Controvérsia
issn 1808-5253

Acessando informações sobre estados mentais epistêmicos por meio de respostas eletrofisiológicas: uma análise de como a eletroencefalografia pode elucidar questões da Filosofia da Mente

Accessing information about epistemic mental states through electrophysiological responses: an analysis of how electroencephalography can elucidate issues from Philosophy of Mind

Ricardo Augusto Perera
Universidade do Vale do Rio dos Sinos
ricardoperera@outlook.com
http://lattes.cnpq.br/3437045512392329

Sofia Inês Albornoz Stein
Universidade do Vale do Rio dos Sinos/CNPq
siastein@me.com
http://lattes.cnpq.br/2045729600668270

Resumo
Neste artigo serão expostos pressupostos teóricos e hipóteses filosóficas que podem ser testadas utilizando eletroencefalografia. Utilizando a técnica de ERP's (Event-Related Potentials), que faz uso de dados eletroencefalográficos, espera-se ser possível encontrar padrões de respostas eletrofisiológicas que "carreguem informação" acerca da posse ou ausência de estados mentais epistêmicos, enquanto os participantes estiverem lendo sentenças filosóficas que terminam de modo verdadeiro ou falso. A ideia é incluir os participantes no grupo dos filósofos (alunos de pós-graduação em Filosofia) ou no grupo dos não-filósofos (alunos de cursos não relacionados) e, pressupondo uma maior familiaridade com fatos filosóficos entre os filósofos, analisar-se-á o sinal evocado nas duas condições (i.e. frases falsas e frases verdadeiras) nos dois grupos, principalmente nos eletrodos localizados nas regiões central e parietal. O componente de interesse do ERP será o N400 (deflexão negativa cuja latência ocorre aproximadamente 400 milissegundos após um estímulo), que está principalmente associado à dificuldade de integrar uma palavra – em termos semânticos – ao contexto em que ela aparece. Da mera resposta cerebral a expectativas semânticas satisfeitas (N400 atenuado) ou violadas (N400 acentuado), conjectura-se ser possível compreender os papéis causais das crenças epistêmicas subjacentes que estão a modular os sinais e o modo como estas são recrutadas durante a leitura para predizer os próximos itens léxicos.

Palavras-chave
Mindreading; Filosofia da Mente; Eletroencefalografia; N400.

Abstract
This article will expose theoretical assumptions and philosophical hypotheses that can be tested using electroencephalography. Using the ERP technique (Event-Related Potentials), which makes use of EEG data, it is expected to be possible to find patterns of electrophysiological responses that "carry information" about the possession or absence of epistemic mental states, while participants are reading philosophical sentences ending in a true or false way. The idea is to include participants in the group of philosophers (graduate students in philosophy) or the group of non-philosophers (students of unrelated courses) and assuming greater familiarity with philosophical facts among philosophers, will be analyzed the signal evoked in the two conditions (i.e. false and true sentences) in both groups, especially in electrodes located in the central and parietal regions. The component of interest of the ERP will be the N400 (negative deflection whose latency occurs about 400 milliseconds after a stimulus), which is mainly

associated with the difficulty of integrating a word - in semantic terms – into the context in which it appears. From the mere brain response to satisfied semantic expectations (attenuated N400) or violated (strong N400), it is conjectured to be possible to understand the causal role of the underlying epistemic beliefs that are modulating the signals and how they are recruited during reading to predict the next lexical items.
Keywords
Mindreading; Philosophy of Mind; Electroencephalography, N400.

O presente trabalho apresenta uma proposta de experimento cujo objetivo é compreender os papéis causais de estados mentais representacionais (e.g. crenças e conhecimento) ao serem recrutados por mecanismos preditivos, assim como medir o tempo que o cérebro demora para detectar a falsidade de uma sentença. Utilizando a técnica de *Event-Related Potentials* (ERP's[1]), espera-se encontrar padrões de respostas eletrofisiológicas-enquanto os participantes estiverem lendo certas sentenças filosóficas verdadeiras ou falsas. Para tanto, Analisar-se-á o sinal evocado nas duas condições (i.e. frases falsas e frases verdadeiras) nos eletrodos posicionados nas regiões central e parietal. O componente de interesse do ERP será o N400 – uma deflexão negativa cuja latência ocorre aproximadamente 400 milissegundos após um estímulo –, que está principalmente associado à dificuldade de integrar uma palavra, em termos semânticos, ao contexto no qual ela aparece.

Na seção 1 é exposta uma breve revisão da literatura em que variações no sinal do componente N400 são encontradas em violações de expectativa semântica e epistêmica. Na seção 2 é apresentado o desenho de um possível experimento em que a amplitude N400 é modulada pelo recrutamento de crenças específicas. Na seção 3 argumenta-se que os dados do experimento podem complementar teorias filosóficas sobre estados mentais representacionais. As premissas assumidas, hipóteses propostas e resultados esperados são formalizados na seção 4.

1. N400, forte ou atenuado

Kutas e Hillyard (1980) descobriram o N400 ao apresentarem sentenças (palavra por palavra) cujos últimos itens eram semanticamente congruentes (e.g. "I shaved off my mustache and *beard*") ou incongruentes (e.g. "I take coffee with cream and *dog*"). Esperando um P300 (uma positividade cuja latência ocorre entre 300 a 800 milissegundos após o estímulo, inversamente correlacionada com a probabilidade subjetiva de um evento) nas anomalias semânticas, os autores se depararam, acidentalmente, com o N400. Desde então, centenas de pesquisas utilizaram o N400 como medida independente, em áreas como processamento da linguagem, objetos, faces, ações e gestos, cognição matemática, memória semântica e de reconhecimento, assim como no estudo de uma variedade de transtornos adquiridos ou relacionados aos desenvolvimento (Kutas e Federmeier, 2011, p. 622). Hagoort et al. (2004) verificaram a presença de N400 também em sentenças que violavam o conhecimento de fatos do mundo: as sentenças falsas nada tinham de semanticamente anômalo (i.e. não continham propriedades inerentemente em conflito, sendo portanto plausíveis), mas apresentavam palavras que

[1] A técnica de Event-Related Potentials (ERP's) consiste em recortar segmentos do EEG bruto, gerando a média de vários *trials* de uma mesma condição. Deste modo é possível atenuar o ruído aleatório presente em cada *trial* (resultante de atividades cerebrais dissociadas dos estímulos apresentados e também de outras interferências externas), retendo apenas aquilo que todos os *trials* possuem em comum (Luck, 2005).

conflitavam com o conhecimento dos participantes. Um holandês, por exemplo, que possui o conhecimento de que os trens da Holanda são brancos, apresenta N400 ao ler "amarelo" após a sentença incompleta "Os trens da Holanda são *amarelos*". A latência do componente evocado pelas frases falsas não diferiu da latência encontrada em violações semânticas, apesar de ser menos negativa (porém, apresentando uma negatividade significativa em relação às sentenças verdadeiras). De acordo com os mesmos autores, verificar que uma palavra torna a sentença falsa não é um processo que ocorreria posteriormente à detecção de que a sentença é semanticamente plausível (não havendo duas etapas, uma de compreensão do estado de coisas representado pela sentença e outra de "olhar para o mundo" para verificar seu valor de verdade).

À luz destes achados, conjectura-se que se o conhecimento de um fato produz um N400 forte em um indivíduo que o possui e que lê algo que frustra sua expectativa (havendo uma resposta atenuada quando lê algo que é esperado dado suas crenças), então seria a princípio possível extrair informações acerca da posse ou ausência de conhecimentos/crenças por meio da análise deste padrão de respostas eletrofisiológicas. Uma pessoa que possua um conjunto C de conhecimentos, ao ler frases cujas palavras finais são compatíveis com as proposições do conjunto C, deverá apresentar um N400 fraco. Já as frases cujas palavras finais são incompatíveis com o conjunto C deverão evocar um N400 acentuado. Por outro lado, um indivíduo que ignore todos os fatos descritos pelo conjunto C não apresentará diferença de N400 entre as condições Verdadeiro e Falso, uma vez que tanto as sentenças falsas quanto as verdadeiras serão para ele igualmente plausíveis. Na ausência de conhecimento de que "A casa do rei do Mali é *vermelha*", por exemplo, não se esperará a palavra "vermelha" após a leitura de "A casa do rei do Mali é [...]". Não obstante, nossa expectativa semântica é de certa forma restrita: esperamos uma palavra que seja plausível tanto em termos semânticos como sintáticos. "Vermelha" e "azul" não diferirão quanto a seus efeitos eletrofisiológicos, mas "homem" ou "salgada" – anomalias semânticas – produzirão N400 mesmo em ignorantes: não é necessário olhar para o mundo para detectar a incongruência da frase "A casa do rei do Mali é *homem* [ou *salgada*]", uma vez que as propriedades "ser homem" e "ser salgado" não são predicáveis de entidades como casas.

Quanto maior a probabilidade subjetiva de determinada palavra P aparecer em uma sentença incompleta (*cloze probability*), maior a negatividade do N400 quando, ao invés de P, outra palavra aparecer (Kutas e Federmeier, 2011). A negatividade é muito forte quando ocorrem anomalias semânticas, sendo mais atenuado o N400 de palavras inesperadas cujos conceitos se relacionam de alguma forma com o conceito da palavra esperada, como "pêra" aparecendo no lugar de "maçã" (Federmeier e Kutas, 1999).

2. Uma proposta de experimento

Vimos, na seção anterior, que o N400 está associado à frustração de uma expectativa semântica específica. Em função disso, preparamos um experimento que conta com um conjunto de sentenças capazes de provocar, cada uma delas, uma expectativa forte quanto à palavra final, em indivíduos que possuem conhecimento relativo à filosofia (filósofos). As sentenças em questão constituem uma espécie de "denominador comum" que idealmente inclui somente aquilo que todos os participantes da categoria 'filósofos' conhecem. Adicionalmente, cada sentença está redigida de forma tal que apenas uma única palavra, a última, a torna verdadeira ou falsa. "Teeteto é um diálogo escrito por *Platão*" é um exemplo de sentença desse tipo, pois

apenas a palavra "Platão" torna a sentença verdadeira, e apenas "Platão" será esperada (expectativa semântica) por aqueles que deste fato estão cientes.

As sentenças descritas no parágrafo anterior são do tipo "A...B". Nessa formalização, (A) designa a primeira parte da sentença e (B) o nome que aparece como palavra final. Para expressar que nenhum outro nome além de B torna a sentença incompleta verdadeira, adicionou-se uma cláusula de unicidade: A...B Λ ¬(∃θ) A...θ Λ ¬(θ=B)[2]. A figura a seguir (Fig. 1) detalha a estrutura da formalização:

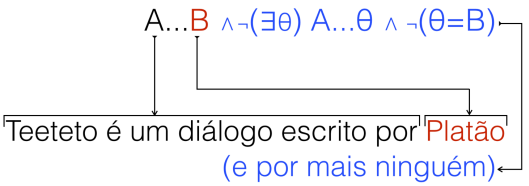

Fig. 1 - Exemplo de sentença do tipo A...B Λ ¬(∃θ) A...θ Λ ¬(θ=B)

Cumpre notar que "Platão escreveu o diálogo Teeteto", apesar de ser acerca do mesmo estado de coisas da sentença "Teeteto é um diálogo escrito por Platão", não restringe consideravelmente a expectativa do filósofo, uma vez que há outras palavras, além de "Teeteto", que tornam a sentença verdadeira, visto que Platão escreveu vários diálogos. Assim, a sentença "Platão escreveu o diálogo Teeteto" não é uma sentença do tipo A...B Λ ¬(∃θ) A...θ Λ ¬(θ=B), mas do tipo A...B Λ (∃θ) A...θ Λ ¬(θ=B):

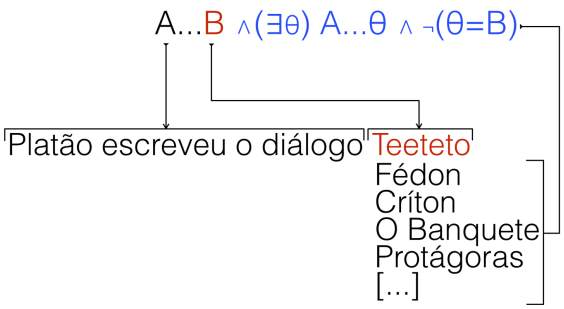

Fig. 2 - Exemplo de sentença do tipo A...B Λ (∃θ) A...θ Λ ¬(θ=B)

O experimento prevê que as 80 sentenças relativas à filosofia – todas elas do tipo A...B Λ ¬(∃θ) A...θ Λ ¬(θ=B), que praticamente todo filósofo sabe serem verdadeiras – serão apresentadas em um monitor, palavra por palavra, aos participantes. 40 sentenças terminarão com palavras diferentes das esperadas (tornando, portanto, as sentenças falsas), porém mantendo-as plausíveis (e.g. "Zenão de Eléia foi discípulo de Tales"). As 80 sentenças pertencem ao conjunto aqui referido como @, resultante da interseção entre o conjunto das sentenças que todo filósofo sabe serem verdadeiras – o conjunto P – e o conjunto das sentenças do tipo A...B Λ ¬(∃θ) A...θ Λ ¬(θ=B) – o conjunto Q:

[2] Uma objeção que pode ser feita a esta formalização é a de que termos correferenciais, como "Túlio" e "Cícero", podem ambos tornar uma sentença incompleta como "Saber envelhecer é um livro de" verdadeira. O mesmo ocorreria com descrições definidas (e.g. "o discípulo de Sócrates e professor de Aristóteles"). Como a escolha deste tipo específico de sentença tem apenas uma finalidade prática (criar nos participantes uma expectativa por uma única palavra), esses problemas se tornam aqui irrelevantes.

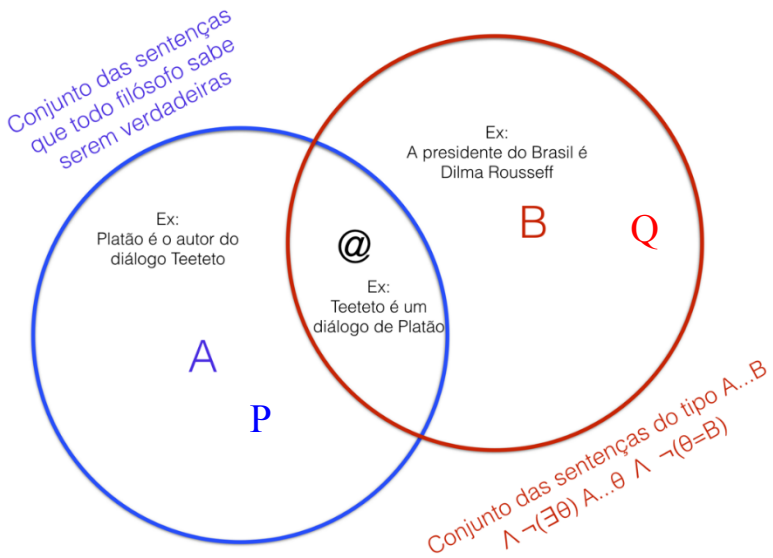

Fig. 3 - Conjunto @ (P ∩ Q): o conjunto das sentenças do tipo A...B Λ ¬(∃θ) A...θ Λ ¬(θ=B) que todo filósofo sabe serem verdadeiras

Nossa proposta de experimento prevê considerar dois grupos dicotômicos, filósofos e não-filósofos, que idealmente conhecem e ignoraram todos os fatos descritos pelas sentenças do conjunto @, respectivamente. Outra possibilidade é comparar os sinais dos ERP's com o desempenho individual dos participantes em um questionário escrito desenhado para avaliar o grau de conhecimento de fatos filosóficos. Deste modo, conjecturamos, seria possível verificar se o conhecimento medido em um teste explícito está correlacionado com um maior *N400 Effect* (a *differential wave* obtida ao subtrair-se o sinal evocado pelas palavras corretas do sinal evocado pelas palavras incorretas, representado a diferença entre as duas condições) no experimento realizado com EEG. A linha de tendência abaixo ilustra o resultado esperado com participantes com um conhecimento parcial do conjunto @:

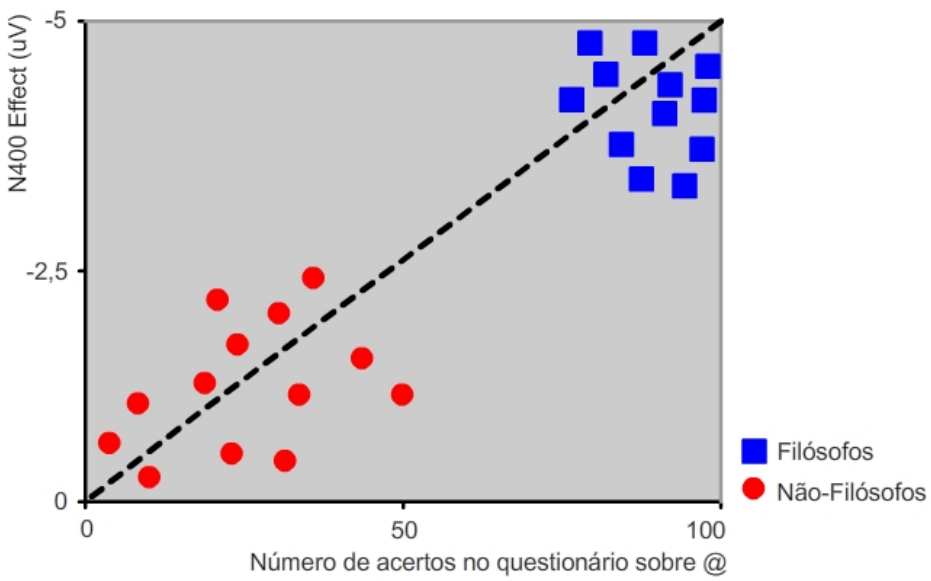

Fig. 4 - Correlação esperada entre medição de conhecimento filosófico em questionário escrito e N400 Effects.

Essa predição descansa sobre a seguinte suposição: apenas quem tem conhecimento dos fatos aludidos pelas sentenças incompletas criará expectativas quanto à palavra seguinte. A predição de que após "Teeteto é um diálogo escrito por..." aparecerá a palavra "Platão" será, no participante filósofo, de altíssima precisão. É possível que uma antecipação deste grau não se limite à expectativa semântica, mas que esta também influencie a expectativa léxica: não apenas penso em Platão, o pensador que escreveu o referido diálogo, mas espero a palavra "Platão" escrita na tela, com determinadas letras e comprimento, e nenhuma outra coisa exceto isso.

Em virtude da magnitude, tal sucesso preditivo elicitará, no participante filósofo, uma resposta atenuada ao ler as palavras esperadas (tanto no P300, que vem sendo associado à expectativa léxica, quanto no N400). Perante uma palavra que falseia a sentença, espera-se que o filósofo apresente um N400 bastante negativo e, possivelmente, um P300 mais positivo:[3]

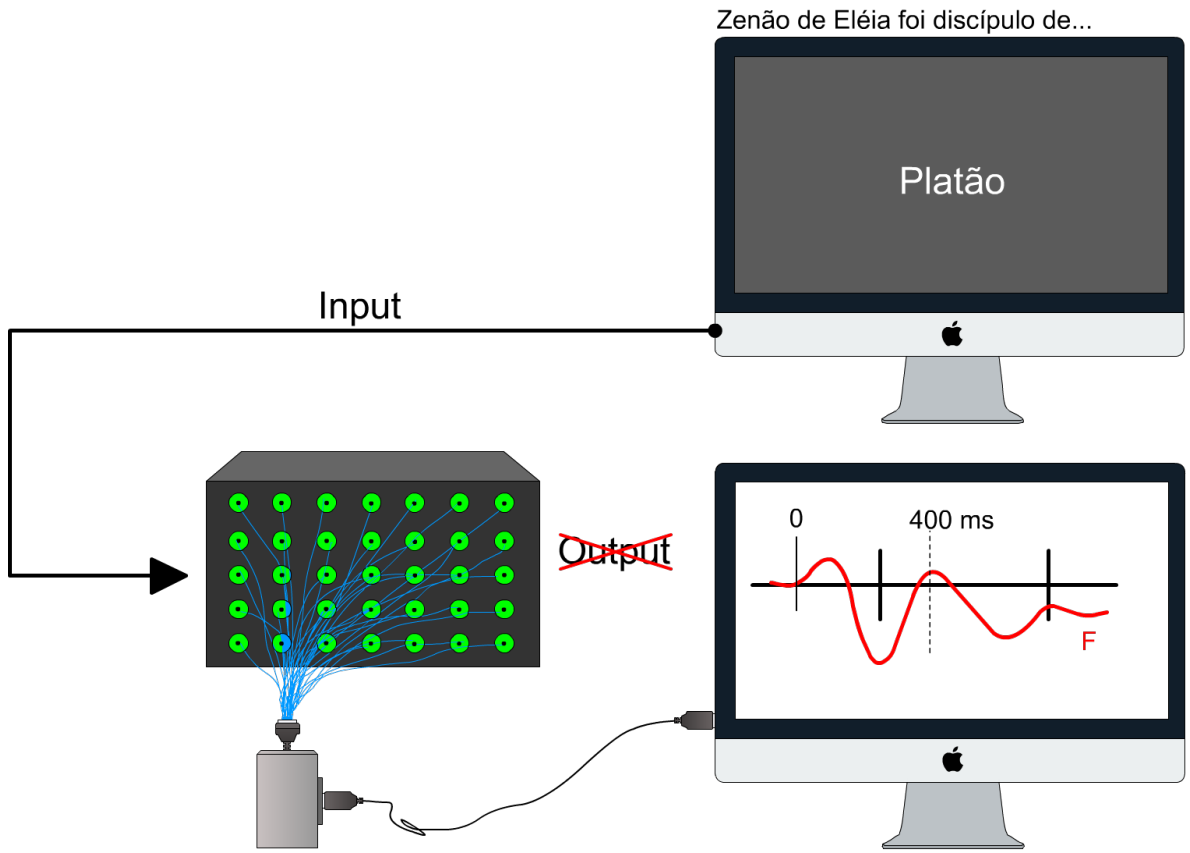

Fig. 5 - Representação do sinal esperado pelos participantes que tiverem conhecimento das sentenças do conjunto @, ao lerem a palavra incorreta: um N400 forte.

[3] Uma possibilidade é a de que o P300 e o N400 se sobreponham e, pelo fato de o primeiro ser positivo e o segundo negativo, venham a se anular. Talvez este problema – caso de fato ocorra – possa ser eliminado ao se utilizar sentenças que não constranjam tanto a expectativa do leitor, não gerando expectativa léxica, como "Descartes foi um filósofo medieval/contemporâneo". O preço desta abordagem que evolve menor precisão é a obtenção de um *N400 Effect* mais fraco (visto que a amplitude do N400 é inversamente proporcional à probabilidade subjetiva de uma palavra), requerendo um maior número de participantes para se obter dados significativos.

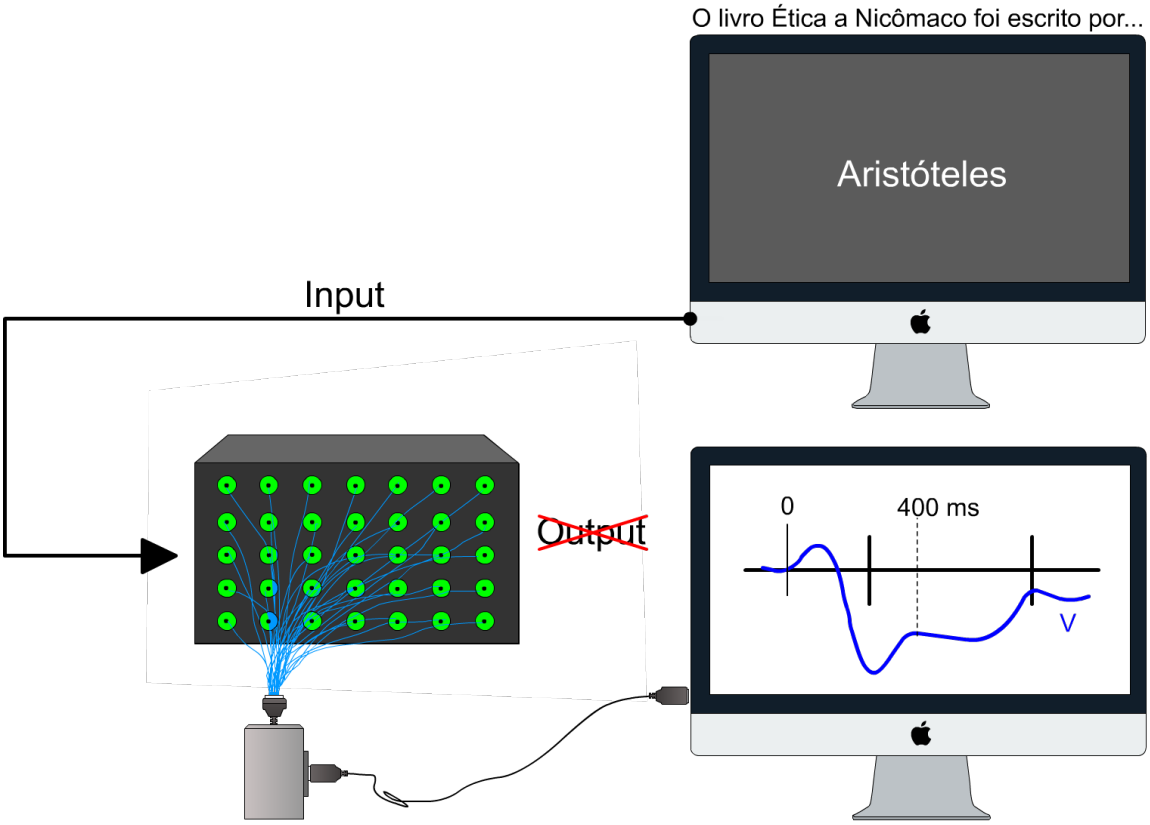

Fig. 6 - Representação do sinal esperado pelos participantes que tiverem conhecimento das sentenças do conjunto @, ao lerem a palavra correta: um N400 fraco.

O participante não-filósofo, que ignore todos os fatos descritos pelas sentenças do conjunto @, não deverá apresentar ERP's com significativa diferença entre as condições verdadeiro e falso (N400 *Effect*), pois suas expectativas não irão muito além de restrições semânticas e sintáticas, e talvez algumas associações rasas (e.g. uma obra que possua um título visivelmente grego ser escrita por um autor de sobrenome alemão poderia elicitar N400). A Figura 7 ilustra os resultados esperados nas duas condições (Verdadeiro e Falso) nas duas categorias (Filósofos e Não-filósofos). Já a Figura 8 apresenta os *N400 Effects* esperados em ambas as categorias. Para análise estatística prevemos calcular a amplitude média dos *N400 Effects* de cada indivíduo do intervalo entre 300 e 500 milissegundos após a apresentação do estímulo crítico.[4]

[4] As etapas de análise dos dados são as seguintes: a) filtrar o sinal do EEG bruto de 0.1 a 30Hz; b) recortar 900 milissegundos (-100 a 800) dos 80 trials do experimento (o instante 0 é o da apresentação da palavra crítica), gerando 40 segmentos de cada condição; c) detectar artefatos e canais ruins, eliminando canais inadequados e trials contendo artefatos ou um número excessivo de canais ruins; d) calcular e média das condições Verdadeiro e Falso; e) re-referenciar para a média aritmética dos eletrodos posicionados nos mastoides direito e esquerdo; f) utilizar os 100 milissegundos que precedem o estímulo crítico como baseline; g) calcular a differential wave resultante da subtração do ERP das palavrar corretas do ERP das palavras incorretas (=N400 Effect); h) Calcular a amplitude média dos N400 Effects de cada indivíduo do intervalo entre 300 e 500 milissegundos após a apresentação do estímulo crítico; i) Realizar análise estatística para verificar se a diferença entre os N400 Effects dos filósofos e dos não-filósofos é significativa. Os participantes das duas categorias que não apresentarem N400 Effect no experimento de anomalias semânticas serão excluídos da análise no experimento sobre conhecimento filosófico.

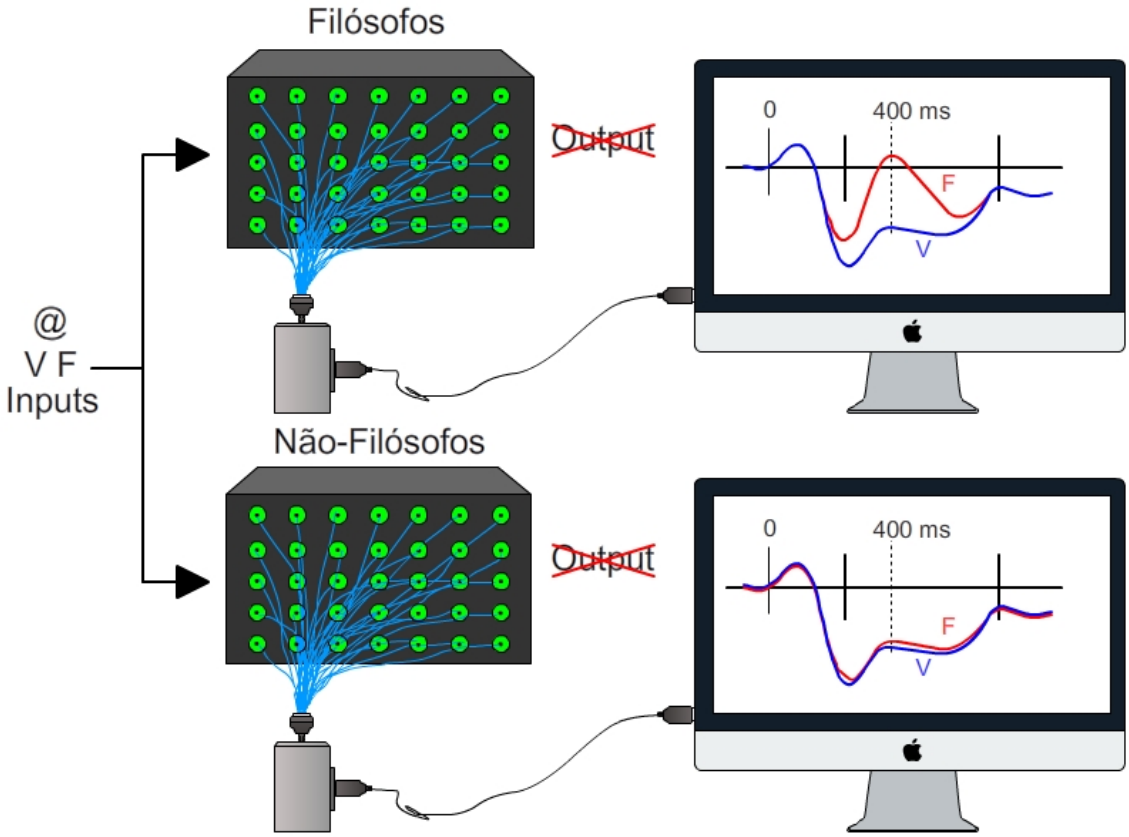

Fig. 7 - Comparação das médias esperadas dos sinais dos filósofos e não-filósofos, nas sentenças verdadeiras e falsas.

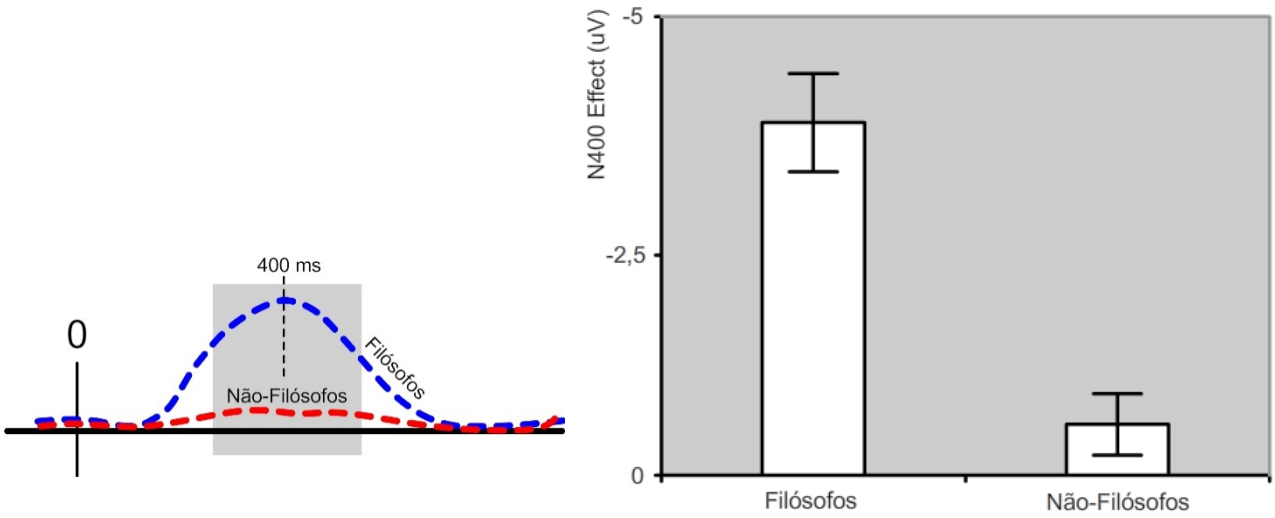

Fig. 8 - Intervalo de análise das amplitudes médias (na figura esquerda, em cinza) e barras de erro esperadas dos N400 Effects dos Filósofos e Não-Filósofos (direita).

O experimento em comento não extrai propriamente dos participantes informações acerca de seus estados mentais, uma vez que tais informações já estão sendo pressupostas: os participantes da categoria 'filósofos' são indivíduos que confirmaram antecipadamente que possuem conhecimento filosófico; os participantes da categoria 'não-filósofos' são indivíduos

que afirmam antecipadamente ignorar a maioria dos assuntos filosóficos. A partir disso, espera-se encontrar N400 *Effects* significativamente diferentes entre os dois grupos, dadas as razões supramencionadas. Se isso ocorrer, será a princípio possível inverter a situação e, em um segundo experimento, ensaiar um passo além: em vez de escolher grupos de conhecedores e ignorantes de um conjunto C de fatos e verificar, utilizando a técnica aqui exposta, se há ou não diferença no *N400 Effect*, poderemos inferir com satisfatória probabilidade de acerto se alguém ignora ou conhece um conjunto C de fatos apenas por meio do *N400 Effect* elicitado em quem lê sentenças que aludem ao conjunto C e que terminam de modo correto ou incorreto (uma inferência reversa). Neste caso, teríamos uma espécie de extração de informação não-pressuposta acerca da mentalidade alheia, permitindo uma "leitura de mentes" que prescinda da observação de outputs comportamentais. Nesse segundo experimento, o participante receberia inputs sem ser instruído a realizar nenhuma tarefa a não ser a leitura passiva do que lhe é exposto no monitor. Dependendo do modo como seu cérebro responde a esses inputs – seriam novamente sentenças do tipo $A...B \Lambda \neg(\exists\theta)\ A...\theta \Lambda \neg(\theta=B)$ – conseguir-se-ia inferir crenças que estão a modular as respostas eletrofisiológicas.[5]

3. Como o experimento proposto poderia ajudar a compreender os papés funcionais de estados mentais intencionais

Teorias filosóficas sobre o significado dos termos que se referem a estados mentais intencionais (como crenças e desejos) costumam variar de eliminativismos (e.g. Churchland, 1981; 1988) a realismos (e.g. Fodor, 2002), havendo posições intermediárias entre estes dois extremos. Como as outras mentes são inacessíveis – aqui a escolha de uma "*black box*" como analogia é proposital –, há dificuldade em fixar a referência dos termos mentais que atribuímos às outras pessoas. Um behaviorista lógico, por exemplo, reduz a fala acerca do mental a uma fala acerca de padrões comportamentais (e.g. Ryle, 2000): a "crença de que vai chover" pode ser definida como uma disposição comportamental do tipo "se Pedro enxergar nuvens negras se aproximando" (input), então "Pedro irá recolher as roupas do varal, fechar as janelas da casa e sair com um guarda-chuva" (output). Se Pedro exibir tal padrão comportamental, então Pedro acredita que irá chover. Mas ao "atribuir" a Pedro esta crença o behaviorista lógico não está a postular alguma entidade inobservável que estaria dentro da *black box*, pois ter uma crença, para ele, é tão-somente apresentar (ou possuir a disposição de apresentar) certos comportamentos em determinadas condições de input. Já David Lewis (1970a; 1970b; 1972; 1980) – um funcionalista – define os estados mentais como entidades reais existentes dentro da *black box*, porém inteiramente definidas por meio do modo como as mesmas se relacionam com inputs e outputs observáveis: uma crença é uma entidade teorética postulada pela teoria [folk] psicológica que é denotada e definida por meio da função que realiza, isto é, o modo como interage com as entidades observáveis – inputs e outputs. Diferente da atitude do behaviorista, o funcionalista postula a existência de algo, e lhe faz referência por meio de uma descrição definida como "o estado, qualquer que seja, que é causado pela visão de nuvens negras se aproximando, interage com n crenças e desejos, causando o comportamento de recolher as roupas do varal, etc.". A circularidade desta prática – como o desiderata de definir uma

[5] Aqui foi utilizado o termo "inferir" num sentido fraco e pragmático. Na verdade, a inferência reversa constitui uma afirmação do consequente – uma falácia –, sendo o salto dos ERP's para as crenças não-válido. Contudo, pode-se considerar que a hipótese de que Pedro sabe que os fatos aludidos pelas sentenças do conjunto @ são o caso (ou uma porcentagem considerável de @) obtém satisfatório grau de corroboração caso as palavras finais corretas das sentenças de @ evoquem em Pedro N400 atenuado e as incorretas N400 acentuado. Se os ERP's de Pedro nas duas condições não apresentarem diferença considerável, então a hipótese é falseada (neste último caso a inferência será válida num sentido estrito, pois teremos uma negação do consequente, conforme é demonstrado no argumento da Fig. 17).

entidade mental por meio de apenas entidades observáveis que, ao definí-la, não consegue evitar a menção a outras entidades mentais inobserváveis – não parece constituir um grande problema, pois as demais entidades mentais referidas na definição também sofrem um mesmo destino. Mas apesar do avanço de aceitar entidades intermediárias entre inputs e outputs, o funcionalista o realiza mantendo a *black box* inteiramente opaca, não se tendo acesso a nenhum dado de seu interior que sinalize a existência de crenças, muito menos o modo como atuariam em processos cognitivos que se mantêm invisíveis do lado de fora. A presente proposta de utilizar eletrodos para captar sinais eletrofisiológicos provenientes do interior da *black box*, como até aqui demostrada, seria capaz não só de detectar a existência/inexistência de crenças (indiretamente, pelo modo como modulam as respostas dos ERP's) prescindindo da verificação de outputs comportamentais, mas também de obter dados empíricos sobre como estes estados mentais integram processos cognitivos mais complexos, como ao "serem utilizados" por um *forward model*[6] para predizer o próximo input quando lemos sentenças relacionadas a estes mesmos estados mentais. O fato de o tempo que nosso cérebro leva pra detectar uma violação de crença/conhecimento ser igual ao tempo que leva para perceber uma anomalia semântica é um excelente exemplo de como a eletroencefalografia pode expandir nosso conhecimento sobre entidades mentais de um modo interdito à tradicional abordagem filosófica. De modo semelhante a um funcionalista, seria igualmente possível fixar a referência de termos mentais por meio de descrições definidas/funcionais, só que, em vez de outputs comportamentais, façam uso do sinal eletroencefalográfico proveniente do interior da *black box*. O conhecimento de que o estado de coisas p é o caso poderia ser denotado como "aquela coisa, qualquer que seja, que, ao ser utilizada por um *forward model*, modula a resposta eletrofisiológica de indivíduos que estejam a ler sentenças incompletas que se referem a p, resultando em N400 atenuado quando a palavra final for correta em relação a p e N400 acentuado quando for incorreta".[7] A hipótese de que nosso cérebro utiliza um *forward model* para antecipar inputs relativos a conhecimentos/crenças poderia explicar o fato de a latência do N400 se manter inalterada nas violações de conhecimento em relação às violações semânticas. Em vez de primeiramente compreendermos o estado de coisas descrito pela sentença, de modo passivo (*bottom-up*), analisando se os conceitos empregados estão ou não em conflito, para depois verificarmos seu valor de verdade à luz de nossas crenças sobre o mundo (o que constituiria uma etapa posterior, resultando em N400 de maior latência), estaríamos, no entanto, a adotar uma postura muito mais profeticamente engajada (*top-down*), antevendo o próximo input antes de sua ocorrência, à luz não só de restrições conceituais, mas também de nossas crenças (tanto as explicitamente representadas como as que estão implícitas nas primeiras).

A estratégia aqui utilizada é a seguinte: primeiramente, a partir dos padrões de respostas eletrofisiológicas encontrados na literatura para anomalias semânticas e violação do conhecimento, conceber uma possível explicação dos processos que estão a ocorrer internamente na *black box* para, a partir do mecanismo conjecturado, predizer novos resultados empíricos que possam vir a corroborá-lo ou refutá-lo. Se há um *forward model* operante enquanto lemos sentenças, o mesmo teria acesso a nossas crenças e memória semântica, gerando suas expectativas à luz destas. A Figura 9 esboça um modelo rudimentar dos supostos processos intermediários entre inputs e ERP's em um indivíduo que possua o conhecimento de que Fédon é um diálogo de Platão (uma crença do conjunto @):

[6] Um mecanismo que prediz o estado futuro de um sistema (Pickering e Clark, 2014).
[7] Ou, aceitando a intencionalidade do estado mental: "aquela coisa, qualquer que seja, *que contém informação acerca do fato p* e que, ao ser utilizada por um *forward model*, modula a resposta eletrofisiológica de indivíduos que estejam a ler sentenças incompletas que se referem a p, resultando em N400 atenuado quando a palavra final for correta em relação a p e N400 acentuado quando for incorreta".

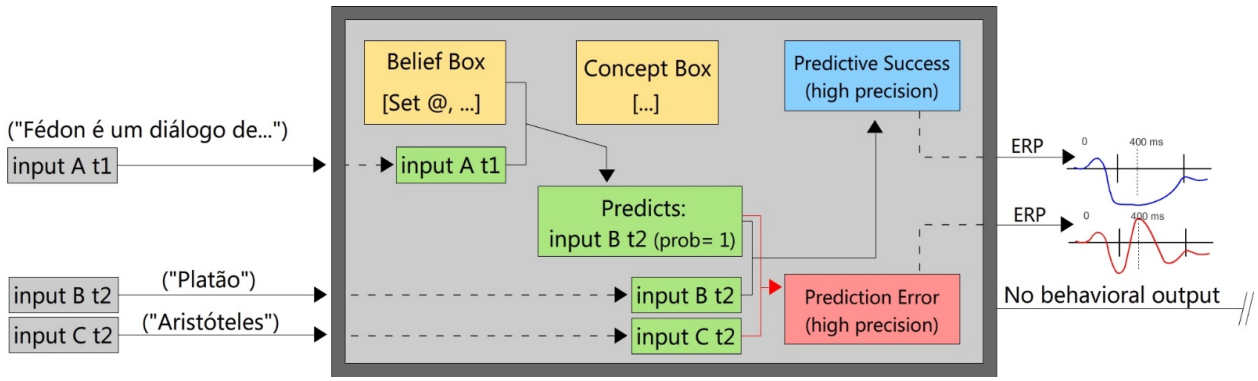

Fig. 9 - Modelo de processo cognitivo que modularia os ERP's dos filósofos em uma condição preditiva de alta precisão.

A leitura da frase incompleta "Fédon é um diálogo de" geraria no indivíduo que sabe que Fédon é um diálogo de Platão uma expectativa de alta precisão de que o próximo input será a palavra "Platão". A *cloze probability* da palavra "Platão" é, em quem está ciente do fato, igual a 1, visto que nenhuma outra palavra é capaz de tornar a sentença verdadeira. Mas se a frase incompleta "Platão escreveu o diálogo" for apresentada ao filósofo, ainda haverá uma expectativa, porém não a de uma palavra específica, mas de uma palavra indeterminada dentro de um escopo conhecido (Fédon, Teeteto, Críton, etc.). Neste caso, a predição não comporta muita precisão quanto ao *token* que aparecerá, porém proíbe o aparecimento de um *token* que não seja de um tipo determinado (diálogo de Platão):

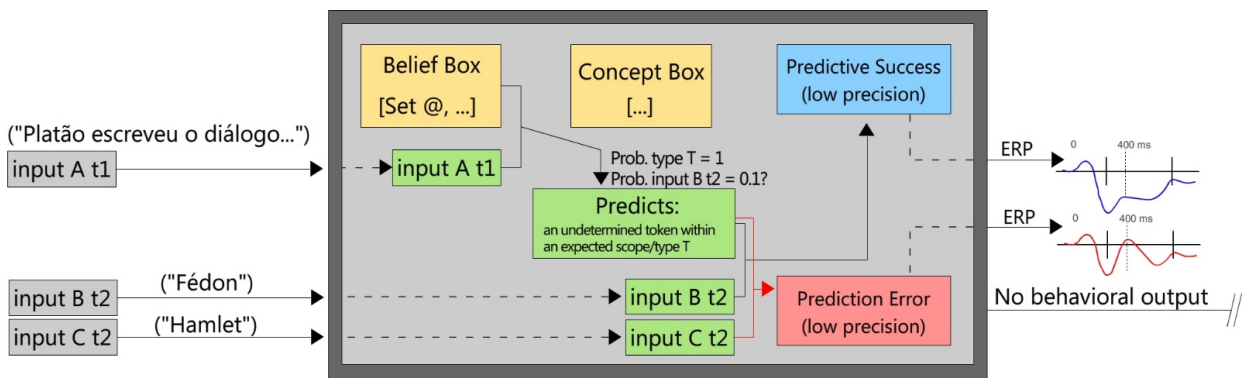

Fig. 10 - Modelo de processo cognitivo que modularia os ERP's dos filósofos em uma condição preditiva de baixa precisão.

Espera-se um maior N400 em conhecedores quando a palavra incorreta estiver numa condição de alta precisão ("Fédon é um diálogo de...") em relação a uma condição de baixa precisão ("Platão escreveu o diálogo...") (Lau, Holcomb e Kuperberg, 2013). Entretanto, como fora mencionado, existe a possibilidade de a expectativa léxica da condição de alta precisão resultar em maior P300 quando violada, encobrindo o N400. O exemplo da Figura 10 provavelmente produziria resultados discrepantes entre as duas condições mesmo em não-filósofos, pois apesar de estes não estarem a esperar um item impreciso de uma disjunção precisa ("Fédon" v "Teeteto" v "Críton" v "O Banquete"...), o conhecimento de um fato não-filosófico (que Hamlet é uma peça de Shakespeare) seria suficiente para causar supresa ao ler-se

"Hamlet" após "Platão escreveu o diálogo" (poder-se-ia dizer que a crença de que Platão não escreveu Hamlet está implícita na crença explícita de que Hamlet é uma peça escrita unicamente por Shakespeare). Mas um não-filósofo que desconheça o fato citado na condição de alta predição ("Fédon é um diálogo de") esperará tão-somente um nome indefinido, dado seu *background knowledge* e restrições semânticas e sintáticas impostas pelo contexto, não elicitanto ERP's diferenciados quando uma palavra incorreta – porém a ele plausível – aparecer:

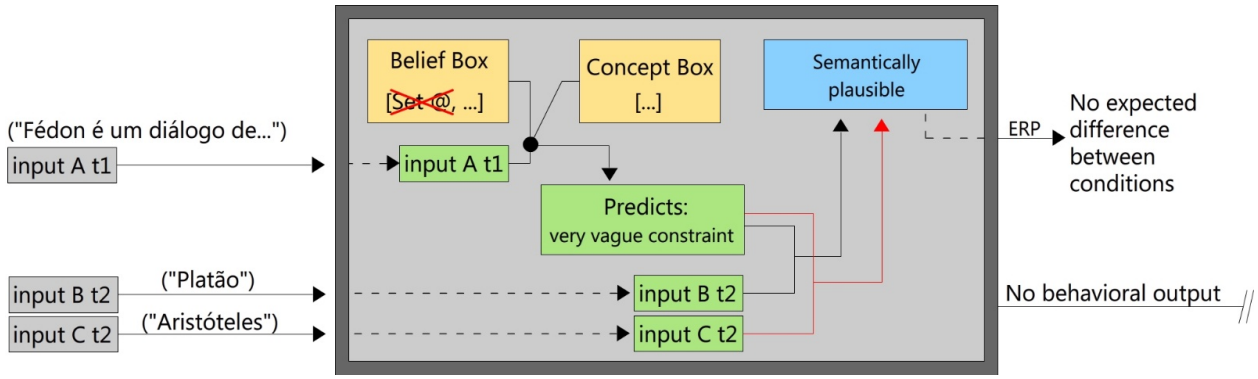

Fig. 11 - Modelo de processo cognitivo que modularia os ERP's dos não-filósofos em uma condição preditiva de alta precisão.

Quando a tarefa do leitor exige que ele apenas compreenda o estado de coisas que é gradualmente descrito, sendo irrelevante seu valor de verdade, a ocorrência de uma anomalia semântica é detectada apenas devido a incompatibilidades entre os conceitos empregados. A expectativa, ainda que vaga, não dependerá da crença de que algum estado de coisas seja o caso. Se lemos a frase incompleta "Minha irmã se casou com um", esperamos uma palavra que expresse um conceito cujos exemplares estão aptos a se casar com irmãs (e.g. "argentino", "escritor", "amigo"). Se a palavra crítica tiver como referência uma entidade que instancie uma propriedade incompatível com a relação matrimonial (e.g. "nariz", "tomate", "número"), então a expectiva semântica será violada:

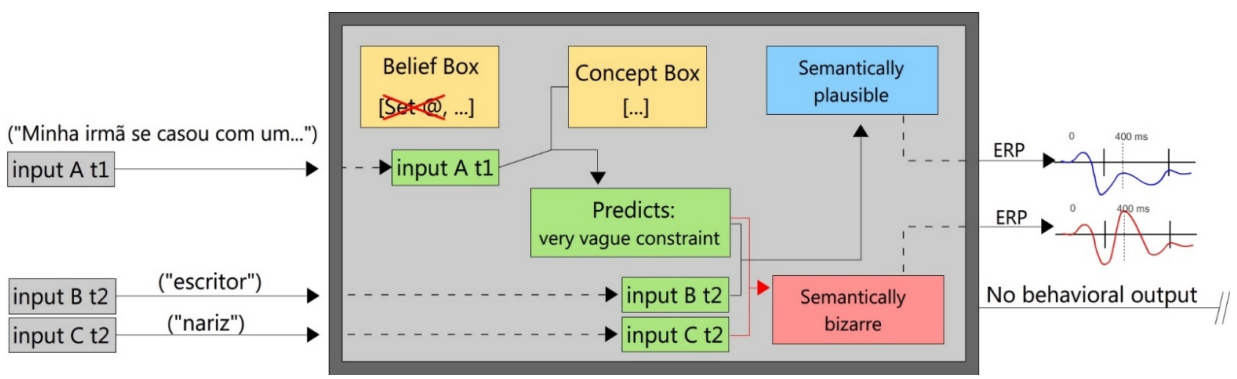

Fig. 12 - Modelo de processo cognitivo que modularia os ERP's de sentenças semanticamente congruentes/incongruentes.

Os quatro modelos esboçados acima possuem em comum o fato de predições satisfeitas elicitarem um N400 mais atenuado que o de predições frustradas: haveria uma conjunção

constante entre frustrações de expectativas (aqui descritas como eventos mentais, subjetivos) e N400 acentuados (aqui descritos como eventos eletrofisiológicos, objetivos)[8], assim como entre expectativas satisfeitas e N400 atenuados. Nas violações de conhecimento, a diferença entre as condições de alta e baixa precisão (figuras 9 e 10, respectivamente) seria o grau de incerteza dos inputs esperados, o que viria a refletir nos ERP's.[9] Já a diferença entre as violações de conhecimento e violações semânticas, não obstante o fato de seus N400 possuírem uma mesma latência, seria relativa ao *tipo de informação* recrutada pelo mecanismo gerador da predição: nas violações de conhecimento, representações de estados de coisas que o indivíduo toma como correspondendo ao mundo; nas violações semânticas, conceitos.

4. Formalização dos argumentos

Pretende-se, nesta seção, explicitar cada etapa do argumento do experimento. Fazendo uso da conjunção conjecturada entre frustrações de expectativas e N400 acentuados e entre expectativas satisfeitas e N400 atenuados, prediz-se, em filósofos, as seguintes consequências:

Premissas Gerais	1	(∀x)(∀ω)(∀φ) x é um filósofo ∧ "ω...φ" é uma sentença do conjunto @ → x acredita que [ω...φ ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)]	P
	2	(∀x)(∀ω)(∀φ)(∀β)(∀t) x acredita que [ω...φ ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)] ∧ "β" é apresentado a x no instante t após x ter lido "ω" ∧ ¬(β=φ) → x tem a sua expectativa semântica frustrada ao ler a palavra "β" que foi apresentada no instante t	P
	3	(∀x)(∀ω)(∀φ)(∀t) x acredita que [ω...φ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)] ∧ "φ" é apresentado a x no instante t após x ter lido "ω" ∧ → x tem a sua expectativa semântica satisfeita ao ler a palavra "φ" que foi apresentada no instante t	P
Hipóteses	4	(∀x)(∀φ)(∀t) x tem a sua expectativa semântica frustrada ao ler a palavra "φ" que foi apresentado no instante t ∧ os sinais eletroencefalográficos de x estão sendo medidos a partir do instante t → os sinais dos eletrodos de x localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após t	H
	5	(∀x)(∀φ)(∀t) x tem a sua expectativa semântica satisfeita ao ler a palavra "φ" que foi apresentado no instante t ∧ os sinais eletroencefalográficos de x estão sendo medidos a partir do instante t → ¬(os sinais dos eletrodos de x localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após t)	H
Fatos particulares	6	Pedro é um filósofo	P
	7	"Platão" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de"	P
	8	Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0	P
	9	Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	P
	10	"Fédon é um diálogo de Platão" é uma sentença do conjunto @	P

[Fig. 13 segue na próxima página]

[8] Não se está assumindo aqui nenhuma forma de dualismo, visto que descrições de níveis distintos podem ser concebidas como tendo um mesmo referente (e.g. "dor" e "disparos de fibras-C"), assim como é possível fazer uso do conhecimento de uma conjunção constante em um determinado sentido sem se comprometer com nenhuma "conexão" entre os conjuntivos (e.g. a generalização indutiva (∀x) (x está com dor) → (as fibras-C de x disparam)).

[9] Resultados que se espera encontrar nos filósofos:
(Amplitude média dos N400 elicitados pelas palavras corretas na condição de alta precisão) < (Amplitude média dos N400 elicitados pelas palavras corretas na condição de baixa precisão);
(Amplitude média dos N400 elicitados pelas palavras incorretas na condição de alta precisão) > (Amplitude média dos N400 elicitados pelas palavras incorretas na condição de baixa precisão);
(Amplitude média dos N400 Effects na condição de alta precisão) > (Amplitude média dos N400 Effects na condição de baixa precisão).

Consequências lógicas das premissas	11	Pedro é um filósofo ∧ "Fédon é um diálogo de Platão" é uma sentença do conjunto @	CONJ. 6,10
	12	Pedro é um filósofo ∧ "Fédon é um diálogo de Platão" é uma sentença do conjunto @ → Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	EU1
	13	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	MP 11,12
	14	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] ∧ "Platão" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" → Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante 0	EU3
	15	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] ∧ "Platão" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de"	CONJ. 7,13
	16	Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante 0	MP 14,15
	17	Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante 0 ∧ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0	CONJ. 8,16
Consequências lógicas das hipóteses 4 e 5	18	Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante 0 ∧ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0 → ¬(os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0)	EU5
	19	¬(Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0)	MP 17,18
Sinal do EEG de Pedro	20	¬(Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0)	P

Fig. 13 - Formalização das consequências lógicas da hipótese 5 em filósofos.

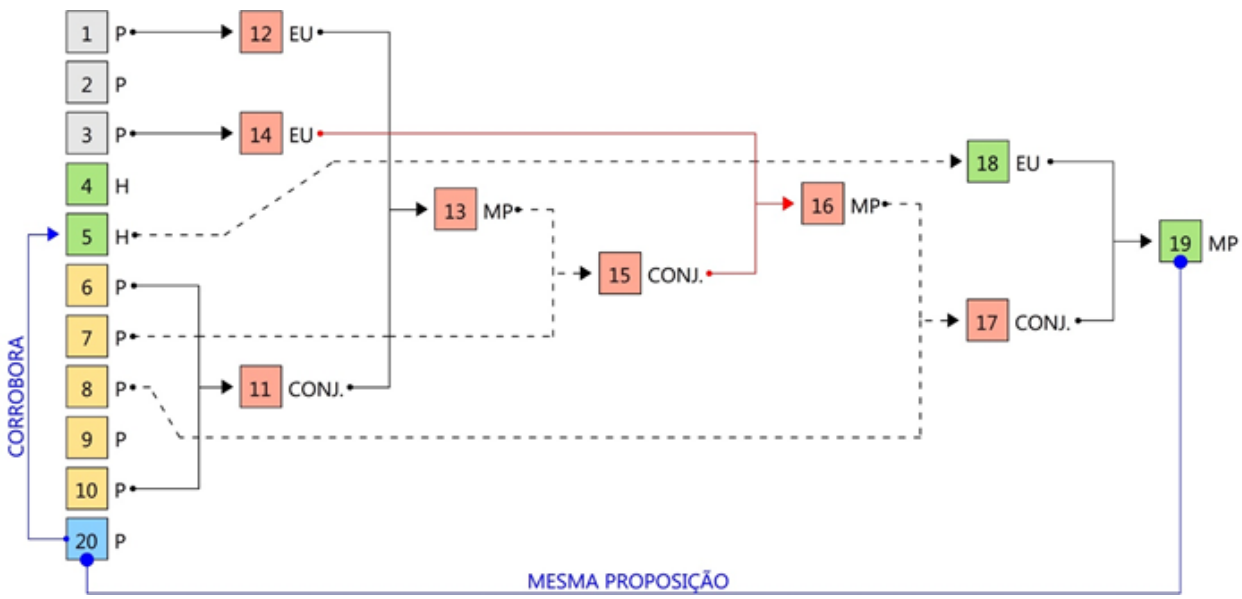

Fig. 14 - Esquema simplificado da estrutura do argumento exposto na Figura 13.

Premissas Gerais	1	(∀x)(∀ω)(∀φ) x é um filósofo ∧ "ω...φ" é uma sentença do conjunto @ → x acredita que [ω...φ ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)]	P
	2	(∀x)(∀ω)(∀φ)(∀β)(∀t) x acredita que [ω...φ ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)] ∧ "β" é apresentado a x no instante t após x ter lido "ω" ∧ ¬(β=φ) → x tem a sua expectativa semântica frustrada ao ler a palavra "β" que foi apresentada no instante t	P
	3	(∀x)(∀ω)(∀φ)(∀t) x acredita que [ω...φ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)] ∧ "φ" é apresentado a x no instante t após x ter lido "ω" ∧ → x tem a sua expectativa semântica satisfeita ao ler a palavra "φ" que foi apresentada no instante t	P

[Fig. 15 segue na próxima página]

Hipóteses	4	(∀x)(∀φ)(∀t) x tem a sua expectativa semântica frustrada ao ler a palavra "φ" que foi apresentado no instante t ∧ os sinais eletroencefalográficos de x estão sendo medidos a partir do instante t → os sinais dos eletrodos de x localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após t	H
	5	(∀x)(∀φ)(∀t) x tem a sua expectativa semântica satisfeita ao ler a palavra "φ" que foi apresentado no instante t ∧ os sinais eletroencefalográficos de x estão sendo medidos a partir do instante t → ¬(os sinais dos eletrodos de x localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após t)	H
Fatos particulares	6	Pedro é um filósofo	P
	7	"Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de"	P
	8	Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0	P
	9	Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	P
	10	Fédon é um diálogo de Platão é uma sentença do conjunto @	P
	11	¬(Kant=Platão)	P
Consequências lógicas das premissas	12	Pedro é um filósofo ∧ Fédon é um diálogo de Platão é uma sentença do conjunto @	CONJ. 6,10
	13	Pedro é um filósofo ∧ "Fédon é um diálogo de Platão" é uma sentença do conjunto @ → Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	EU1
	14	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	MP 12,13
	15	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] ∧ "Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" ∧ ¬(Kant=Platão) → Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0	EU2
	16	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] ∧ "Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" ∧ ¬(Kant=Platão)	CONJ. 7,11,14
	17	Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0	MP 15,16
	18	Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0 ∧ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0	CONJ. 8,17
Consequências lógicas das hipóteses 4 e 5	19	Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0 ∧ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0 → os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0	EU4
	20	Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0	MP 18,19
Sinal do EEG de Pedro	21	Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0	P

Fig. 15 - Formalização das consequências lógicas da hipótese 4 em filósofos.

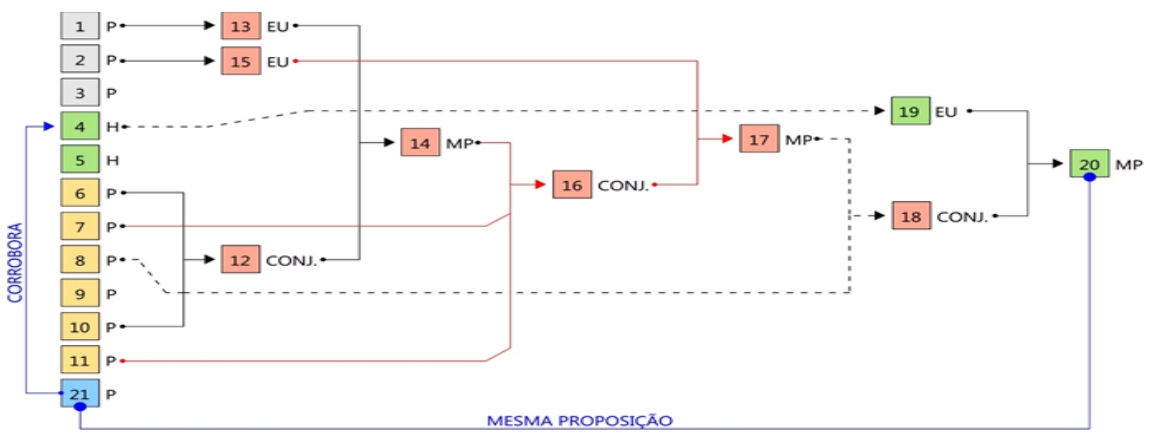

Fig. 16 - Esquema simplificado da estrutura do argumento exposto na Fig. 15.

Os dois argumentos acima demonstram quais as consequências empíricas deduzidas das hipóteses 4 e 5. A estrutura do argumento não permite certeza quanto à veracidade das hipóteses, pois os resultados esperados, ainda que se verifiquem, podem apenas corroborá-las (pois são afirmações do consequente). Entretanto, se os dados eletrofisiológicos não forem os esperados, então poderemos inferir que as hipóteses são falsas (via negação do consequente). Faz-se necessário enfatizar que a resposta evocada por um único *trial* não se faz visível em sua gravação por haver sobreposição de ruído aleatório dissociado do estímulo, e que somente a média de vários *trials* de uma mesma condição permite atenuar os ruídos individuais, permitindo enxergar o ERP. Também não é possível determinar com certeza se o indivíduo que apresentar significativo N400 *Effect* possui um conhecimento específico (e.g. que Teeteto é um diálogo de Platão), visto que é possível que, dos 80 fatos filosóficos mencionados, a pessoa venha a ignorar uma pequena porcentagem.

Uma vez corroboradas as hipóteses 4 e 5, pode-se arriscar assumi-las como premissas, testando novas hipóteses. O argumento da Fig. 17 detalha como seria possível inferir que um indivíduo não é filósofo ao ser demonstrado que da hipótese de que o mesmo indivíduo é filósofo deduz-se uma contradição (por meio de *reductio ad absurdum*):

Premissas Gerais	1	(∀x)(∀ω)(∀φ) x é um filósofo ∧ "ω...φ" é uma sentença do conjunto @ → x acredita que [ω...φ ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)]	P
	2	(∀x)(∀ω)(∀φ)(∀β)(∀t) x acredita que [ω...φ ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)] ∧ "β" é apresentado a x no instante t após x ter lido "ω" ∧ ¬(β=φ) → x tem a sua expectativa semântica frustrada ao ler a palavra "β" que foi apresentada no instante t	P
	3	(∀x)(∀ω)(∀φ)(∀t) x acredita que [ω...φ∧ ¬(∃θ) ω...θ ∧ ¬(θ=φ)] ∧ "φ" é apresentado a x no instante t após x ter lido "ω" ∧ → x tem a sua expectativa semântica satisfeita ao ler a palavra "φ" que foi apresentada no instante t	P
	4	(∀x)(∀φ)(∀t) x tem a sua expectativa semântica frustrada ao ler a palavra "φ" que foi apresentado no instante t ∧ os sinais eletroencefalográficos de x estão sendo medidos a partir do instante t → os sinais dos eletrodos de x localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após t	P
	5	(∀x)(∀φ)(∀t) x tem a sua expectativa semântica satisfeita ao ler a palavra "φ" que foi apresentado no instante t ∧ os sinais eletroencefalográficos de x estão sendo medidos a partir do instante t → ¬(os sinais dos eletrodos de x localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após t)	P
	6	(∀π) ¬(x acredita que π) ∧ π → x ignora que π	P
Hipótese	7	Pedro é um filósofo	H
Fatos particulares	8	"Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de"	P
	9	Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0	P
	10	Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	P
	11	"Fédon é um diálogo de Platão" é uma sentença do conjunto @	P
	12	¬(Kant=Platão)	P
Consequências lógicas das premissas	13	Pedro é um filósofo ∧ "Fédon é um diálogo de Platão" é uma sentença do conjunto @ → Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	EU1
	14	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] ∧ "Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" ∧ ¬(Kant=Platão) → Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0	EU2
	15	Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0 ∧ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0 → os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0	EU4

[Fig. 17 segue na próxima página]

Consequências lógicas das hipóteses 7	16	Pedro é um filósofo ∧ "Fédon é um diálogo de Platão" é uma sentença do conjunto @	CONJ. 7,11
	17	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	MP 13,16
	18	Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] ∧ "Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" ∧ ¬(Kant=Platão)	CONJ. 17,8,12
	19	Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0	MP 14,18
	20	Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0 ∧ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0	CONJ. 19,9
	21	Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0	MP 15,20
Sinal do EEG de Pedro	22	¬(Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0)	P
Contradição deduzida da hipótese 7	23	(Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0) ∧ ¬(Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0)	C 21,22
Conclusões	24	¬(Pedro é um filósofo)	RAA 7-23
	25	¬(Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0 ∧ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0)	MT 15,22
	26	¬(Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0)	9,25
	27	¬{Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] ∧ "Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" ∧ ¬(Kant=Platão)}	MT 14,26
	28	¬{Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]}	8,12,27
	29	¬{Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]} ∧ Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)] → Pedro ignora que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	EU6
	30	¬{Pedro acredita que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]} ∧ Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	CONJ. 10,28
	31	Pedro ignora que [Fédon é um diálogo de Platão ∧ ¬(∃θ) Fédon é um diálogo de θ ∧ ¬(θ=Platão)]	MP 29,30

Fig. 17 - Demonstração de como os ERP's individuais permitiriam a identificação de não-membros de uma determinada categoria C a partir da detecção de ignorância de fatos cujo conhecimento é condição necessária para ser membro de C.

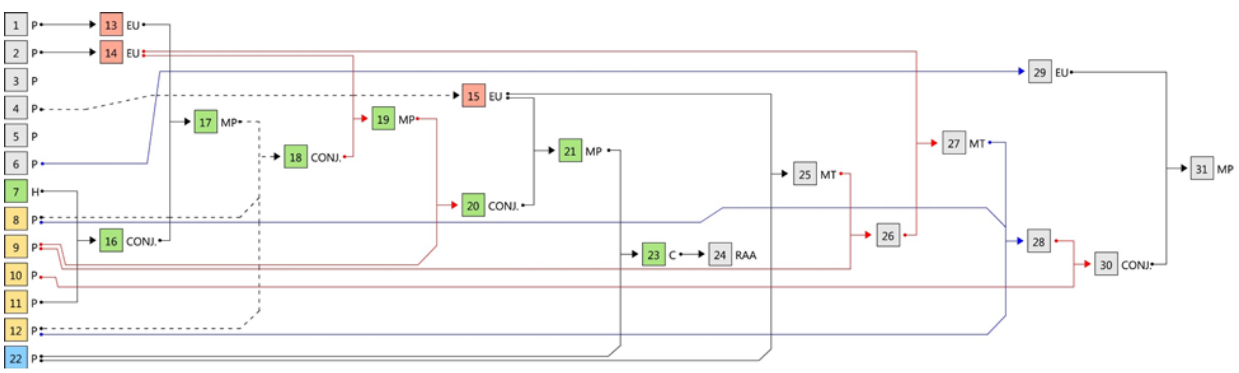

Fig. 18 - Esquema simplificado da estrutura do argumento exposto na Fig. 17.

5. Considerações finais

Para explicitar a estrutura da presente argumentação sem complexificações desnecessárias, algumas idealizações tiveram que ser feitas, como a pressuposição de que todos os filósofos conheçam todos os fatos referidos pelas sentenças do conjunto @ e que todos os não-filósofos os ignorem inteiramente (algo que é obviamente falso). Considerando ser suficiente, na prática, a existência de regularidades que se verificam "na maioria dos casos" (em vez de leis estritas), espera-se encontrar resultados que, não obstante a presença de *outliers*, sejam compatíveis com os dados previstos. O experimento referido é uma tentativa de alcançar evidências empíricas que possam corroborar, aprimorar ou refutar teorias filosóficas sobre representações mentais, processamento semântico e léxico, a partir do modo como representações de estados de coisas e eventos vão sendo mentalmente modelados de modo gradual e preditivo durante uma tarefa de leitura. Se for realmente possível, utilizando a técnica aqui exposta, extrair de um indivíduo um conjunto de crenças por ele possuídas sem que ele voluntariamente nos forneça tal informação (apenas ao engajá-lo em uma tarefa de leitura), então seria a princípio também possível detectar conhecimento ou ignorância de [um conjunto de] fatos em uma pessoa que os quer intencionalmente omitir. Uma possível aplicação prática da técnica é a sua utilização como detector de mentiras: basta que se substitua, nos argumentos acima, "Filósofos" e "Não-Filósofos" por "Criminosos" e "Inocentes", e "conhecimento filosófico" (=@) por "conhecimento de fatos de um crime".[10] No entanto, é difícil que um criminoso real se lembre perfeitamente de um número significativamente grande (e.g. 80) de fatos relacionados a seu crime. No caso dos filósofos, supõe-se que as crenças filosóficas selecionadas sejam nesses muito enraizadas, e também que a escolha por sentenças que impõem grandes restrições quanto ao próximo input contribuam para um maior *N400 Effect.* Neste caso, a inferência reversa se afigura uma prática muito mais confiável – algo que poderá ser sustentado caso a porcentagem de categorizações corretas for maior que a dos experimentos que utilizaram sentenças de menor grau preditivo.

Referências

BOAZ, T. L.; PERRY, N. W.; RANEY, G; FISCHLER, I. S.; SHUMAN, D. Detection of guilty knowledge with event-related potentials. *Journal of Applied Psychology*, v. 76, n. 6, p. 788-795, 1991.

CHURCHLAND, P. M. Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, v. 78, n. 1, p. 67-90, 1981.

CHURCHLAND, P. M. Folk psychology and the explanation of human behavior. *Proceedings of the Aristotelian Society*, v. 62, p. 209-221, 1988.

FEDERMEIER, K. D.; KUTAS, M. Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, v. 8, n. 3, p. 373-392, 1999.

FODOR, Jerry A. Propositional attitudes. In: CHALMERS, D. (Ed.) *Philosophy of mind*: classical and contemporary readings. New York: Oxford University Press, 2002. p. 542-55.

HAGOORT, P.; HALD, L. A.; BASTIAANSEN, M. C. M.; PETERSSON, K. M. Integration of word meaning and world knowledge in language comprehension. *Science*, v. 304, n. 5669, p. 438-441, 2004.

[10] Algo semelhante já foi realizado por Boaz et al. (1991), em que 70% dos participantes foram categorizados corretamente como "inocentes" ou "culpados" (os participantes não cometeram nenhum crime, mas assistiram a vídeos filmados em primeira pessoa que mostravam ou a ação de um criminoso ou o mero passeio de uma pessoa inocente).

KUTAS, M.; FEDERMEIER, K. D. Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, v. 62, p. 621-647, 2011.

KUTAS, M.; HILLYARD, S. A. Reading senseless sentences: brains potentials reflect semantic incongruity. *Science*, v. 207, p. 203-205, 1980.

LAU, E. F.; HOLCOMB, P. J.; KUPERBERG, G. R. Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, v. 25, n. 3, p. 484-502, 2013.

LEWIS, D. An argument for the identity theory. *Journal of Philosophy*, v. 63, n. 2, p. 17-25, 1970. (1970a)

LEWIS, D. How to define theoretical terms. *Journal of Philosophy*, v. 67, n. 13, p. 427-446, 1970. (1970b)

LEWIS, D. Mad pain and Martian pain. In: BLOCK, N. (Ed.) *Readings in the Philosophy of Psychology* – v. 1. Cambridge: Harvard University Press, 1980. p. 216-222.

LEWIS, D. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, v. 50, p. 249-258, 1972.

LUCK, S. J. *An introduction to the Event-Related Potential technique*. Cambridge: MIT Press, 2005.

PICKERING, M. J.; CLARK, A. Getting ahead: forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, v. 18, n. 9, p. 451-456, 2014.

RYLE, G. *The concept of mind*. London: Penguin Books, 2000.