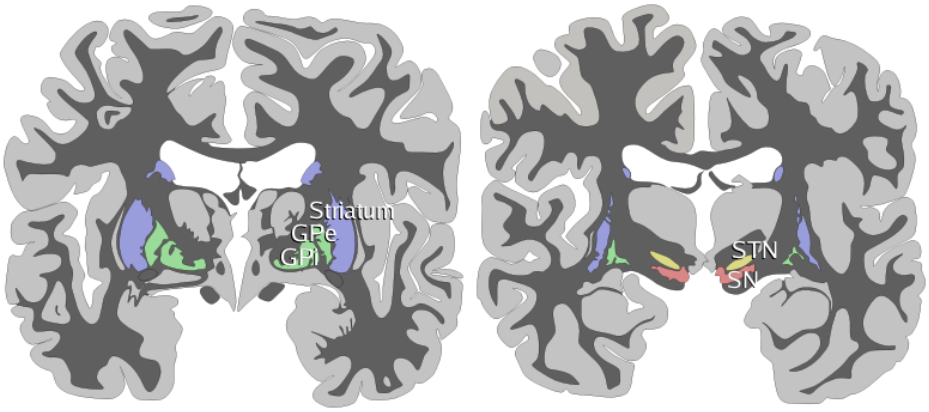


FIGURA 2.0.1. Fatias coronais (resultante de cortes de um lado do corpo ao outro) mostrando os núcleos da base. Na fatia à esquerda, em posição anterior (próxima à face), é possível ver o estriado e o globo pálido (GPe e GPi). Na fatia à direita, em posição posterior, é possível ver o núcleo subtalâmico e a substância negra. Por Andrew Gillies, CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/deed.en), via Wikimedia Commons.

são, assim, controladas por suas consequências, enquanto que os hábitos são controlados pelos estímulos precedentes.

2.0.1. Núcleos da base e dopamina. Estudos com animais de diferentes espécies usando métodos variados sugerem que uma estrutura do sistema nervoso conhecida como *estriado* exerce um papel fundamental no aprendizado de ações habituais [**27**]. Ele é parte dos *núcleos da base,* um conjunto de núcleos na base do cérebro que também engloba o globo pálido, a substância negra e o núcleo subtalâmico (Figura 2.0.1).

Os núcleos da base em geral são considerados componentes do sistema motor e estão envolvidos na regulação dos movimentos voluntários. Em seres humanos, as doenças que afetam os núcleos da base, como a doença de Parkinson ou a doença de Huntington, resultam em movimentos voluntários lentos ou no desenvolvimento de movimentos involuntários [**28**]. Adicionalmente, os núcleos da base exercem funções não-motoras, incluindo cognição, emoção e possivelmente outras.

Apesar de a participação dos núcleos da base no controle dos movimentos não estar totalmente esclarecida, uma teoria popular afirma que eles agem para permitir a execução automática, ou seja, habitual, de sequências aprendidas de movimentos, bem como o seu aprendizado [**28**]. Mudanças na atividade de certos grupos de neurônios do estriado e da substância negra podem ser detectadas conforme uma tarefa é aprendida e se torna automática ou quando estímulos ou eventos novos são introduzidos. Além disso, sabe-se que lesões no estriado ou a doença de Parkinson prejudicam este tipo de aprendizado. Assim, de acordo com esta proposta, os núcleos da base podem ser considerados como a implementação neural da Lei do Efeito

de Thorndike, sendo responsáveis por formar a associação entre estímulos e ações que resultam na obtenção de recompensas [**27**].

O sinal de reforço usado no aprendizado parece ser fornecido pela *dopamina,* um neurotransmissor produzido e secretado em várias áreas do cérebro, incluindo a substância negra e a área tegmental ventral. Esta última se localiza em posição inferior à dos núcleos da base e se projeta para o estriado, exercendo um papel central na sinalização de recompensa e outros aspectos da motivação [**22**]. Todas as drogas de abuso, por exemplo, agem direta ou indiretamente no sistema dopaminérgico. A nicotina, a morfina e o etanol ativam os neurônios dopaminérgicos direta ou indiretamente. Já a cocaína e a anfetamina bloqueiam a retirada de dopamina das sinapses e assim aumentam a ação natural do neurotransmissor. A anfetamina, além disso, causa a liberação de dopamina nas sinapses. Desta forma, estas drogas aumentam o nível de dopamina nos núcleos da base e se acredita que isso lhes confere o seu caráter viciante, formador de hábitos.

Medidas experimentais da atividade fásica dos neurônios dopaminérgicos são consistentes com a proposta de que o sinal transmitido corresponde ao erro de predição da recompensa, ou seja, à diferença entre a recompensa esperada, dadas as experiências anteriores, e a recompensa obtida em dado momento. Algoritmos computacionais como o Q-Learning, discutido no capítulo seguinte, também calculam o erro de predição da recompensa para possibilitar o aprendizado e o planejamento. Assim, por sua inspiração biológica, usaremos tais algoritmos para criar agentes computacionais que realizam a tarefa de escolha binária repetida.

2.0.2. Memória de curto prazo. Para buscar e aprender padrões na tarefa de escolha binária repetida, no entanto, a existência de um mecanismo com capacidade de aprendizado por tentativa e erro não é suficiente. É necessário, além disso, que haja o armazenamento dos resultados anteriores em algum tipo de memória.

Quando nós pensamos em memória, geralmente pensamos na ampla quantidade de informações persistentes às quais temos acesso consciente, mesmo quando elas não são recuperadas por um longo período. Tal memória é chamada de memória de longo prazo explícita. No entanto, existem outros tipos de memória que não possuem todas estas qualidades. Como vimos, a informação sobre ações executadas e recompensas obtidas que é armazenada durante a criação de um hábito não pode ser acessada conscientemente — é um tipo de memória de longo prazo implícita.

Já um outro tipo de memória, a memória de curto prazo ou memória operacional, pode ser acessada conscientemente, mas armazena apenas por segundos ou minutos uma quantidade pequena de informações relevantes para os objetivos presentemente considerados. Além disso, tais informações, ao deixarem de ser relevantes, serão descartadas, a menos que sejam transferidas para a memória de longo prazo. A memória de curto prazo é usada quando, por exemplo, repetimos mentalmente um número de telefone para que não nos esqueçamos dele até que ele seja anotado. Em humanos, a memória de curto prazo é composta por pelo menos dois subsistemas — um para informações verbais e outro para informações visuoespaciais [**29**]

— e há evidências de que o subsistema verbal se localiza no hemisfério esquerdo do cérebro, enquanto que o visuoespacial se localiza no hemisfério direito [**30, 31, 32**].

Neste trabalho, vamos assumir que cada agente computacional ou voluntário humano que realiza a tarefa de escolha binária repetida armazena um número fixo de resultados anteriores na memória de curto prazo. Esta informação será, então, usada para prever o resultado da próxima apresentação.

CAPíTULO 3

Aprendizado por Reforço

Nas ciências da computação, o estudo da tomada de decisão pode dar-se como planejamento, um ramo da inteligência artificial que se concentra em resolver o problema de calcular sequências de decisões para que um agente inteligente atinja seu objetivo. Uma abordagem importante em planejamento é o aprendizado por reforço, no qual agentes aprendem a tomar decisões que maximizam a obtenção de recompensas ao interagir com o ambiente [**25**]. O aprendizado por reforço foi inspirado pelo estudo da tomada de decisão em outra disciplina, a psicologia — as ideias sobre o sistema habitual, discutidas no capítulo anterior, inspiraram, nos anos 1980, a criação de agentes computacionais que buscam descobrir quais ações resultam nas maiores recompensas executando tais ações; ou seja, eles aprendem por tentativa e erro.

O problema de aprendizado por reforço é uma tentativa de formalizar o problema de aprender pela interação com o ambiente a fim de atingir um objetivo [**25**]. O aprendiz e tomador de decisões é chamado de *agente* e tudo o que é externo ao agente constitui o *ambiente.* O agente e o ambiente interagem continuamente – o agente selecionando uma *ação* e o ambiente respondendo a tal ação por meio de *recompensas* e apresentando novas situações, ou *estados* (Figura 3.0.1). Mais especificamente, a cada passo de tempo $t = 0,1,2,\ldots$ o agente recebe uma representação $S_t \in \mathcal{S}$ no passo t do estado do ambiente, onde $\mathcal{S}$ é o conjunto de estados possíveis. O agente, por sua vez, seleciona uma ação $A_t \in \mathcal{A}$, onde $\mathcal{A}$ é o conjunto de ações disponíveis. No próximo passo de tempo $t+1$, o agente recebe uma recompensa numérica $R_{t+1} \in \mathbb{R}$ e se encontra no próximo estado S_{t+1}.

A probabilidade de o agente selecionar uma ação a em um estado s no instante t é dada por $\pi_t(a|s)$, o valor da função π_t, chamada de *política,* naquele ponto. Os métodos de aprendizado por reforço determinam como o

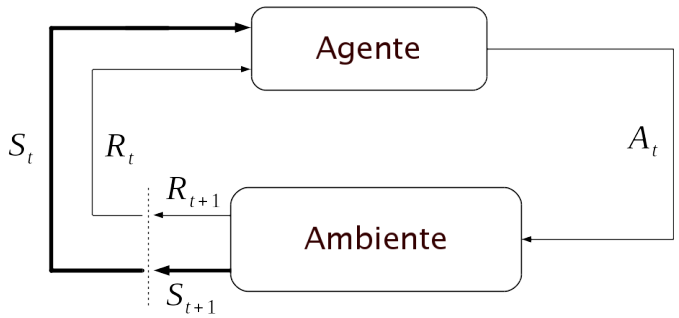


FIGURA 3.0.1. Aprendizado por reforço

agente muda sua política como resultado de sua experiência. O objetivo do agente é maximizar o *retorno descontado* esperado G_t, dado pela soma das recompensas descontadas que ele receberá no futuro:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}, \tag{3.0.1}$$

onde γ é um parâmetro, $0 \leq \gamma \leq 1$, chamado de *fator de desconto*. O fator de desconto determina o valor presente das recompensas futuras, tal que uma recompensa a ser recebida em k passos vale somente γ^k vezes o que ela valeria se fosse recebida imediatamente.

Quase todos os algoritmos de aprendizado por reforço envolvem determinar *funções valor*, que estimam o quão bom é para um agente estar em um determinado estado, em termos das futuras recompensas esperadas. Como tais valores dependem das ações que o agente irá tomar em cada estado, as funções valores são definidas em relação a uma política. Para este trabalho, será estimado o valor de tomar uma ação em um determinado estado. O valor de tomar a ação a em um estado s sob uma política π, dado por $q_\pi(s,a)$, é definido como o retorno esperado começando em s, tomando a ação a e posteriormente seguindo a política π:

$$q_\pi(s,a) = \mathrm{E}\left[G_t | S_t = s, A_t = a\right]. \tag{3.0.2}$$

A função q_π é chamada de função ação-valor para a política π.

Em muitos problemas de aprendizado por reforço, é possível provar que existe pelo menos uma política π^*, chamada de *política ótima*, que é melhor ou igual a todas as outras políticas. As políticas ótimas compartilham a mesma função ação-valor ótima $q*$, definida como

$$q*(s,a) = \max_\pi q_\pi(s,a), \tag{3.0.3}$$

para todo estado $s \in \mathcal{S}$ e toda ação $a \in \mathcal{A}$.

3.0.3. Exploração *versus* exploitação. Durante um problema de aprendizado, um agente deve se concentrar nas ações de maior valor a fim de obter uma boa recompensa total. Se o agente mantém estimativas dos valores das ações e seleciona uma ação de valor estimado máximo, a ação é chamada de *gulosa* e diz-se que o agente está realizando *exploitação*. No entanto, em muitos problemas, é necessário tomar ações não-gulosas a fim melhorar a estimativa do valor daquelas ações. Neste caso, diz-se que o agente está realizando uma *exploração*. A exploitação é a melhor estratégia para maximizar a recompensa esperada naquele passo de tempo, mas a exploração pode produzir a maior recompensa total a longo prazo.

Assim, em um algoritmo de aprendizado por reforço, pode-se realizar exploitação na maior parte do tempo e com uma pequena probabilidade ε, selecionar uma ação aleatória com igual probabilidade, sem considerar as estimativas atuais de seus valores. Tal método de seleção de ação é chamado de *ε-guloso*. Alternativamente, o método *softmax* de seleção de ação usa a distribuição de Boltzmann para atribuir uma probabilidade de seleção $p_{A,t}$ a cada ação A no passo t de acordo com seus valores estimados:

$$p_{A,t} = \frac{e^{Q_t(A)/\tau}}{\sum_{a \in \mathcal{A}} e^{Q_t(a)/\tau}}, \tag{3.0.4}$$

Algoritmo 1 Q-Learning

Inicialize $Q(s,a) = 0, \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$
Inicialize S aleatoriamente
Repita:
Escolha A em S usando uma política derivada de Q (ε-gulosa ou *softmax* com a distribuição de Boltzmann).
Tome a ação A, observe R, S'
$Q(S_t,A_t) \leftarrow Q(S_t,A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)]$
$S \leftarrow S'$

onde τ é um parâmetro positivo chamado de *temperatura.* Quanto maior é a temperatura, maior a exploração.

3.0.4. Q-Learning. O Q-Learning é um algoritmo de aprendizado por reforço que vem sendo muito utilizado para modelar decisões por seres humanos (por exemplo, [**33, 34, 35, 22**]). Ele mantém uma estimativa Q do valor ótimo $q*$ de uma ação em um dado estado e atualiza esta estimativa após cada apresentação de acordo com a regra

$$Q(S_t,A_t) \leftarrow Q(S_t,A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)], \tag{3.0.5}$$

tendo dois parâmetros: α, a taxa de aprendizado, e γ, a fator de desconto. O termo $R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)$, que corresponde à diferença entre a recompensa prevista ($Q(S_t,A_t)$) e a recompensa recebida ($R_{t+1} + \gamma \max_a Q(S_{t+1},a)$), é chamado de *diferença temporal.*

O Algoritmo 1 é um algoritmo de Q-Learning básico.

3.0.5. Aprendizado por reforço e neurociências. Devido à sua inspiração biológica, algoritmos de aprendizado por reforço passaram a ser utilizados por neurocientistas como modelos matemáticos da tomada de decisão em seres humanos e outros animais. Acredita-se que grupos distintos de neurônios nos núcleos da base codificam o valor da recompensa esperada, bem como erros positivos e negativos na previsão de recompensas, e há maior ativação quando a entrega de recompensa depende da ação do indivíduo [**22**]. Em particular, há evidências experimentais de que a diferença temporal corresponde ao sinal de reforço transmitido pela atividade fásica de neurônios dopaminérgicos [**36**].

CAPíTULO 4

Modelos Computacionais de Decisão Binária

Para que os agentes computacionais que construímos fossem capazes de realizar uma tarefa de escolha binária repetida e procurar padrões, eles precisaram se lembrar dos resultados das apresentações anteriores. O modelo geral que utilizamos para dar aos agentes esta habilidade é baseado no conceito de Cadeias de Markov.

Cadeias de Markov ou cadeias de Markov de ordem 1, como iremos chamá-las, são sequências de variáveis aleatórias nas quais o valor do elemento seguinte só depende do valor do elemento anterior. Por exemplo, o resultado de cada apresentação na tarefa de escolha binária repetida pode ser representado por uma variável aleatória X_i, onde i é o número da apresentação começando em 1. O valor de X_i será 0 se o estímulo i vier de um dado lado e 1 se o estímulo i vier do outro lado. Se o valor da próxima apresentação X_{i+1} não depender de nenhum outro valor além de X_i , então este experimento pode ser completamente descrito por uma cadeia de Markov de ordem 1; ou seja, $p(X_i|X_{i-1},\ldots,X_1) = p(X_i|X_{i-1})$. Além de cadeias de Markov de ordem 1, podemos considerar cadeias de Markov de ordens superiores. Em uma cadeia de ordem 2, por exemplo, o resultado de cada apresentação depende das duas apresentações anteriores: $p(X_i|X_{i-1},X_{i-2},\ldots,X_1) = p(X_i|X_{i-1},X_{i-2})$. De forma geral, em uma cadeia de ordem L, cada apresentação só depende das L apresentações anteriores: $p(X_i|X_{i-1},X_{i-2},\ldots,X_1) = p(X_i|X_{i-1},X_{i-2},...,X_{i-L})$.

Para modelar o aprendizado em um experimento de escolha binária repetida, como descrito no Capítulo 1, e no qual pode existir um padrão, foram criados agentes de Inteligência Artificial com a tarefa de prever o próximo elemento de uma sequência binária que constitui uma cadeia de Markov de ordem L. Seja t o instante de tempo considerado e $\eta_1^L(t)$ o conjunto dos últimos L elementos da sequência, ou seja, $(X_{t-1},X_{t-2},\ldots,X_{t-L})$. Pela descrição anterior, este conjunto de variáveis aleatórias caracteriza o estado do sistema que gera a sequência binária a ser prevista, ou seja, $p(X_i|X_{i-1},X_{i-2},\ldots,X_1) = p(X_t|\eta_1^L(t))$.

O agente, ao tentar prever o próximo elemento da sequência, no entanto, só é capaz de se lembrar do passado recente, o que inclui somente os K últimos elementos da sequência por ele observada. Assim, seu objetivo é determinar qual ação deve ser tomada após um passado $\eta_1^K(t)$. Para este fim, dois algoritmos foram utilizados: (1) o algoritmo Q-Learning de aprendizado por reforço e (2) o algoritmo Bayesiano proposto por Dobay, Caticha e Baldo (modelo DCB) [**37**]

4.1. Aprendizado por reforço

O algoritmo de aprendizado por reforço que usaremos é o Algoritmo 1 adaptado para realizar a tarefa de escolha binária repetida, usando uma política *softmax* com a distribuição de Boltzmann e os parâmetros $\gamma = 0{,}99$, $\alpha = 0{,}5$, $\tau = 1$. Quando o agente acertava uma previsão na tarefa de escolha binária repetida, recebia a recompensa $R_a = 1$. Tais valores foram selecionados por resultarem em um bom desempenho (estratégia perseveradora) no caso em que $K = 0$ e não por meio de dados biológicos. Os parâmetros que variamos nos experimentos abaixo foram p, a probabilidade da alternativa em maioria, K, o tamanho da memória, e R_e, a recompensa recebida pelo agente quando ele errava o valor do próximo elemento da sequência.

Segundo Sutton e Barto [**25**], a representação do ambiente que constitui um estado pode ser determinada exclusivamente por informações sensoriais, ser baseada na memória do agente ou mesmo ser totalmente subjetiva. No experimento de escolha binária repetida, nenhuma informação sensorial é relevante para determinar a escolha a ser feita, mas informações sobre os resultados anteriores podem ser importantes se o agente acredita que a sequência segue um padrão, dado por uma cadeia de Markov de ordem L. Assim, os estados do problema serão definidos como os estados da memória do agente, que contém os K últimos elementos da sequência que ele tenta prever, ou seja, $S = \eta$. Assim, existem 2^K estados no qual o agente pode se encontrar. Por exemplo, se $K = 2$ e os dois últimos elementos da sequência binária foram 11, então o agente se encontra no estado $S = 11$. Outros estados no qual este agente poderia se encontrar são 00, 01 e 10.

4.2. Modelo DCB

Outro algoritmo que usamos em experimentos computacionais foi recentemente proposto por Dobay, Caticha e Baldo para realizar tarefas de escolha binária repetida [**37**].

O algoritmo proposto por esses autores, que chamaremos de modelo DCB, tenta estimar por métodos Bayesianos a probabilidade p_η de o próximo elemento da sequência ser 1 após um passado η. Partindo da distribuição *a priori* de p estabelecida para a situação inicial, antes do início da tarefa — uma distribuição uniforme no intervalo [0,1] — o teorema de Bayes é aplicado após cada apresentação, resultando em uma distribuição *a posteriori*. Seja μ o número de vezes que o passado η foi observado e j o número de vezes que esse passado foi seguido por 1. É possível mostrar que a distribuição *a posteriori* é a distribuição beta com parâmetros $(j + 1, \mu - j + 1)$. Podemos tomar, então, a média dessa distribuição como uma estimativa p^* de p a ser usada para tomar uma decisão na próxima ocorrência do passado η, dada por:

$$p^* = \frac{t+1}{\mu+2}. \tag{4.2.1}$$

Se $p^* > \frac{1}{2}$, na próxima ocorrência de η, diremos que o próximo elemento é 1, se $p^* < \frac{1}{2}$, diremos que é 0 e se $p^* = \frac{1}{2}$, diremos que o próximo elemento é 0 ou 1 com igual probabilidade.

CAPíTULO 5

Experimentos com Agentes de Inteligência Artificial

A seguir, serão descritos cinco experimentos computacionais que realizamos para testar nossa proposta, em particular a afirmação de que quanto maior é a capacidade de memória a curto prazo, menor é a velocidade de aprendizado. Ela é testada diretamente nos experimentos 4 e 5 usando agentes Q-Learning e DCB com diversos valores de K. Em relação aos três experimentos computacionais precedentes, os dois primeiros mostram como o desempenho do Q-Learning é afetado pela recompensa obtida quando o agente erra o resultado da escolha binária. O experimento seguinte testa o desempenho do Q-Learning em tarefas com diferentes valores de p, o que também influencia a velocidade de aprendizado.

5.1. Experimento 1: Q-Learning com $p = 0{,}7$, $K = 0$ e $R_e = 0$

Este experimento consistiu em medir o desempenho do Q-Learning com parâmetros $p = 0{,}7$, $K = 0$ e $R_e = 0$ na tarefa de escolha binária repetida com 300 tentativas. Tais valores foram considerados para uma avaliação inicial do modelo, pois $p = 0{,}7$ e $R_e = 0$ foram usados no experimento com seres humanos discutido no início desta monografia, cujos resultados podem ser vistos nas Figuras 1.0.1 (página 8) e 1.0.2; $K = 0$ é a escolha ótima, pois, na tarefa realizada, o resultado de cada apresentação não depende dos resultados das apresentações anteriores.

Os resultados de 1000 repetições deste experimento podem ser visualizados nas Figuras 5.1.1 e 5.1.2. Se compararmos a Figura 5.1.1 (Q-Learning) com a Figura 1.0.1 (seres humanos), notamos que o desempenho médio do algoritmo é superior ao dos seres humanos. No entanto, a Figura 5.1.2 mostra que às vezes o algoritmo aprende a perseverar na alternativa em *minoria,* o que leva a um desempenho pior do que o de seres humanos — a Figura 1.0.2 é análoga à Figura 5.1.2, mas usa dados de seres humanos e mostra que estes não perseveram na alternativa em minoria.

5.2. Experimento 2: Q-Learning com $p = 0{,}7$, $K = 0$ e $R_e = -1$

Em relação ao experimento anterior, alteramos o valor de R_e, a recompensa que o agente recebe quando erra a previsão de um resultado, para -1.

Como pode ser visto nas Figuras 5.2.1 e 5.2.2, o desempenho do agente é melhor quando $R_e = -1$ do que quando $R_e = 0$, como visto no experimento anterior. O principal motivo é que a mudança deste valor impede que o algoritmo aprenda a perseverar na alternativa em *minoria,* o que também o torna mais próximo do comportamento humano. Por isso, foi usado

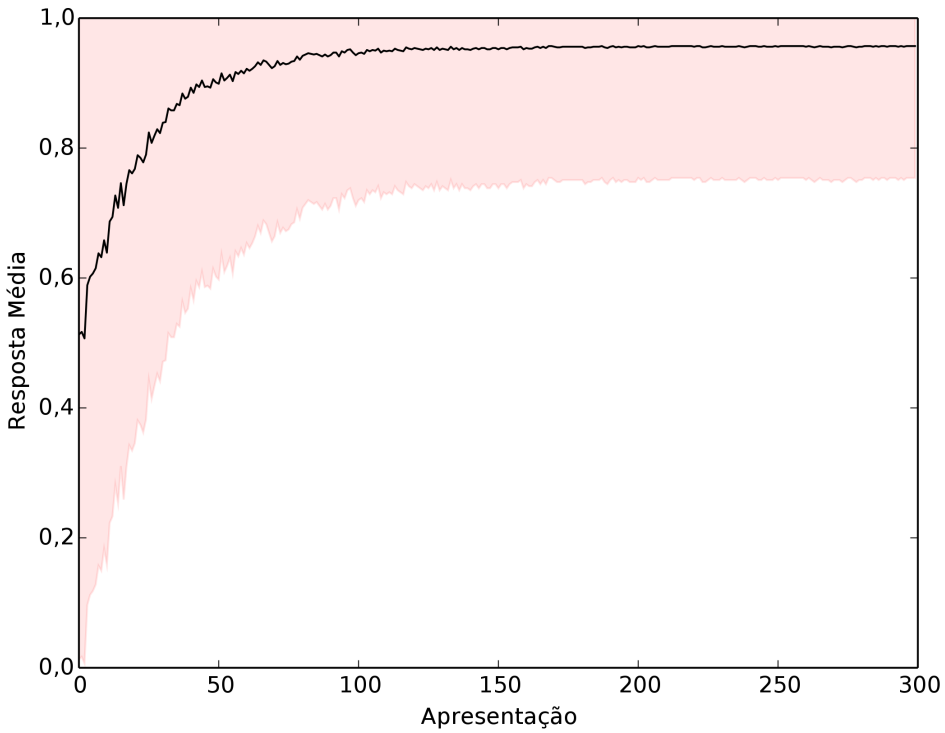


FIGURA 5.1.1. Resposta média do Q-Learning com parâmetros $p = 0{,}7$, $K = 0$ e $R_e = 0$. A área colorida corresponde ao desvio padrão. $N = 1000$.

o valor $R_e = -1$ nos experimentos computacionais seguintes que usam o Q-Learning.

5.3. Experimento 3: Q-Learning com p variável, $K = 3$ e $R_e = -1$

Neste experimento, a resposta média do agente foi calculada em tarefas de escolha binária repetida com p variável de 0,6 a 1,0. Usamos $K = 3$ e não $K = 0$ para que as diferenças entre as curvas pudessem ser melhor observadas. Como é possível ver na Figura 5.3.1, o agente aprende a perseverar mais rapidamente conforme o valor de p aumenta.

5.4. Experimento 4: Q-Learning com K variável, $p = 0{,}7$ e $R_e = -1$

A seguir, fizemos um teste computacional da afirmação de que quanto maior é a capacidade de memória a curto prazo, menor é a velocidade de aprendizado. Para isso, comparamos a resposta média do Q-Learning ao longo de 300 apresentações com $p = 0{,}7$ e K variando de 0 a 5. Os resultados podem ser visualizados na Figura 5.4.1. De fato, conforme aumenta o valor de K, mais lento é o crescimento da resposta média ao longo do tempo.

5.5. Experimento 5: DCB com K variável e $p = 0{,}7$

Os resultados obtidos no experimento anterior foram reproduzidos utilizando o outro algoritmo de aprendizado de sequências binárias estudado,

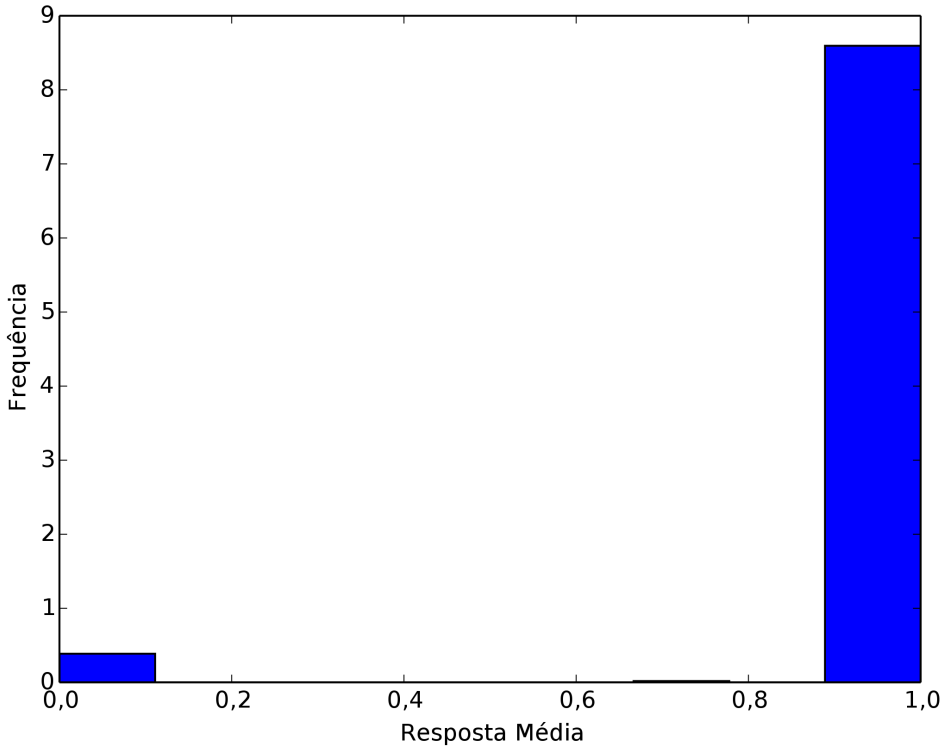


FIGURA 5.1.2. Histograma da resposta média do Q-Learning nas 100 últimas apresentações da tarefa, com parâmetros $p = 0{,}7$, $K = 0$ e $R_e = 0$. $N = 1000$.

o modelo DCB. Comparamos a resposta média dos agentes DCB ao longo de 300 apresentações com $p = 0{,}7$ e K variável de 1 a 5 (o comportamento deste modelo não é bem definido quando $K = 0$). Os resultados podem ser visualizados na Figura 5.5.1. Novamente, conforme aumenta o valor de K, mais lento é o crescimento da resposta média ao longo do tempo.

5.6. Discussão Parcial

Os resultados dos experimentos com agentes computacionais apoiam a nossa proposta. O modelo DCB sempre aprende a perseverar na tarefa de escolha binária repetida e o Q-Learning, quando usamos parâmetros apropriados, também.

Nos Experimentos 1 e 2, vimos que, para o Q-Learning se comportou de maneira mais semelhante aos seres humanos, nunca perseverando do lado em minoria em 300 tentativas, alteramos o valor de R_e de 0 para -1, o que indica que, quando há um erro na previsão do próximo resultado, a recompensa é negativa e não nula. Quando seres humanos fazem a tarefa de escolha binária repetida, eles não aprendem a perseverar do lado em minoria mesmo quando não perdem dinheiro ou pontos pelo erro. Foi proposto que o cérebro atribui um valor negativo (chamado de "erro fictício"), que pode ser considerado como arrependimento, quando o indivíduo percebe que a oportunidade de

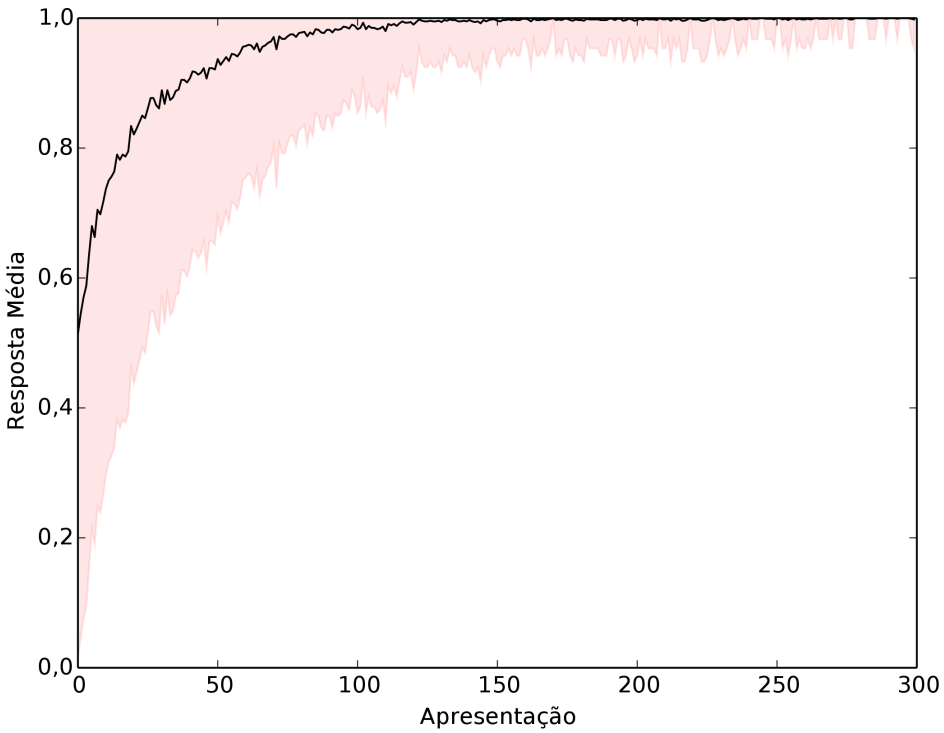


FIGURA 5.2.1. Resposta média do Q-Learning com parâmetros $p = 0{,}7$, $K = 0$ e $R_e = -1$. A área colorida corresponde ao desvio padrão. $N = 1000$.

obter uma recompensa foi perdida [**33**]. Há evidências experimentais de sinais correspondentes ao erro fictício nos núcleos da base [**34, 38, 39**].

Ambos os algoritmos tentam prever o próximo resultado com base nos resultados anteriores, o que pode ser descrito como uma busca por padrões, e verificou-se que quanto maior é o valor de K, menor é a velocidade de aprendizado. Além disso, o Experimento 3 mostra que a velocidade de aprendizado também é afetada pelo valor de p. Os resultados deste experimento apoiam a nossa proposta de que o pareamento de probabilidades observado em seres humanos pode ser simplesmente o resultado de diferentes taxas de aprendizado da estratégia ótima de perseveração. Se o desempenho de cada agente fosse avaliado após, por exemplo, 100 apresentações, poderíamos erroneamente concluir que a resposta média do agente reflete o valor de p e chamar isso de pareamento de probabilidades. No entanto, nenhum dos algoritmos pareia a probabilidade de escolha das alternativas com a probabilidade de elas estarem corretas; eles apenas demoram mais tempo para aprender a perseverar quando os valores de p são menores.

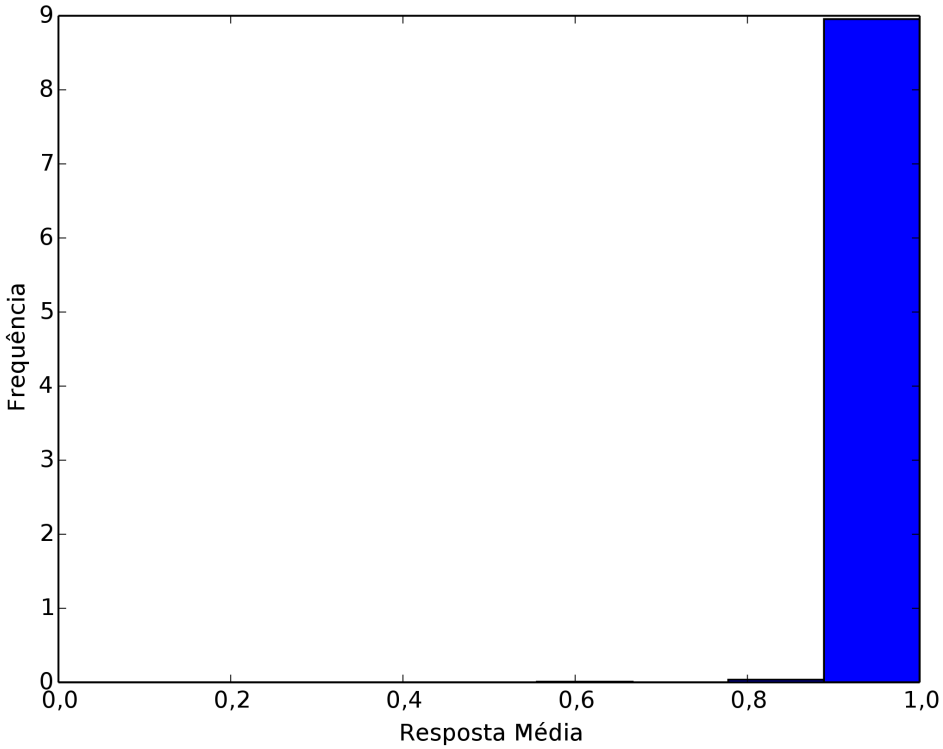


FIGURA 5.2.2. Histograma da resposta média do Q-Learning nas 100 últimas apresentações da tarefa, com parâmetros $p = 0{,}7$, $K = 0$ e $R_e = -1$. $N = 1000$.

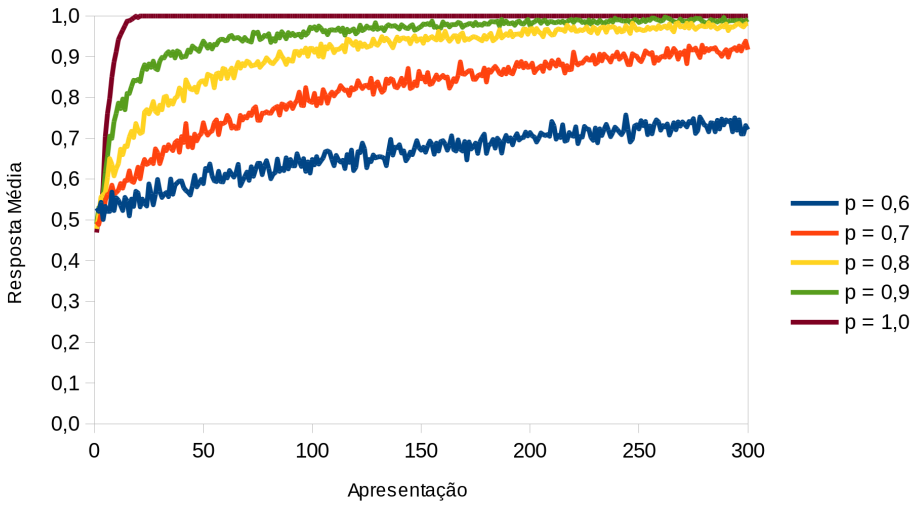


FIGURA 5.3.1. Resposta média do Q-Learning com p variável, $K = 3$ e $R_e = -1$. $N = 1000$.

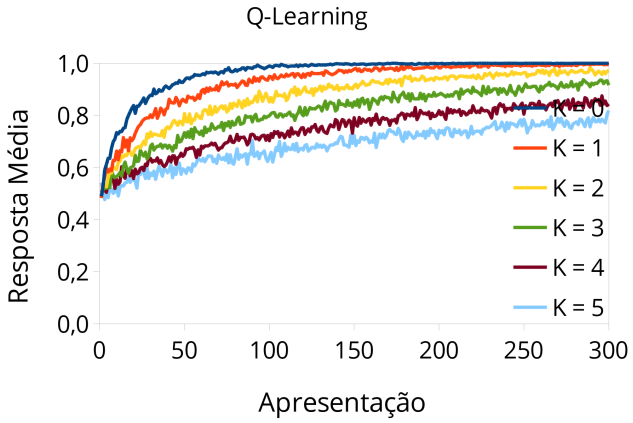


FIGURA 5.4.1. Resposta média de agentes Q-Learning ao longo de 300 apresentações com $p = 0{,}7$ e K variando de 0 a 5. $N = 1000$.

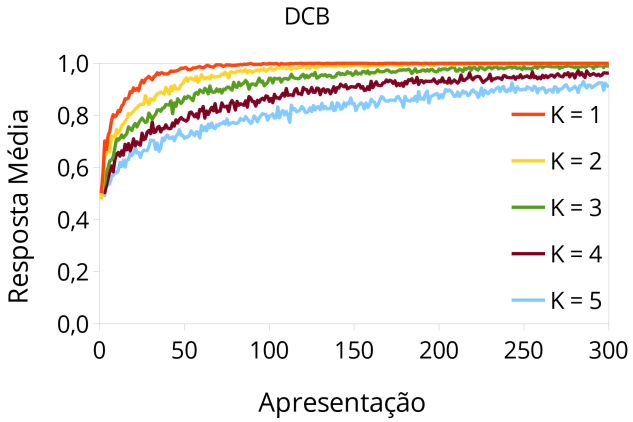


FIGURA 5.5.1. Resposta média de agentes DCB ao longo de 300 apresentações com $p = 0{,}7$ e K variando de 1 a 5. $N = 1000$.

CAPíTULO 6

Experimento com Voluntários Humanos

6.1. Proposta

Neste trabalho, verificou-se que, quando modelos de aprendizado foram usados para realizar a tarefa escolha binária repetida, o aprendizado é mais lento quanto maior é a memória dos resultados das apresentações anteriores. Assim, se os seres humanos tomarem decisões baseados nos resultados das apresentações anteriores por acreditarem que existe um padrão, infere-se que eles também demorarão mais tempo para aprender a perseverar. Essa proposta foi testada através do experimento descrito abaixo.

Mais especificamente, o objetivo do experimento foi avaliar a velocidade de aprendizado em uma tarefa de escolha binária repetida em função da capacidade da memória de curto prazo. Para isso, um grupo de 12 voluntários se sentou em frente ao computador e realizou uma tarefa que alternava uma escolha binária com a memorização e reprodução de uma sequência espacial com comprimento 1 ou 4. A ideia do experimento foi fazer com que os voluntários usassem uma parte menor ou maior de sua memória de curto prazo visuoespacial para armazenar a sequência espacial a ser reproduzida e assim ter sua memória dos resultados anteriores na escolha binária prejudicada.

6.2. Métodos

O voluntário se sentava em frente ao computador e realizava a seguinte tarefa 200 vezes:

(1) Dois quadrados aparecem na tela, um de cada lado (Figura 6.2.1). O voluntário tem 1 segundo para tentar adivinhar dentro de qual quadrado aparecerá um círculo, apertando as teclas A ou S do teclado com a mão esquerda, que se referem ao quadrado da esquerda e ao quadrado da direita, respectivamente. A posição na qual o

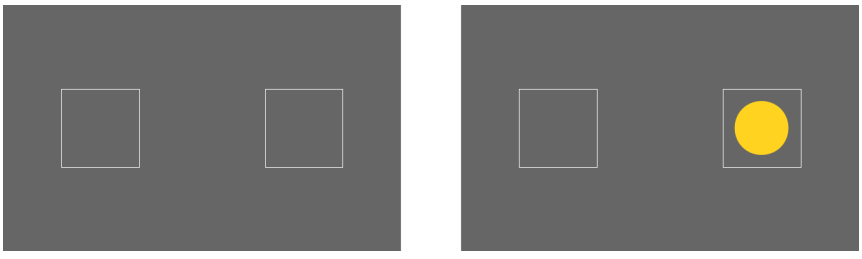


FIGURA 6.2.1. Escolha binária realizada pelos voluntários. Dois quadrados apareciam na tela, um de cada lado (à esquerda). O voluntário tentava adivinhar dentro de qual quadrado apareceria um círculo (à direita).

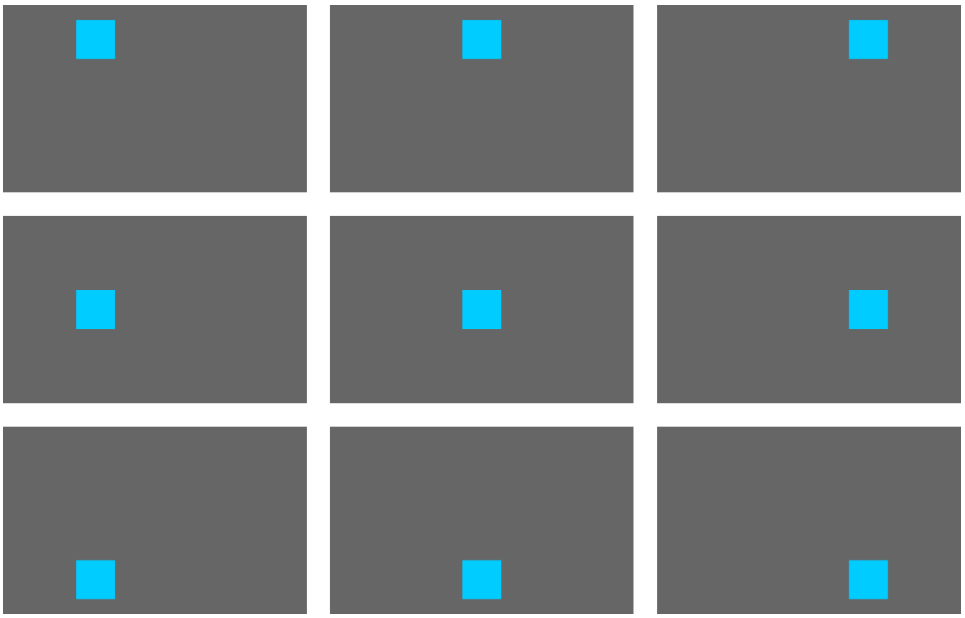


FIGURA 6.2.2. Na tarefa de memorização, em quadrado azul aparece por 500 ms na tela em uma entre nove possíveis posições.

círculo aparece é determinada aleatoriamente, como descrito no Capítulo 1, com $p = 0{,}7$. O lado em maioria é o esquerdo para metade dos voluntários e o direito para a outra metade.

(2) O círculo aparece na tela. Se o voluntário não respondeu, ele perde 10 pontos; se escolheu o lado certo, ele ganha 1 ponto e se escolheu o lado errado, não ganha nem perde pontos. A recompensa obtida é exibida na tela, bem como a recompensa total que o voluntário acumulou até aquele momento, por 1 segundo.

(3) Um quadrado azul aparece por 500 ms na tela em uma entre nove possíveis posições (Figura 6.2.2). Ele repetidamente desaparece e aparece por mais 500 ms em uma posição diferente, formando uma sequência de tamanho 4. A sequência de posições é determinada aleatoriamente a cada apresentação por um sorteio sem reposição.

(4) O voluntário deve memorizar a última posição desta sequência (tarefa fácil) ou todas as posições dela (tarefa difícil) e a seguir digitar sua resposta com a mão direita, usando as teclas 1 a 9 do teclado numérico. Ele tem até 2 segundos para fazer isso e pode ganhar até 1 ponto por sua resposta proporcionalmente à quantidade de elementos da sequência que acertou. A recompensa obtida é exibida na tela, bem como a recompensa total que o voluntário acumulou até aquele momento, por 1 segundo.

As instruções escritas dadas aos voluntários estão no Apêndice.

Os voluntários foram divididos em dois grupos, correspondentes às duas dificuldades da tarefa de memória: fácil ou difícil. O experimento podia durar até 25 minutos. Ao final do experimento, cada voluntário preencheu

um questionário contendo a seguinte pergunta: "Qual foi a sua estratégia para adivinhar o resultado da tarefa de escolha binária?"

O software usado no experimento foi desenvolvido em Python 2.7.8, usando a biblioteca PsychoPy 1.80.07.

6.2.1. Amostra. A amostra foi composta por 12 alunos de graduação da Universidade de São Paulo, os quais faziam os cursos de Física (6 alunos), Design (1 aluno), Letras (1 aluno), Engenharia Civil (1 aluno), Engenharia da Computação (1 aluno), Ciências da Computação (1 aluno) e Medicina (1 aluno).

6.2.1.1. *Critério de Inclusão.* O voluntário deveria ser aluno de graduação e ter entre 18 e 25 anos.

6.2.1.2. *Critério de Exclusão.* Não há.

6.2.1.3. *Riscos.* Não há.

6.2.1.4. *Benefícios.* O voluntário recebeu um chocolate por sua participação.

6.2.2. Análise. Foi analisado o número de vezes que o voluntário escolheu a alternativa em maioria nas últimas 50 apresentações do experimento.

6.3. Resultados esperados

Foi esperado que os voluntários escolhessem a alternativa em maioria em maior proporção quando eles memorizavam as quatro posições do quadrado azul do que quando eles memorizavam somente a última delas.

6.4. Resultados obtidos e discussão parcial

Os resultados dos experimentos podem ser vistos na Figura 6.4.1. Eles foram o contrário do esperado — o grupo de voluntários que realizou a tarefa de memória fácil teve uma resposta média nas 50 últimas apresentações da tarefa de escolha binária superior à do grupo de voluntários que realizou a tarefa de memória difícil. É importante observar também que a resposta média dos voluntários que realizaram a tarefa de memória fácil foi superior à obtida em outros experimentos, como por exemplo a do experimento com 72 alunos discutido anteriormente (Figura 1.0.1).

Além disso, somente um dos voluntários do experimento, pertencente ao grupo que realizou a tarefa de memória fácil, escreveu no questionário que sua estratégia para prever o próximo resultado da tarefa de escolha binária foi procurar padrões. Outros voluntários citaram outras estratégias, como "escolha aleatória" e "chutar sempre o mesmo quadrado". Uma explicação possível para tal fato é a de que os alunos de Exatas, que compõem a maior parte da amostra estudada, têm menos tendência a acreditar na existência de padrões do que os alunos de outras áreas. No entanto, os três alunos da amostra que não eram de Exatas não mencionaram a busca de padrões; o único voluntário que fez isso é aluno de Física.

É possível que nossos resultados tenham sido obtidos, porque a tarefa de memória "fácil" é na verdade difícil e impediu que os resultados anteriores da escolha binária fossem memorizados e que os voluntários procurassem padrões na sequência. Já a tarefa de memória difícil exigiu concentração de-

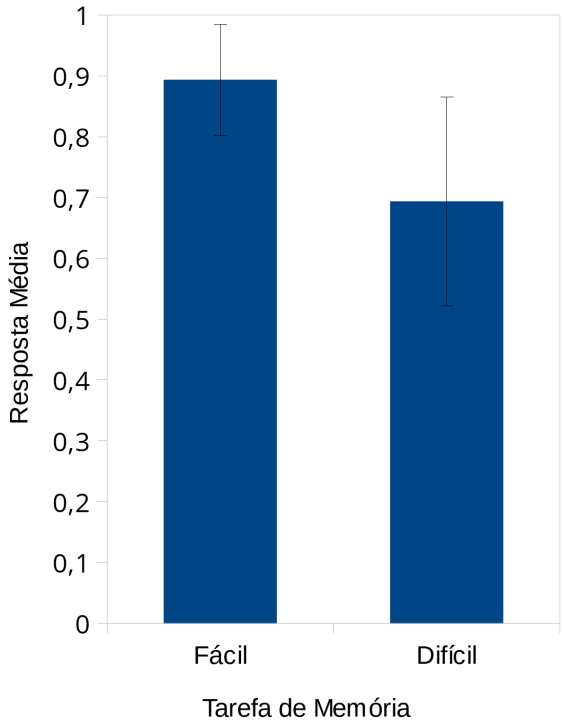


FIGURA 6.4.1. Resposta média (proporção de escolha do lado em maioria) dos voluntários dos dois grupos (tarefa de memória fácil ou difícil) nas últimas 50 apresentações da tarefa de escolha binária repetida.

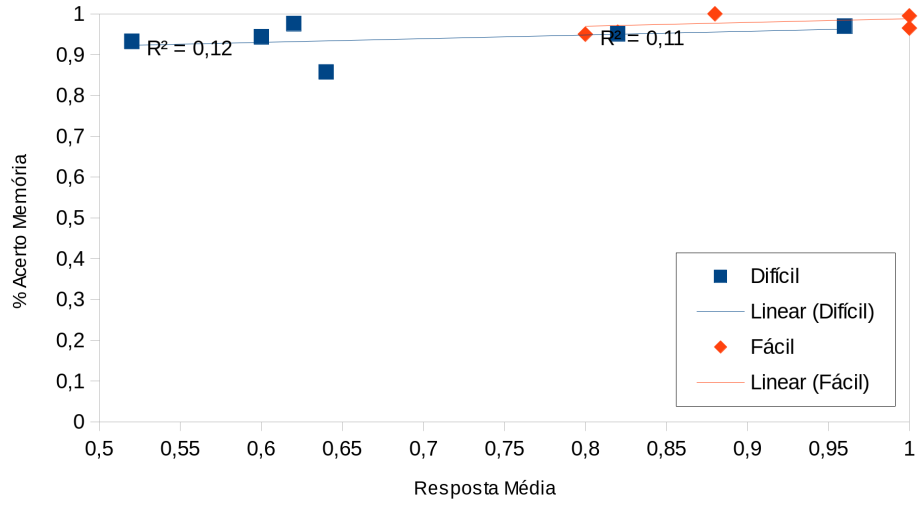


FIGURA 6.4.2. Resposta média dos voluntários dos dois grupos (tarefa de memória fácil ou difícil) nas últimas 50 apresentações da tarefa de escolha binária repetida *versus* a sua proporção de acertos na tarefa de memória.

masiada de muitos voluntários, fazendo com que eles não prestassem a atenção necessária à escolha binária e obtivessem uma resposta média abaixo até mesmo do pareamento de probabilidades, como é mostrado na Figura 6.4.2. Um dos voluntários, inclusive, teve uma resposta média de 0,52, o que poderia ter sido facilmente obtido fazendo escolhas completamente aleatórias, sem se importar com o resultado. No entanto, se a resposta média dos voluntários na tarefa de escolha binária reflete a dificuldade que eles tiveram na tarefa de memória, ela não pode ser evidenciada pela comparação entre desempenho dos voluntários nas duas tarefas: a Figura 6.4.2 mostra que não houve relação entre o desempenho dos voluntários de ambos os grupos na tarefa de memória e a sua resposta média na tarefa de escolha binária.

CAPíTULO 7

Discussão e Conclusões

Neste trabalho, estudamos como os seres humanos tomam decisões em uma tarefa de escolha binária repetida, que se tornou famosa em psicologia e economia, porque, apesar de ela ser muito simples, os seres humanos não aprendem a estratégia ótima de perseveração para maximizar seu ganho, mesmo após centenas de apresentações, ao contrário de outros animais [**1**]. Considera-se que a estratégia empregada por seres humanos é o pareamento de probabilidades; no entanto, nossa proposta é a de que os seres humanos apenas demoram para aprender a perseverar na alternativa em maioria. Uma explicação para tais resultados tinha sido anteriormente sugerida: os seres humanos buscam padrões na sequência de resultados do experimento, uma habilidade que depende da capacidade de sua memória de curto prazo. Baseando-nos nesta ideia, propusemos que quanto maior é o número de resultados armazenados na memória de curto prazo e usados para tomar as próximas decisões, menor é a velocidade de aprendizado. Para testar essa proposta, desenvolvemos agentes computacionais e realizamos experimentos com voluntários humanos.

O modelo de decisão binária que utilizamos assume que os seres humanos armazenam um número fixo, K, de resultados anteriores na memória de curto prazo e usam tais informações para prever o resultado da próxima apresentação de forma independente para cada uma das 2^K combinações de resultados. Tais suposições são simplificações em maior ou menor grau do que de fato ocorre; em particular, é provável que (1) o número de resultados anteriores armazenados não seja fixo, mas varie com o nível de atenção e cansaço do voluntário ao longo da tarefa, e (2) o voluntário use outras informações para prever o resultado da próxima apresentação além dos resultados anteriores, como suas próprias respostas e a frequência observada de cada alternativa. Além disso, é possível que alguns resultados de apresentações anteriores possam ser recuperados de uma memória de longo prazo. No entanto, o modelo que utilizamos, mesmo que simplificado, nos permitiu tirar conclusões interessantes sobre como os seres humanos tomam decisões.

Dois algoritmos de tomada de decisão foram utilizados no desenvolvimento dos agentes computacionais: (1) o Q-Learning, que foi inspirado na maneira pela qual os animais aprendem sequências habituais de ações frequentemente associadas a recompensas no passado, e (2) o modelo DCB, que foi desenvolvido no IFUSP especificamente para realizar a tarefa considerada, usando métodos Bayesianos para procurar padrões. Cinco experimentos com agentes computacionais foram realizados, nos quais foi verificado que ambos os algoritmos aprendem a perseverar, embora as velocidades de aprendizado sejam diferentes. Principalmente, observamos que a velocidade de aprendizado aumenta com o valor de p, o que pode explicar o pareamento

de probabilidades, e diminui com o valor de K, o que apoia diretamente a nossa proposta.

Outro resultado interessante obtido foi que o Q-Learning tem melhor desempenho quando há uma recompensa negativa, ou punição, por cada previsão errada. De fato, foi proposto que o cérebro atribui um valor negativo quando a oportunidade de obter uma recompensa é perdida [**33**]. Nosso resultado ilustra como este mecanismo pode impedir o aprendizado de uma estratégia ruim, neste caso, a perseveração no lado em minoria.

A seguir, fizemos um experimento com doze voluntários humanos, alunos de graduação da USP, cujo objetivo foi avaliar a velocidade de aprendizado em uma tarefa de escolha binária repetida em função da capacidade da memória de curto prazo. Para isso, cada voluntário se sentou em frente ao computador e realizou uma tarefa que alternava uma escolha binária com a memorização e reprodução de uma sequência espacial com comprimento 1 (tarefa de memória fácil) ou 4 (tarefa de memória difícil). A ideia foi fazer com que os voluntários usassem uma parte menor ou maior de sua memória de curto prazo para armazenar a sequência espacial a ser reproduzida e assim terem menos espaço para armazenar os resultados anteriores da escolha binária. De acordo com nossa proposta, era esperado que os voluntários tivessem uma maior resposta média quando memorizavam as quatro posições do quadrado azuis do que quando memorizavam somente a última delas.

Os resultados obtidos, no entanto, foram contrários ao que era esperado — o grupo de voluntários que realizou a tarefa de memória fácil teve uma resposta média na tarefa de escolha binária superior à do grupo de voluntários que realizou a tarefa de memória difícil. Além disso, o grupo que realizou a tarefa de memória fácil teve um comportamento diferente do que é geralmente observado em outros experimentos: a resposta média deles foi maior e somente um dos voluntários disse que sua estratégia para prever o próximo resultado da tarefa de escolha binária foi procurar padrões. A conclusão preliminar que pudemos tirar foi a de que tais resultados foram obtidos porque a tarefa de memória fácil foi na verdade difícil o suficiente para impedir que os resultados anteriores da escolha binária fossem memorizados e que os voluntários procurassem padrões na sequência. Já a tarefa de memória difícil foi tão difícil que eles não puderam se concentrar na tarefa de escolha binária. No entanto, só será possível tirar conclusões mais satisfatórias após a realização de novos experimentos. Três variações do experimento que foi descrito aqui já foram preparadas, mas infelizmente os resultados não puderam ser obtidos a tempo para serem incluídos no presente trabalho.

Podemos, então, concluir que:

(1) Os resultados com agentes computacionais apoiam a nossa proposta de que uma maior capacidade da memória de curto prazo torna o aprendizado mais lento.
(2) Isso poderia explicar por que seres humanos têm um desempenho em geral ruim na tarefa de escolha binária repetida: eles procuram padrões na sequência de resultado e para isso usam os resultados das apresentações anteriores, armazenados na memória de curto prazo, para tentar prever o resultado da próxima apresentação.

(3) O resultado do nosso experimento com voluntários humanos, no entanto, foi contrário ao esperado. É possível que a dificuldade das tarefas que criamos tenha sido inadequada.
(4) Novos experimentos com voluntários humanos devem ser realizados a fim de responder as perguntas que ficaram em aberto.

Parte 2

Subjetiva

CAPíTULO 8

Desafios e Frustrações

Há alguns anos eu ajudei uma aluna de mestrado do ICB em seu projeto, cujo objetivo era descobrir como as estratégias de decisão de um ser humano mudam ao longo da vida; para isso, ela fez experimentos de escolha binária repetida com voluntários de 4 a 70 anos de idade [**40**]. Em sua dissertação, ela concluiu que uma parte das diferenças que ela observou podia ser explicada por uma tendência das crianças a explorar mais do que exploitar e dos idosos a exploitar mais do que explorar.

Inspirada por este trabalho e ciente de que muitos aspectos dos dados coletados não haviam sido abordados, meu objetivo inicial para o TCC era usar Inteligência Artificial para analisar estes resultados com maior profundidade. Tendo um modelo de como os seres humanos tomam decisões, eu pensava em ajustá-los aos dados, estimando valores para os parâmetros, e assim chegar a uma conclusão mais interessante sobre as diferenças entre crianças, jovens e idosos. No entanto, este objetivo sempre me pareceu vago. Não há uma teoria sendo testada ou uma expectativa em relação aos resultados que poderia obter. Experimentos científicos em geral são feitos para testar uma teoria, mas neste caso os experimentos já tinham sido feitos e o meu papel seria o de criar uma teoria *a posteriori*.

Felizmente, já nos passos iniciais do projeto, que consistiram em fazer pesquisas bibliográficas e ler os trabalhos mais relevantes encontrados, tivemos a ideia da proposta atual. Agora, com uma teoria a ser testada em mente, havia um caminho mais claro à frente.

Mas, como em geral se faz, eu cometi o erro de achar as coisas seriam mais simples do que elas de fato acabaram sendo. Para testar os modelos computacionais inicialmente, eu explorei valores arbitrários para os parâmetros, o que levou aos mais diversos resultados. Sendo assim, quais valores, de fato, eu deveria usar? Ao tentar usar dados de experimentos anteriores com seres humanos para estimar esses valores, descobri que os modelos na verdade não se ajustavam bem aos dados. Cheguei a uma conclusão à qual chegaria outras vezes durante a execução do trabalho: as pessoas são mais complicadas do que os nossos modelos matemáticos preveem que elas são. Após pensar muito sobre como resolver este problema, decidi não resolvê-lo e simplesmente usar o modelo como uma demonstração de como a memória afeta a velocidade de aprendizado e não como um modelo fidedigno de como as pessoas de fato tomam decisão.

Novos desafios foram surgindo para criar os experimentos com seres humanos. Como fazer com que as pessoas não se lembrassem do que aconteceu? Que tamanho deveria ter a sequência a ser memorizada para que isso acontecesse? Após testar versões preliminares dos experimentos com alunos de pós-graduação do grupo de Mecânica Estatística do IFUSP, eu notei que, ao

contrário do que é em geral observado, esses alunos em geral assumiam que a sequência binária não seguia um padrão, que ela era aleatória. Disseram-me que "o pessoal da Mecânica Estatística acha que tudo é aleatório" e que portanto eu havia escolhido a amostra errada para o meu experimento. Isso me levou a generalizar a ideia e pensar que talvez o pessoal de Exatas, com melhor formação em Estatística e melhor entendimento de probabilidades, seja de fato diferente do resto da população.

Continuando o desenvolvimento do experimento, na tentativa de estimar o tamanho da sequência necessário para impedir que os voluntários procurem padrões, criei tarefas em que a sequência binária não era aleatória; ao contrário, ela seguia um padrão gerado por uma cadeia de Markov de ordem variada. Após vários testes no qual um amigo e eu fomos voluntários, percebi que ambos estávamos "sabotando" o experimento: nós éramos capazes de descobrir estratégias alternativas para repetir o padrão com poucos erros, mesmo quando não nos lembrávamos de todos os resultados anteriores teoricamente necessários para isso. A conclusão foi que as estratégias das pessoas para encontrar padrões são mais complicadas do que uma cadeia de Markov. Desisti da ideia de usar padrões e voltei à escolha binária simples, aleatória e independente.

Eventualmente eu cheguei a um experimento que me parecia satisfatório para testar a proposta; é o experimento descrito nesta monografia. No entanto, o resultado foi contrário ao esperado e, assim como nos testes anteriores com os alunos de pós-graduação do IFUSP, só um voluntário procurou um padrão na sequência de resultados da tarefa de escolha binária. Dessa vez, nem todos os voluntários eram de Exatas, e os que não eram tiveram resultados muito parecidos, de modo que eu acho que posso descartar esta ideia de que o pessoal de Exatas é diferente e não procura padrões. Para obter mais informações, vou ter que fazer mais experimentos e minha última frustração foi não ter tido tempo de fazê-los antes da entrega final desta monografia.

CAPíTULO 9

Disciplinas Relevantes

Muitas disciplinas do BCC foram relevantes para este trabalho, mas aquelas cuja importância foi mais direta foram:

(1) Inteligência Artificial (MAC0425): é uma matéria optativa e tenho muita sorte de ter tido a ideia de prestá-la, pois foi assim que aprendi sobre o estudo de tomada de decisão em Ciências da Computação e conheci a Prof. Leliane, que orientou este TCC.

(2) Todas as disciplinas de Estatística que prestei (MAE0121, MAE0212, MAE0228), sem as quais eu não teria aprendido a fazer as análises dos meus dados e não entenderia o funcionamento de modelos probabilísticos.

CAPíTULO 10

Próximos Passos

Pretendo continuar a trabalhar neste projeto durante o próximo ano pelo menos. Os próximos passos serão:

(1) Realização de mais experimentos com seres humanos a fim de testar as novas perguntas que surgiram deste trabalho.
(2) Reprodução dos experimentos com humanos usando um número maior de voluntários e recompensa em dinheiro por cada acerto.
(3) Preparação de um manuscrito em inglês para publicação dos resultados em uma revista revisada por pares.

Referências Bibliográficas

[1] VULKAN, N. An Economist's Perspective on Probability Matching. *Journal of Economic Surveys*, v. 14, n. 1, p. 101–118, fev. 2000. ISSN 0950-0804. Disponível em: <http://www.blackwell-synergy.com/links/doi/10.1111/1467-6419.00106>.

[2] Feher da Silva, C. *Abordagem computacional e psicofísica da alocação atencional e tomada de decisão*. 123 p. Tese (Tese de Doutorado) — Universidade de São Paulo, 2011. Disponível em: <http://www.teses.usp.br/teses/disponiveis/42/42137/tde-04102011-163948/en.php>.

[3] PARDUCCI, A.; POLT, J. Correction vs. noncorrection with changing reinforcement schedules. *Journal of Comparative and Physiological Psychology*, v. 51, n. 4, p. 492–495, 1958. ISSN 0021-9940. Disponível em: <http://content.apa.org/journals/com/51/4/492>.

[4] GRAF, V.; BULLOCK, D. H.; BITTERMAN, M. E. Further experiments on probability-matching in the pigeon. *Journal of the Experimental Analysis of Behavior*, v. 7, n. 2, p. 151–157, mar. 1964. ISSN 0022-5002. Disponível em: <http://www.pubmedcentral.gov/articlerender.fcgi?artid=1404312>.

[5] BEHREND, E. R.; BITTERMAN, M. E. Probability-Matching in the Fish. *The American Journal of Psychology*, v. 74, n. 4, p. 542–551, 1961. Disponível em: <http://www.jstor.org/stable/1419664>.

[6] KAREEV, Y.; LIEBERMAN, I.; LEV, M. Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, v. 126, n. 3, p. 278–287, 1997. ISSN 1939-2222. Disponível em: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-3445.126.3.278>.

[7] WOLFORD, G.; MILLER, M. B.; GAZZANIGA, M. S. The Left Hemisphere's Role in Hypothesis Formation. *The Journal of Neuroscience*, v. 20, n. RC64, p. 1–4, 2000.

[8] WOLFORD, G. et al. Searching for Patterns in Random Sequences. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, v. 58, n. 4, p. 221–228, dez. 2004. ISSN 1196-1961. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/15648726 http://vitallongevity.utdallas.edu/cnl/wp-content/uploads/2014/04/Wolford_etal_2004_CanJExpPsychol.pdf http://doi.apa.org/getdoi.cfm?doi=10.1037/h0087446>.

[9] UNTURBE, J.; COROMINAS, J. Probability matching involves rule-generating ability: a neuropsychological mechanism dealing with probabilities. *Neuropsychology*, v. 21, n. 5, p. 621–30, set. 2007. ISSN 0894-4105. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/17784810>.

[10] GAISSMAIER, W.; SCHOOLER, L. J.; RIESKAMP, J. Simple predictions fueled by capacity limitations: when are they successful? *Journal of experimental psychology. Learning, memory, and cognition*, v. 32, n. 5, p. 966–82, set. 2006. ISSN 0278-7393. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16938040>.

[11] GAISSMAIER, W.; SCHOOLER, L. J. The smart potential behind probability matching. *Cognition*, Elsevier B.V., v. 109, n. 3, p. 416–22, dez. 2008. ISSN 1873-7838. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/19019351>.

[12] GAISSMAIER, W.; SCHOOLER, L. J. An ecological perspective to cognitive limits: Modeling environment-mind interactions with ACT-R. *Judgment and Decision Making*, v. 3, n. 3, p. 278–291, 2008. Disponível em: <http://journal.sjdm.org/bn7/bn7.html>.

[13] KOEHLER, D. J.; JAMES, G. Probability matching in choice under uncertainty: intuition versus deliberation. *Cognition*, Elsevier B.V., v. 113, n. 1, p. 123–7, out. 2009. ISSN 1873-7838. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/19664762>.

[14] SHANKS, D. R.; TUNNEY, R. J.; MCCARTHY, J. D. A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, v. 15, n. 3, p. 233–250, jul. 2002. ISSN 0894-3257. Disponível em: <http://doi.wiley.com/10.1002/bdm.413>.

[15] NIV, Y. et al. Evolution of Reinforcement Learning in Uncertain Environments: A Simple Explanation for Complex Foraging Behaviors. *Adaptive Behavior*, v. 10, n. 1, p. 5–24, jan. 2002. ISSN 1059-7123. Disponível em: <http://adb.sagepub.com/cgi/doi/10.1177/10597123020101001>.

[16] HARDY-VALLéE, B. Artificial life, natural rationality and probability matching. In: *Artificial Life, 2007. ALIFE'07. IEEE Symposium on*. IEEE, 2007. p. 123–129. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4218877>.

[17] SETH, a. K. Modeling Group Foraging: Individual Suboptimality, Interference, and a Kind of Matching. *Adaptive Behavior*, v. 9, n. 2, p. 67–89, jun. 2001. ISSN 1059-7123. Disponível em: <http://adb.sagepub.com/cgi/doi/10.1177/105971230200900204>.

[18] SETH, A. K. The ecology of action selection: insights from artificial life. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, v. 362, n. 1485, p. 1545–58, set. 2007. ISSN 0962-8436. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2440771&tool=pmcentrez&rendertype=abstract>.

[19] nA, D. E. A.; SCHRATER, P. Structure Learning in Human Sequential Decision-Making. *PLoS computational biology*, v. 6, n. 12, p. e1001003, jan. 2010. ISSN 1553-7358. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2996460&tool=pmcentrez&rendertype=abstract>.

[20] BEREBY-MEYER, Y.; EREV, I. On Learning To Become a Successful Loser: A Comparison of Alternative Abstractions of Learning Processes in the Loss Domain. *Journal of Mathematical Psychology*, v. 42, n. 2-3, p. 266–286, jun. 1998. ISSN 00222496. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0022249698912147>.

[21] PUTERMAN, M. L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. [S.l.]: John Wiley and Sons, Inc., 2005. 684 p.

[22] GLIMCHER, P. W.; FEHR, E. (Ed.). *Neuroeconomics: Decision Making and the Brain*. Second. [S.l.]: Academic Press, 2014. 560 p.

[23] GLäSCHER, J. et al. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, Elsevier, v. 66, n. 4, p. 585–95, maio 2010. ISSN 1097-4199. Disponível em: <http://www.cell.com/neuron/abstract/S0896-6273(10)00287-4 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2895323&tool=pmcentrez&rendertype=abstract>.

[24] JOZEFOWIEZ, J.; STADDON, J. E. R. Operant Behavior. In: MENZEL, R.; BYRNE, J. H. (Ed.). *Learning and Memory: A Comprehensive Reference, Volume 1: Learning Theory and Behaviour*. [S.l.]: Academic Press, 2008. cap. 1.06, p. 75–102.

[25] SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction*. Second. [S.l.]: MIT Press, 2012.

[26] RANGEL, A.; CAMERER, C.; MONTAGUE, P. R. A framework for studying the neurobiology of value-based decision making. *Nature reviews. Neuroscience*, v. 9, n. 7, p. 545–56, jul. 2008. ISSN 1471-0048. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/18545266>.

[27] YIN, H. H.; KNOWLTON, B. J. The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, v. 7, n. 6, p. 464–76, jun. 2006. ISSN 1471-003X. Disponível em: <http://dx.doi.org/10.1038/nrn1919 http://www.ncbi.nlm.nih.gov/pubmed/16715055>.

[28] MINK, J. W. The Basal Ganglia. In: SQUIRE, L. et al. (Ed.). *Fundamental Neuroscience*. Fourth. [S.l.]: Academic Press, 2012. cap. 30, p. 653–676.

[29] KANDEL, E. R. et al. (Ed.). *Principles of Neural Science*. Fifth. [S.l.]: McGraw-Hill, 2013. 1760 p. ISBN 978-0071390118.

[30] SMITH, E. E.; JONIDES, J.; KOEPPE, R. A. Dissociating Verbal and Spatial Working Memory Using PET. *Cerebral Cortex*, v. 6, n. 1, p. 11–20, jan. 1996. ISSN 1047-3211. Disponível em: <http://www.cercor.oxfordjournals.org/cgi/doi/10.1093/cercor/6.1.11>.

[31] D'ESPOSITO, M. et al. Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research*, v. 7, n. 1, p. 1–13, jul. 1998. ISSN 09266410. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0926641098000044>.

[32] AWH, E.; JONIDES, J. Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, v. 5, n. 3, p. 119–126, mar. 2001. ISSN 13646613. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S136466130001593X>.

[33] MONTAGUE, P. R.; KING-CASAS, B.; COHEN, J. D. Imaging valuation models in human choice. *Annual review of neuroscience*, v. 29, p. 417–48, jan. 2006. ISSN 0147-006X. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16776592>.

[34] LOHRENZ, T. et al. Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, n. 22, p. 9493–8, maio 2007. ISSN 0027-8424. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1876162&tool=pmcentrez&rendertype=abstract>.

[35] LI, J.; DAW, N. D. Signals in human striatum are appropriate for policy update rather than value prediction. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 31, n. 14, p. 5504–11, abr. 2011. ISSN 1529-2401. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3132551&tool=pmcentrez&rendertype=abstract>.

[36] FIORILLO, C. D.; TOBLER, P. N.; SCHULTZ, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science (New York, N.Y.)*, v. 299, n. 5614, p. 1898–902, mar. 2003. ISSN 1095-9203. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/12649484>.

[37] DOBAY, E. S. *Complexidade e tomada de decisão*. 81 p. Tese (Dissertação de Mestrado) — Universidade de São Paulo, 2014.

[38] CHIU, P. H.; LOHRENZ, T. M.; MONTAGUE, P. R. Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nature neuroscience*, v. 11, n. 4, p. 514–20, abr. 2008. ISSN 1097-6256. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/18311134>.

[39] BüCHEL, C. et al. Ventral striatal signal changes represent missed opportunities and predict future choice. *NeuroImage*, v. 57, n. 3, p. 1124–30, ago. 2011. ISSN 1095-9572. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/21616154>.

[40] VICTORINO, C. G. *Estudo do desenvolvimento de estratégias decisionais em escolhas binárias repetidas*. 97 p. Tese (Dissertação de Mestrado) — Universidade de São Paulo, 2012.

Apêndice

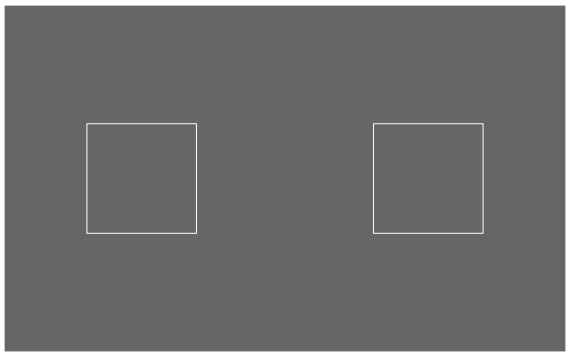
Instruções

O objetivo deste experimento é estudar memória e tomada de decisão em seres humanos. Pela sua participação, você ganhará um chocolate, além de estar contribuindo para o avanço da ciência. A duração do experimento é de até 25 minutos.

O experimento é formado por duas tarefas, uma de escolha binária e uma de memória, que se intercalam em 200 repetições.

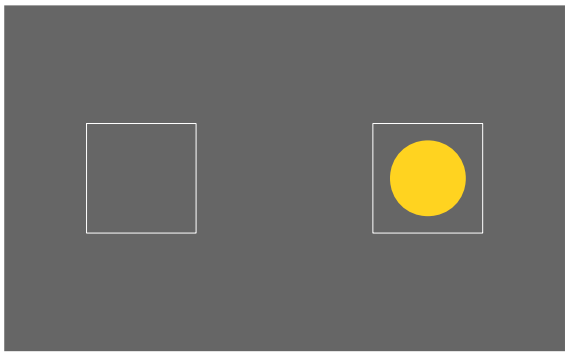
Tarefa de Escolha Binária

Dois quadrados serão apresentados em lados opostos da tela e você deverá prever qual deles contém uma recompensa.



Pressione a tecla A se você deseja escolher o quadrado esquerdo e a tecla S se você deseja escolher o quadrado direito. Use a mão esquerda para fazer a escolha. Você tem 1 segundo para responder; caso não responda, perderá 10 pontos.

A escolha correta será indicada por uma bola dentro do respectivo quadrado.



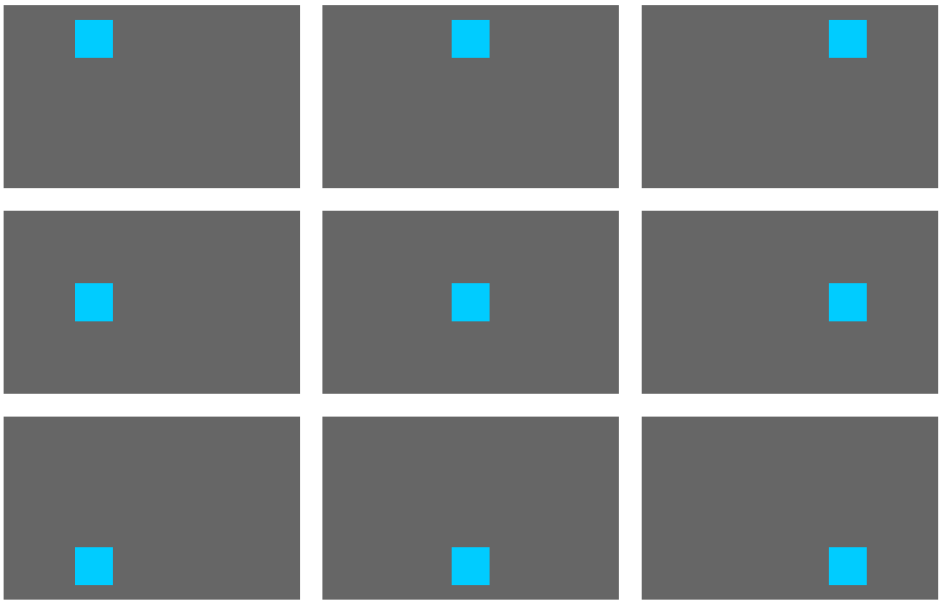
Se você escolheu corretamente, ganhará 1 ponto. Caso tenha escolhido o quadrado errado, não ganhará nem perderá nada.

Repare que se você não responder, perderá 10 pontos, mas se responder errado, não perderá nada. Assim, é sempre melhor responder do que não responder, mesmo que seja errado.

A seguir, uma mensagem aparecerá na tela indicando quantos pontos você ganhou ou perdeu por sua escolha e a pontuação total que você tem até agora.

Tarefa de Memória

Um quadrado aparecerá na tela em uma entre nove posições possíveis.



Ele desaparecerá e aparecerá em outro destes lugares três vezes, formando uma sequência de tamanho 4. Você deve memorizar a última posição e digitá-la a seguir, usando as teclas 1 a 9 do teclado numérico. Use a mão direita para dar sua resposta. Você tem até 2 segundos para fazer isso. Você ganhará 1 ponto se sua resposta estiver correta e não ganhará nem perderá nada se não responder ou se a sua resposta estiver incorreta.

Importante

1. As suas ações e sequência a ser memorizada não influenciam os resultados da tarefa de escolha binária – eles são predeterminados. Lembre-se disso, pois caso contrário seu desempenho será ruim e os seus resultados terão que ser descartados.
2. Antes de o experimento começar, você poderá treinar as tarefas sem limite de tempo.

Boa sorte!

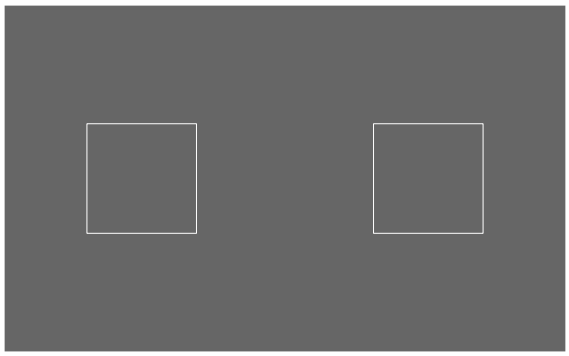
Instruções

O objetivo deste experimento é estudar memória e tomada de decisão em seres humanos. Pela sua participação, você ganhará um chocolate, além de estar contribuindo para o avanço da ciência. A duração do experimento é de até 25 minutos.

O experimento é formado por duas tarefas, uma de escolha binária e uma de memória, que se intercalam em 200 repetições.

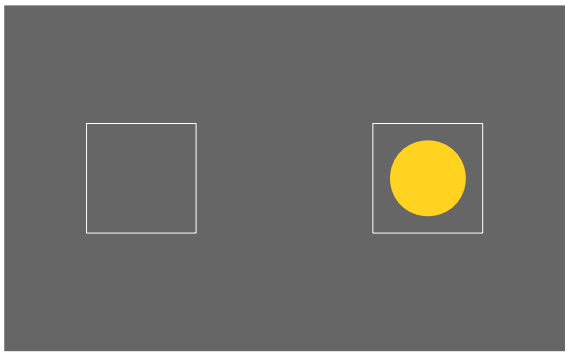
Tarefa de Escolha Binária

Dois quadrados serão apresentados em lados opostos da tela e você deverá prever qual deles contém uma recompensa.



Pressione a tecla A se você deseja escolher o quadrado esquerdo e a tecla S se você deseja escolher o quadrado direito. Use a mão esquerda para fazer a escolha. Você tem 1 segundo para responder; caso não responda, perderá 10 pontos.

A escolha correta será indicada por uma bola dentro do respectivo quadrado.



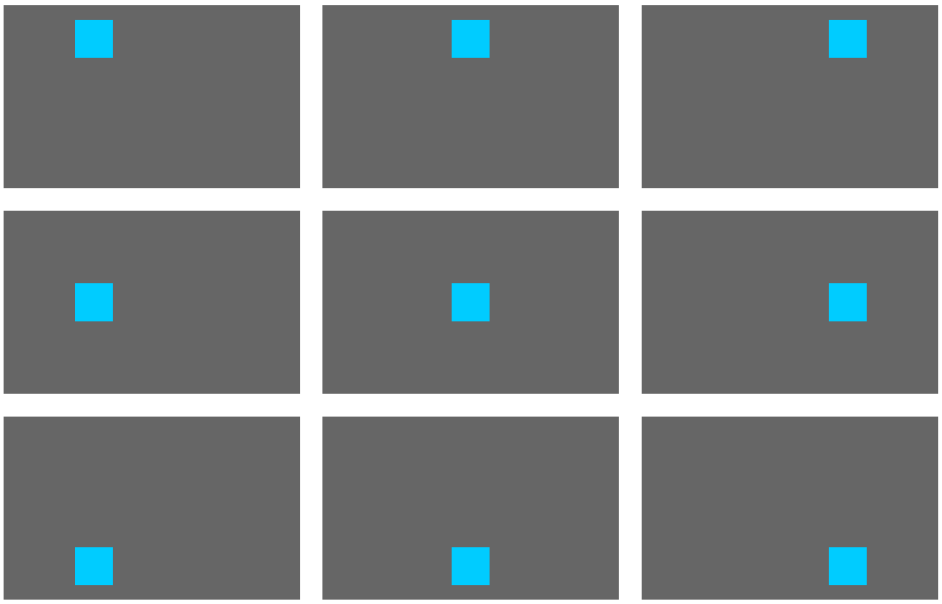
Se você escolheu corretamente, ganhará 1 ponto. Caso tenha escolhido o quadrado errado, não ganhará nem perderá nada.

Repare que se você não responder, perderá 10 pontos, mas se responder errado, não perderá nada. Assim, é sempre melhor responder do que não responder, mesmo que seja errado.

A seguir, uma mensagem aparecerá na tela indicando quantos pontos você ganhou ou perdeu por sua escolha e a pontuação total que você tem até agora.

Tarefa de Memória

Um quadrado aparecerá na tela em uma entre nove posições possíveis.



Ele desaparecerá e aparecerá em outro destes lugares três vezes, formando uma sequência de tamanho 4. Você deve memorizar esta sequência e digitá-la a seguir, usando as teclas 1 a 9 do teclado numérico. Use a mão direita para dar sua resposta. Você tem até 2 segundos para fazer isso; quando este intervalo tiver terminado, somente o que você tiver digitado até aquele momento será levado em conta. Você poderá ganhar até 1 ponto, dependendo do quão acurada for a sua resposta.

Importante

1. As suas ações e sequência a ser memorizada não influenciam os resultados da tarefa de escolha binária – eles são predeterminados. Lembre-se disso, pois caso contrário seu desempenho será ruim e os seus resultados terão que ser descartados.
2. Antes de o experimento começar, você poderá treinar as tarefas sem limite de tempo.

Boa sorte!