

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
NÍVEL MESTRADO**

RICARDO AUGUSTO PERERA

**ACESSANDO REPRESENTAÇÕES MENTAIS PARA PREDIZER ESTÍMULOS:
COMO CRENÇAS MODULAM SINAIS CEREBRAIS**

São Leopoldo

2017

RICARDO AUGUSTO PERERA

ACESSANDO REPRESENTAÇÕES MENTAIS PARA PREDIZER ESTÍMULOS:
COMO CRENÇAS MODULAM SINAIS CEREBRAIS

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre, pelo Programa de Pós-Graduação em Filosofia da Universidade do Vale do Rio dos Sinos - UNISINOS

Orientadora: Prof. Dra. Sofia Inês Albornoz Stein

São Leopoldo

2017

P437a Perera, Ricardo Augusto.
Acessando representações mentais para prever
estímulos : como crenças modulam sinais cerebrais /
Ricardo Augusto Perera. – 2017.
104 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio
dos Sinos, Programa de Pós-Graduação em Filosofia, 2017.
“Orientadora: Prof. Dra. Sofia Inês Albornoz Stein.”

1. Representações mentais. 2. Crenças. 3. Mente
preditiva. 4. Eletroencefalografia. I. Título.

CDU 1

RICARDO AUGUSTO PERERA

ACESSANDO REPRESENTAÇÕES MENTAIS PARA PREDIZER ESTÍMULOS:
COMO CRENÇAS MODULAM SINAIS CEREBRAIS

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre, pelo Programa de Pós-Graduação em Filosofia da Universidade do Vale do Rio dos Sinos - UNISINOS

Aprovado em 03 de março de 2017

BANCA EXAMINADORA

Dra. Sofia Inês Albornoz Stein (orientadora) – UNISINOS

Dr. Adriano Naves de Brito – UNISINOS

Dr. Gabriel José Corrêa Mograbi – UFMT

AGRADECIMENTOS

A presente dissertação foi possibilitada pelo esforço conjunto dos membros do grupo de Pesquisa Social Brains e Chiron, liderados pela Prof. Dra. Sofia Inês Albornoz Stein e pelo Prof. Dr. Adriano Naves de Brito, e apresenta o primeiro resultado das pesquisas empíricas realizadas no Laboratório de Filosofia Experimental e Estudos da Cognição. Os temas estudados nas reuniões do grupo Social Brains, assim os assuntos tratados nos eventos organizados e artigos produzidos, foram a base para a escolha do meu objeto de pesquisa.

Agradeço aos colegas Ana Paula Emmendorfer, João Machado Silveira, Samuel Telles e Viviane Braga pela colaboração durante a realização dos experimentos; ao Dr. Patrick Haggard (UCL), ao Dr. Günther Gerhardt (UCS) e ao Dr. Adolfo García (INECO), que me auxiliaram com a interpretação dos dados e sugeriram melhorias.

Agradeço aos professores avaliadores, Dr. Adriano Naves de Brito (UNISINOS), Dr. John Bolender (UNISINOS) e Dr. Gabriel José Corrêa Mograbi (UFMT), pelos valiosos comentários, críticas e sugestões.

Agradeço aos 35 voluntários que participaram do experimento, cujos dados permitiram a obtenção dos resultados aqui expostos.

Agradeço à CAPES pela bolsa concedida.

Em especial agradeço à minha orientadora, Prof. Dra. Sofia Inês Albornoz Stein, por todo o aprendizado, colaboração e incentivo nesses 5 anos de orientação.

RESUMO

Utilizando o framework teórico denominado Predictive Processing (Clark, 2013), que assume que cérebros são essencialmente máquinas que predizem seus próximos inputs e o fazem minimizando o erro de predição (diferença entre o input previsto e o recebido) de modo semelhante a uma inferência bayesiana, procurou-se, nesta dissertação, encontrar padrões de atividade eletrofisiológica que sinalizassem o recrutamento de crenças epistêmicas. Conjeturou-se que enquanto frases são lidas, representações de estados de coisas e eventos são mentalmente modeladas de modo gradual e preditivo, e que os estímulos vindouros são antecipados com base naquilo que se afigura “o mais provável” em termos epistêmicos, semânticos, sintáticos, léxicos e perceptuais. Devido ao fato de já serem conhecidas as respostas eletrofisiológicas que são moduladas por violações de expectativa dos níveis semântico, sintático, léxico e perceptual, assim como o preciso intervalo temporal em que se manifestam, foi possível estudar como a expectativa gerada por crenças modula os sinais cerebrais. Os experimentos realizados no Laboratório de Filosofia Experimental e Estudos da Cognição, localizado na UNISINOS, apresentavam frases em um monitor (palavra por palavra) que aludiam a fatos conhecidos apenas por um grupo de participantes (filósofos), estruturadas de modo que apenas uma única palavra, aparecendo no final de cada sentença, fosse capaz de as tornar verdadeiras (e.g. “Teeteto é um diálogo escrito por Platão”). O modelo proposto assume que, antes de aparecer a palavra final, o significado dos termos é acessado (no caso exemplificado, um texto específico e uma relação), um estado de coisas representado (i.e. existir um x tal que x escreveu Teeteto), e uma busca iniciada pelo melhor candidato a x, tendo as crenças do sujeito como alvo do rastreamento. A crença relevante é então recrutada e a representação do x identificado (Platão) acessada. Informações sobre a melhor predição semântica (o filósofo Platão) são utilizadas para selecionar o mais provável item léxico que tenha Platão como significado. Os participantes aos quais eram atribuídas crenças filosóficas apresentaram *Event-Related Potentials* correlacionados a processamento léxico-semântico significativamente diferentes dos participantes do grupo dos não-filósofos. Considerando que a única diferença relevante entre os dois grupos era a conjecturada posse ou ausência de determinadas crenças, os resultados foram interpretados como sinalizando o recrutamento de crenças em processos preditivos subjacentes à compreensão textual. Os resultados contrariam posições eliminativistas que consideram que o vocabulário mentalista acerca de crenças, intenções e desejos carece de significado, uma vez que a diferença encontrada sugere que de fato há algo em cérebros que é denotado por meio dessas expressões (ainda que de modo vago e grosseiro) e que está a modular os sinais. Ainda que não compreendamos prontamente o exato modo como se efetiva a instanciação de crenças por cérebros, podemos estudar o modo como elas são recrutadas e integram diversos processos cognitivos (i.e. seus papéis causais).

Palavras-chave: Representações Mentais. Crenças. Mente Preditiva. Eletroencefalografia

ABSTRACT

Using the theoretical framework called Predictive Processing (Clark, 2013), which assumes that brains are essentially machines that predict their next inputs and do so by minimizing prediction error (difference between predicted and received input) similar to a Bayesian inference, it was sought, in this dissertation, to find patterns of electrophysiological activity that signal the recruitment of epistemic beliefs. It has been conjectured that as sentences are read, representations of states of affairs and events are mentally modeled in a gradual and predictive fashion, and that the upcoming stimuli are anticipated on the basis of what appears to be "most likely" in epistemic, semantic, syntactic, lexical and perceptual terms. Due to the fact that the electrophysiological responses that are modulated by violations of expectation of the semantic, syntactic, lexical and perceptual levels are known, as well as the precise time interval in which they are manifested, it was possible to study how the expectation generated by beliefs modulates the cerebral signals. The experiments carried out in the Laboratory of Experimental Philosophy and Studies of Cognition, located at UNISINOS, presented sentences on a monitor (word by word) that alluded to facts known only by a group of participants (philosophers), structured in a way that only a single word, appearing at the end of each sentence, would be able to make them true (e.g. "Theaetetus is a dialogue written by Plato"). The proposed model assumes that, before the final word appears, the meaning of the terms is accessed (in the exemplified case, a specific text and a relation), a state of affairs represented (i.e. there is an x such that x wrote Theaetetus), and a search initiated to find the best candidate for x , taking the subject's beliefs as the target of tracking. The relevant belief is then recruited and the representation of the identified x (Plato) accessed. Information on the best semantic prediction (the philosopher Plato) is used to select the most likely lexical item that has Plato as meaning. Participants for which philosophical beliefs were ascribed presented Event-Related Potentials correlated to lexical-semantic processing significantly different from participants in the non-philosophers group. Considering that the only relevant difference between the two groups was the conjectured possession or absence of certain beliefs, the results were interpreted as signaling the recruitment of beliefs in predictive processes underlying textual comprehension. The results contradict eliminativist positions that consider that the mentalist vocabulary about beliefs, intentions, and desires is meaningless, since the difference found suggests that there is indeed something in brains that is denoted by these expressions (albeit vaguely and coarse) and that is modulating the signals. Although we do not readily understand exactly how belief instantiation by brains is effective, we can study how they are recruited and integrate various cognitive processes (i.e. their causal roles).

Keywords: Mental Representations. Beliefs. Predictive Mind.
Electroencephalography

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema argumentativo da dissertação.	17
Figura 2 – Representação bidimensional de superfícies côncavas e convexas.....	22
Figura 3 – Esquema representando o fluxo de informação <i>bottom-up</i> e <i>top-down</i> entre diversos níveis de processamento linguístico.	34
Figura 4 – Hipóteses possíveis acerca da causa dos inputs recebidos.	36
Figura 5 – Sequência de inputs do False Belief Task verbal.....	42
Figura 6 – Resumo dos resultados dos principais experimentos que utilizaram versões do FBT. Os resultados marcados com asterisco não correspondem a nenhum experimento realizado, consistindo apenas em estimativas inferidas dos demais experimentos.	45
Figura 7 - Exemplo de sentença do tipo $A...B \wedge \neg(\exists\theta) A... \theta \wedge \neg(\theta=B)$	55
Figura 8 – Exemplo de sentença do tipo $A...B \wedge (\exists\theta) A... \theta \wedge \neg(\theta=B)$	56
Figura 9 - Conjunto @ ($P \cap Q$): o conjunto das sentenças do tipo $A...B \wedge \neg(\exists\theta) A... \theta \wedge \neg(\theta=B)$ que todo filósofo sabe serem verdadeiras	57
Figura 10 – Correlação esperada entre medição de conhecimento filosófico em questionário escrito e N400 Effects.....	58
Figura 11 – Representação do sinal esperado pelos participantes que tiverem conhecimento das sentenças do conjunto @, ao lerem a palavra incorreta: um N400 forte.....	59
Figura 12 - Representação do sinal esperado pelos participantes que tiverem conhecimento das sentenças do conjunto @, ao lerem a palavra correta: um N400 fraco.	60
Figura 13 – Comparação das médias esperadas dos sinais dos filósofos e não-filósofos, nas sentenças verdadeiras e falsas.	61
Figura 14 – 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).	61
Figura 15 – Modelo de processo cognitivo que modularia os ERP's dos filósofos em uma condição preditiva de alta precisão.	64
Figura 16 – Modelo de processo cognitivo que modularia os ERP's dos filósofos em uma condição preditiva de baixa precisão.	65

Figura 17 – Modelo de processo cognitivo que modularia os ERP's dos não-filósofos em uma condição preditiva de alta precisão.	66
Figura 18 – Modelo de processo cognitivo que modularia os ERP's de sentenças semanticamente congruentes/incongruentes.	66
Figura 19 – Formalização das consequências lógicas da hipótese 5 em filósofos....	68
Figura 20 – Esquema simplificado da estrutura do argumento exposto na Figura 19.	69
Figura 21 - Formalização das consequências lógicas da hipótese 4 em filósofos. ...	70
Figura 22 - Esquema simplificado da estrutura do argumento exposto na Figura 21.	71
Figura 23 – 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.....	73
Figura 24 - Esquema simplificado da estrutura do argumento exposto na figura 23.	74
Figura 25 – Variação da amplitude do <i>N400 effect</i> em função do grau de conhecimento da cada grupo.	78
Figura 26 - Mapa topográfico do <i>N400 Effect</i> do experimento envolvendo anomalias semânticas, replicando metodologia de KUTAS et al. (1980).	79
Figura 27 – Eletrodos selecionados para análise estatística do N400.	80
Figura 28 – Diferença nos ERP's do grupo dos filósofos e dos não-filósofos em 13 eletrodos posicionados na região centro-parietal, nas frases verdadeiras e falsas...	81
Figura 29 – Mapa topográfico dos <i>N400 Effects</i> dos filósofos e não-filósofos aos 400 ms.	82
Figura 30 – Tabelas ANOVA, médias e erros padrão dos N400 Effects dos filósofos e não filósofos.	83
Figura 31 – Tabelas ANOVA, médias e erros padrão dos <i>N400 Effects</i> dos filósofos com maior tempo de estudo, não-filósofos e calouros.	85
Figura 32 – Valores das médias individuais dos filósofos, não-filósofos e calouros..	85
Figura 33 - Eletrodos selecionados para análise estatística do Recognition Potential.	87
Figura 34 – Tabelas ANOVA, médias e erros padrão dos <i>RP Effects</i> dos filósofos e não-filósofos.	88

Figura 35 – Correlação de Pearson entre os <i>RP Effects</i> e os <i>N400 Effects</i> de todos os grupos.....	90
Figura 36 – Diferenças entre contextos preditivos e não-preditivos.....	91

SUMÁRIO

1 INTRODUÇÃO.....	10
2 FRAMEWORKS TEÓRICOS.....	18
2.1 FRAMEWORK 1 - MODULARIDADE DA MENTE	19
2.2 FRAMEWORK 2 - PREDICTIVE BRAIN	26
2.2.1 Implementação neural e evidências provenientes de respostas corticais	29
2.2.2 Tipo de computação implementada e exemplos corroboradores	35
2.3 ESCOLHA DE FRAMEWORK.....	39
3 SOBRE A CAPACIDADE DE ATRIBUIR ESTADOS MENTAIS REPRESENTACIONAIS.....	41
3.1 FALSE BELIEF TASK VERBAL	41
3.2 O FALSE BELIEF TASK NÃO-VERBAL.....	43
4 SOBRE A ONTOLOGIA DE ESTADOS MENTAIS REPRESENTACIONAIS.....	47
5 ACESSANDO INFORMAÇÕES SOBRE CRENÇAS POR MEIO DE RESPOSTAS ELETROFISIOLÓGICAS	53
5.1 PROJETO DO EXPERIMENTO	53
5.2 FORMALIZAÇÃO DOS ARGUMENTOS.....	67
5.3 RESULTADOS	76
5.3.1 N400.....	76
5.3.2 Recognition Potential	86
5.3.3 Correlação entre o N400 e o <i>Recognition Potential</i>	89
5.4 DIREÇÕES FUTURAS: RECRUTAMENTO DE CRENÇAS EM CONTEXTOS NÃO-PREDITIVOS.....	90
CONSIDERAÇÕES FINAIS	92
REFERÊNCIAS.....	96

1 INTRODUÇÃO

Entre as questões clássicas da Filosofia da Mente, há o problema da relação mente-corpo, que interroga como fenômenos mentais estão relacionados a fenômenos físicos, e o problema das outras mentes, concernente à possibilidade de conhecermos os estados mentais de outras pessoas. Um tema que permeia esta discussão é aquilo que os filósofos chamam de intencionalidade, isto é, a suposta capacidade da mente de representar e “ser acerca de” objetos e estados de coisas presentes ou não no mundo e que, para alguns, consistiria “a marca distintiva do mental” (e.g. CRANE, 2001; BRENTANO, 2014; HUSSERL, 1970). A estrutura geral da intencionalidade envolve uma relação psicológica específica (atitude, modo) que um sujeito possui com certos conteúdos intencionais – usualmente formalizada naquilo que o filósofo Bertrand Russell nomeou de “atitudes proposicionais”. Os dois termos mentais de uso mais frequente da linguagem intencional – crenças e desejos – referir-se-iam, respectivamente, àquelas representações de estados de coisas que um sujeito toma como sendo o caso e às representações de estados de coisas possíveis – não factuais – que ele quer que se concretizem. Entretanto, para alguns filósofos da mente contemporâneos, como Paul e Patrícia Churchland, que procuram abordar a natureza mental de uma perspectiva de terceira pessoa própria da atividade científica, termos como crenças, desejos e intenções constituiriam “entidades fantasmagóricas”, resíduos dualistas a serem eliminados pelas explicações das neurociências com o avanço destas. O materialismo eliminativista (*eliminative materialism*) que defendem considera que as teorias acerca da mente adotadas pelo senso comum consistem em explicações radicalmente inadequadas de nossas atividades cerebrais, “defeituosas demais para sobreviver à redução interteórica” (CHURCHLAND, 1981, p. 72), devendo, portanto, ser simplesmente eliminadas (e não suavemente reduzidas) por uma neurociência completa, do mesmo modo como o desenvolvimento da ciência eliminou entidades que nada designavam como “o flogisto” e “os espíritos vitais”. Posição oposta é a do filósofo Jerry Fodor, para quem não haveria nada de fundamentalmente errado com a visão do senso comum. Ao contrário, a concepção central da teoria atacada por Paul e Patrícia Churchland – a atividade cognitiva como consistindo na manipulação de atitudes proposicionais – é tomada por Fodor como fundamental a sua teoria do pensamento. Sua “linguagem do pensamento” e a teoria computacional da mente a ela associada assumem que

crenças e desejos são informações, encarnadas como configurações de símbolos. Steven Pinker (1998) descreve que “os símbolos são os estados físicos de bits de matéria, como os chips de um computador ou os neurônios do cérebro [...] e simbolizam coisas do mundo porque são desencadeados por essas coisas via órgãos dos sentidos e devido ao que fazem depois de ser desencadeados” (PINKER, 1998, p. 35). Esta estratégia, se correta, resolveria em parte o problema mente-corpo, uma vez que conectaria as à-primeira-vista-entidades-etéreas como "o desejo de ir à casa da avó" a um pedaço físico de matéria como o cérebro, sem comprometer os que adotam uma visão realista de termos mentais com nenhuma forma de dualismo (mesmo a de propriedades).

Teorias intermediárias entre os extremados realismos e o eliminativismo acerca do vocabulário mental, como o behaviorismo lógico (ou disposicionalismo) e certos tipos de funcionalismo, pretendem analisar o significado exato dos termos, à maneira científica, voltada ao observável e objetivo. Behavioristas como Gilbert Ryle (2000) argumentaram que se o uso de termos e predicados mentais ordinários como crenças, competências e traços característicos pessoais são supostos como possuidores de significados, não poderiam ser estes eventos inobserváveis, muito menos propriedades possuídas por uma substância imaterial cartesiana (mito que Ryle pretende eliminar). Alternativamente a isto, o significado das sentenças que invocam tais entidades, após a devida análise, traduzir-se-ia, sem nenhum resquício, a dados observáveis (ou potencialmente observáveis) de comportamentos ou disposições para se comportar sob várias circunstâncias de *inputs*. Acreditar em uma proposição “P” nada mais é que possuir uma ou várias disposições comportamentais tocantes a “P”. Achar que termos mentais designam outra coisa senão comportamentos e disposições seria um erro categorial de supor a mente uma coisa/substância à parte, resultado de confusões e do uso abusivo da linguagem ordinária.

Já David Lewis (1970, 1972) – um funcionalista – considera a prática de atribuição de estados mentais uma teoria que pretende explicar a regularidade entre estímulos e respostas encontradas no comportamento humano. Termos mentais – definidos pela função que realizam – seriam termos teóricos (*T-terms*) introduzidos pela teoria em questão e que, não obstante o fato de serem inobserváveis, adquirem significado por meio de relações que possuem com outros estados mentais, estímulos sensoriais e comportamentos (estes dois últimos termos observáveis, ou *O-terms*).

Entidades teóricas são definidas como “os ocupantes dos papéis causais especificados pela teoria, como as entidades, quaisquer que sejam, que possuem certas relações causais umas com as outras e com os referentes dos *O-terms*” (LEWIS, 1972, p. 211). Lewis pretende com isso demonstrar que, se a teoria psicológica for verdadeira, o que for que existir que tenha relações causais com os *O-terms* de fato existe, mesmo que ainda não sejamos capazes de identificar com exatidão as entidades que “realizam a função” em nossos cérebros (pois só possuiríamos descrições definidas / definições funcionais delas). De acordo com Knobe e Prinz (2008), assim como um cientista postula entidades inobserváveis com o objetivo de prever e explicar o eventos observáveis, também o “*mindreader*” postularia estados mentais inobserváveis como modo de prever e explicar o comportamento humano.

As teorias filosóficas supracitadas concentram-se, principalmente, na ontologia dos termos mentais, indagando se eles possuem ou não referência e, caso haja, de que natureza ela é (disposições comportamentais, realização de funções, informações presentes no cérebro, estados possuidores de intencionalidade, ficções úteis, etc.). Porém, independentemente de sermos capazes ou não de nos referirmos com exatidão àquilo que se passa em mentes alheias, utilizamos no dia a dia, de modo satisfatório, o vocabulário intencional e, embora a teoria subjacente a esta capacidade cognitiva possa vir a revelar-se falsa, parece improvável que nós a abandonemos, ainda mais se a considerarmos parte integrante da cognição humana (da mesma forma que não abandonamos a nossa “*folk physics*” com o advento de teorias científicas, uma vez que a primeira nos permite antecipar eventos de modo satisfatório e rápido, enquanto a segunda demandaria raciocínios complexos – inviáveis quando as circunstâncias exigem rápida resposta e dispensam esforços supérfluos). “Teoria da Mente” (*Theory of Mind*) é o termo que se convencionou chamar a nossa capacidade de atribuir estados mentais possuidores de intencionalidade – como crenças e desejos – a outros indivíduos, conjuntamente com o uso de uma teoria “*folk*” que estipularia o modo como estas entidades teóricas tipicamente relacionam-se entre si, com estímulos sensoriais e comportamentos. É necessário destacar que, diferentemente das teorias científicas, que buscam as causas *reais* de um fenômeno, os estudos em Teoria da Mente concentram-se no modo como as pessoas *pensam* que o comportamento é causado. Pode ser verdadeiro que tenhamos uma teoria do senso comum T para explicar e prever o comportamento de outras pessoas, ainda

que T seja uma teoria falsa (como alerta Paul Churchland). Experimentos recentes encontraram evidências de que crianças de poucos meses já seriam capazes de “ler mentes alheias” muito antes de possuírem os recursos cognitivos necessários para adquirir explicitamente meios de inferir estados mentais e das experiências necessárias para que tal aprendizagem se efetive (ver ONISHI & BAILLARGEON, 2005; SOUTHGATE & VERNETTI, 2014; SOUTHGATE et al, 2007). Não obstante a complexidade das teorias filosóficas, essas crianças já possuiriam uma apreciação tácita de conceitos abstratos como crenças e desejos, sendo ao mesmo tempo capazes de computar –ainda que de modo não-consciente – seus abstrusos papéis causais. A familiaridade com “outras mentes”, desta forma, se daria muito antes de sermos capazes de refletir explicitamente sobre mentes alheias.

A presente dissertação, cujo tema se insere no contexto acima exposto, procura acrescentar novos dados empíricos à discussão que possam atualizar as teorias acerca de representações mentais e servir de evidência para decidir entre teorias rivais. Adotando inicialmente uma postura neutra quanto à ontologia de crenças e outros estados mentais representacionais, buscar-se-á, por meio experimentos com eletroencefalografia, padrões de atividade cerebral que estejam correlacionados com crenças ou que de algum modo são por estas modulados. Pretende-se encurtar a lacuna¹ existente entre estados mentais concebidos em primeira pessoa (privados e subjetivos) e estados cerebrais acessados por meio de sinais eletrofisiológicos provenientes de potenciais pós-sinápticos de células piramidais (estes públicos e objetivos). À primeira vista salta aos olhos um problema inerente a esta abordagem: a circularidade ao utilizar como critério de posse ou ausência de determinadas crenças as atribuições tradicionais da *folk psychology* (capacidade adquirida socialmente por meio de uma linguagem pública) para em seguida verificar se há diferenças cerebrais dependendo da atribuição (i.e. pressupõe-se que cérebros aos quais imputamos estados de ontologia duvidosa de fato os possuam, sendo que é exatamente isso que supostamente se procura provar). Esta dificuldade, entretanto, dissolve-se se considerarmos as crenças assim concebidas como meras hipóteses. Se a conjecturada ausência ou posse de determinadas crenças realmente modular as

¹ O termo "lacuna" aqui se refere à "lacuna explicativa", termo introduzido por Joseph Levine (1983) acerca da dificuldade de teorias fisiológicas explicarem fenômenos psicológicos. Levine foca no problema da consciência, embora seu argumento se refira também a outras funções mentais em que a experiência qualitativa não seja a característica mais saliente.

respostas eletrofisiológicas a um nível estatisticamente significativo, então devemos suspeitar que de fato há algo nos cérebros que é denotado por meio dessas expressões (mesmo que de modo grosseiro) e que está a causar a diferença. Ainda que não compreendamos prontamente o exato modo como se efetiva a instanciação de crenças por cérebros, podemos estudar a forma como elas são recrutadas e integram diversos processos cognitivos (i.e. seus papéis causais). Estando firmado um método confiável que indique o recrutamento de crenças específicas e conceitos por cérebros humanos, faz-se possível avançar no estudos de temas caros à filosofia da mente de um modo interdito à abordagem tradicional:

(1) Quanto tempo leva para o cérebro de sujeitos neurotípicos confrontar o significado de uma sentença lida com um corpo de crenças previamente possuído a fim de determinar seu valor de verdade?

(2) É visível no cérebro a compatibilidade/incompatibilidade daquilo que se lê com aquilo em que se crê?

(3) É possível inferir de um sinal cerebral a posse ou ausência de uma crença ou conjunto de crenças?

(4) Quais os níveis de processamento que possuem penetrabilidade cognitiva?

(5) São crenças individuais recrutadas isoladamente ou em conjunto com outras crenças a elas enredadas?

(6) Há diferenças cerebrais no recrutamento de crenças explícitas (que são representadas) e implícitas (que não são representadas, porém dedutíveis das explícitas), indicando um passo inferencial extra nas últimas?

(7) Conceitos são símbolos atômicos amodais independentes uns dos outros ou constituídos por redes de representações multimodais compartilhadas com outros conceitos?

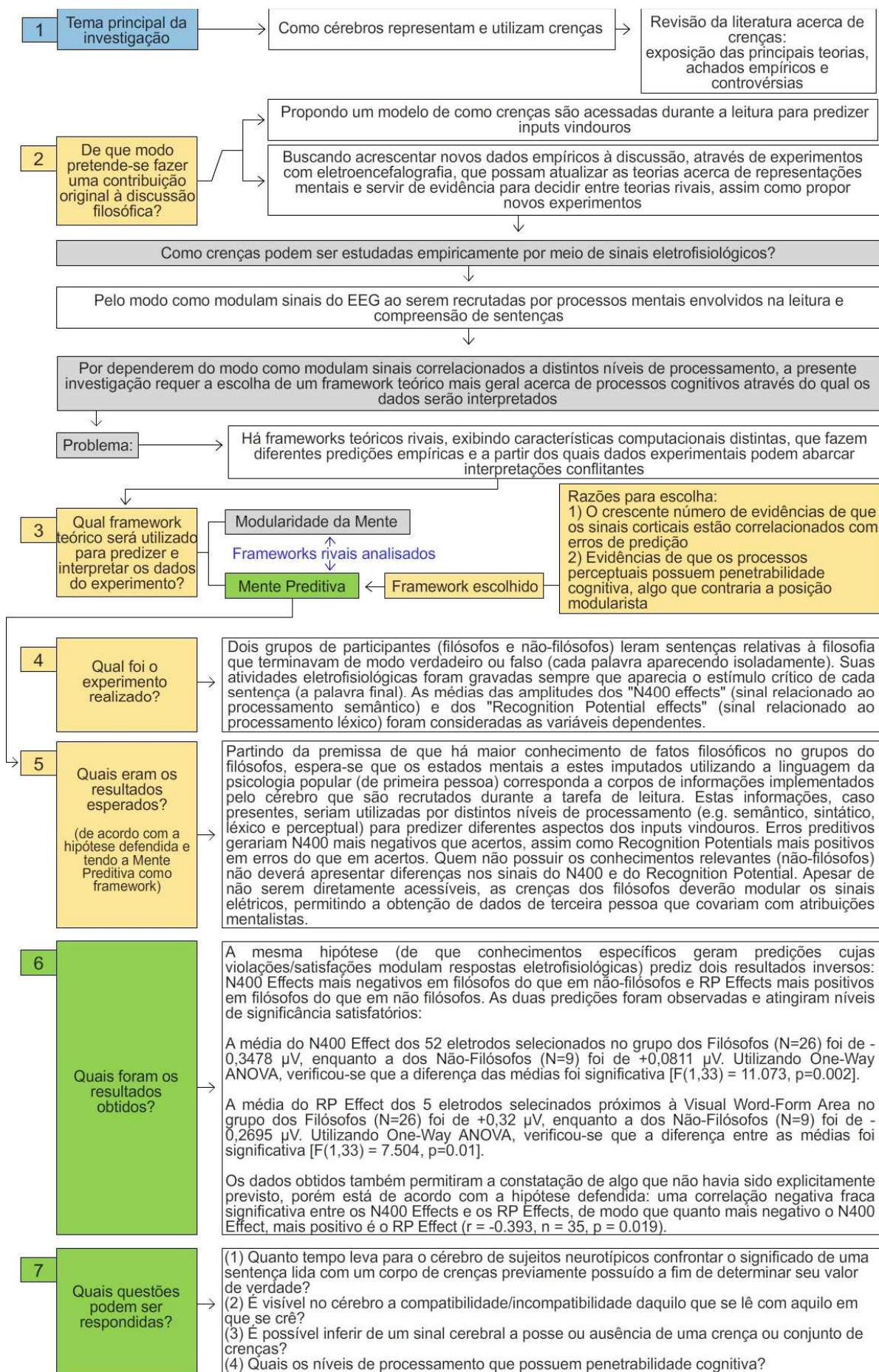
(8) Como o contexto (sentencial ou de inserção do indivíduo) constrange as representações que são acessadas pela memória de trabalho?

Espera-se, no decorrer da argumentação dos capítulos seguintes, demonstrar a viabilidade da presente abordagem filosófico-experimental para ajudar a elucidar essas questões. Os resultados do experimento exposto no capítulo 5 constituem a base para responder às perguntas (1), (2), (3) e (4). Já as questões (5), (6), (7) e (8), embora não possam ser prontamente respondidas pelos resultados atuais, são os

desiderata que poderiam ser solucionados pelo uso da mesma técnica e pressupostos teóricos em experimentos futuros.

Uma vez que este é um estudo acerca de *estados mentais* que, embora sejam informações armazenadas em cérebros que estão à disposição para recrutamento, não são visíveis nos instrumentos de aferição da atividade cerebral enquanto estão num estado de latência, é necessário estudá-los enquanto integram *processos mentais*, estes encadeados por eventos críticos e ocorrendo em intervalos fixos de tempo e de modo serial. É através do recrutamento de crenças por processos mentais que elas, ainda que indiretamente, manifestam-se ao modularem os sinais fisiológicos objetivamente analisáveis. Devido ao fato da presente investigação acerca de crenças depender do modo como elas modulam sinais correlacionados a distintos níveis de processamento, é necessário a utilização de um *framework* teórico mais geral acerca de processos mentais através do qual os dados serão interpretados. Como há *frameworks* rivais – a Modularidade da Mente e a Mente Preditiva –, exibindo características computacionais distintas, que fazem diferentes previsões empíricas e a partir dos quais dados experimentais podem abarcar interpretações conflitantes, será necessário escolher aquele que, após a análise de suas virtudes explicativas e deficiências, se afigurar como o mais adequado para experimentos com eletroencefalografia.

A estrutura da argumentação desta dissertação está resumida no esquema da Figura 1, que expõe de modo simplificado como será abordado o tema principal da investigação, assim como demonstra as relações de dependência que são razões para que se trate de temas paralelos. O esquema também apresenta detalhes sobre o experimento realizado, quais eram resultados esperados de acordo com o *framework* teórico utilizado (i.e. Mente Preditiva), uma prévia dos dados obtidos, assim como procura destacar que questões ele pode elucidar e quais podem ser tratadas em experimentos futuros.



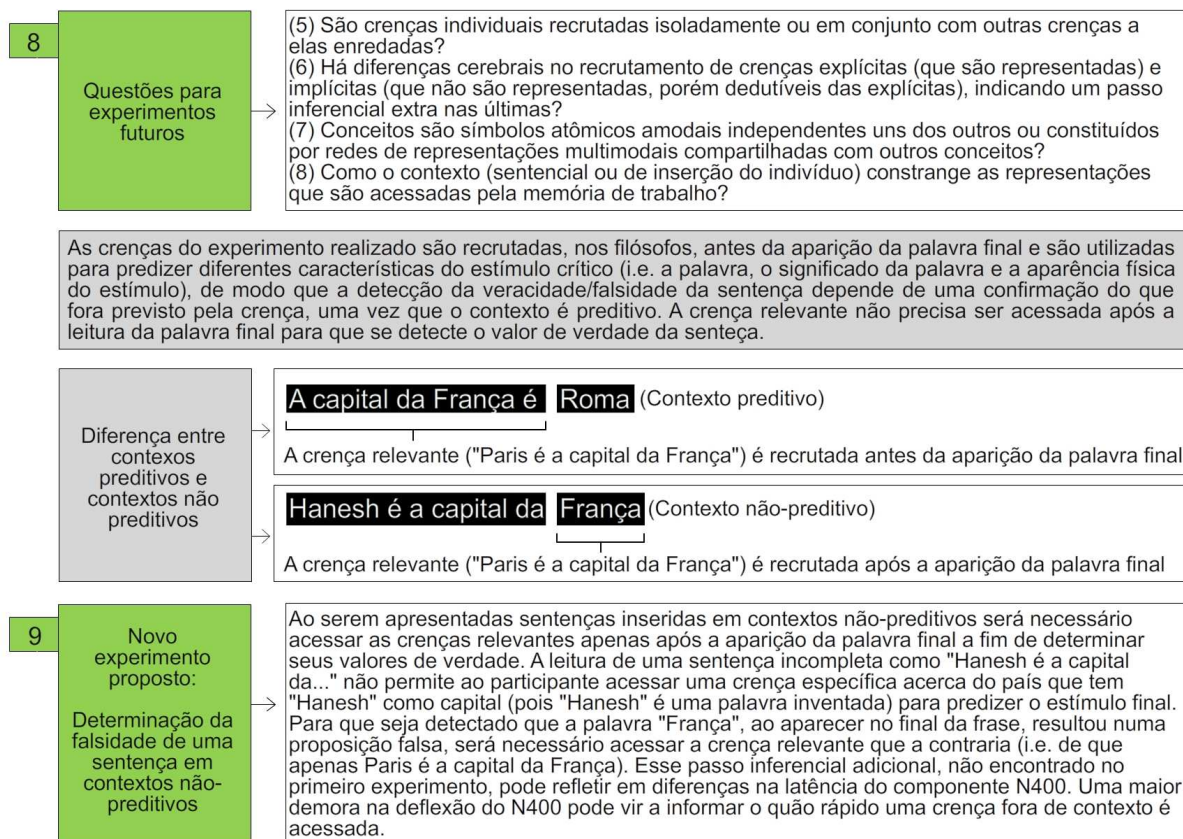


Figura 1 – Esquema argumentativo da dissertação.

2 FRAMEWORKS TEÓRICOS

Uma teoria que tem por escopo explicar como representações são implementadas por cérebros e recrutadas por diversos processos cognitivos necessita situar-se em um *framework* teórico mais geral que a constranja e lhe dê sustentação. Se cérebros humanos operam de acordo com certos princípios, não deve uma teoria acerca de um processo cognitivo específico violá-los, mas, ao contrário, ser construída de modo a encaixar-se no paradigma selecionado. Não se busca apenas formular modelos computacionalmente viáveis, mas também neuralmente realizáveis. Os dois frameworks teóricos que aqui foram escolhidos para análise – a Modularidade da Mente de Jerry Fodor e a Mente Preditiva defendida por Karl Friston, Andy Clark e Jakob Hohwy – são modelos rivais sobre como a mente opera e exibem características computacionais distintas. Neste trabalho será dada ênfase a uma diferença específica: a penetrabilidade cognitiva² dos sistemas perceptuais. Para o modularista, a percepção é informacionalmente encapsulada³ e processa os inputs de modo *bottom-up*⁴. Já para os proponentes da Mente Preditiva, os inputs recebidos são comparados com os inputs previstos, estes gerados de modo *top-down* por processos de níveis superiores.

Dado que o tema da presente dissertação também é abordado empiricamente através de experimentos com respostas cerebrais e que muitos dos sinais eletrofisiológicos medidos são modulados por diferentes níveis de processamento (recrutamento de crenças, ativação de redes semânticas, processamento sintático, léxico, percepção de formas), faz-se necessário selecionar um *framework* a partir do qual os dados serão interpretados. Uma mente modular faz predições empíricas distintas de uma mente preditiva no que concerne ao modo como processos perceptuais interagem com processos mais gerais. Dependendo do framework, um mesmo dado pode abarcar explicações contraditórias. A interpretação de um sinal

² A definição de "penetrabilidade cognitiva" utilizada é a de Pylyshyn (1999): Um sistema perceptual é cognitivamente penetrável se "a função que ele computa é sensível, numa maneira semanticamente coerente, às crenças e metas do organismo, isto é, que possam ser alteradas em uma maneira que tenha alguma relação lógica com o que a pessoa sabe" (PYLYSHYN, 1999, p. 343).

³ A definição de encapsulamento informacional é encontrada na página 20.

⁴ Processos *bottom-up* são impulsionados por informações sensoriais provenientes do mundo físico. Processos *top-down* ativamente buscam informações sensoriais e são impulsionados por nossos conhecimentos, crenças, expectativas e metas.

relacionado a um processo perceptual depende da aceitação da premissa de que os sistemas perceptuais são encapsulados (posição modularista). Havendo suficiente evidência de que mesmo a percepção *low-level* tem acesso a informações localizadas em outros níveis (como regras sintáticas e redes semânticas) deve-se rejeitar a premissa assumida pela teoria modular. Como o experimento tratado no capítulo 5 procura analisar como as crenças individuais modulam respostas cerebrais associadas a diferentes níveis de análise, diferenças de atividade cerebral em função de crenças em processos não-globais constituiriam contraevidências ao encapsulamento, contrariando a modularidade destes processos.

2.1 FRAMEWORK 1 - MODULARIDADE DA MENTE

No livro *Modularity of Mind* (1983), Jerry Fodor defende que a mente é formada tanto por sistemas modulares (*modular systems*) geneticamente determinados – cujo funcionamento é autônomo, inconsciente e de domínio específico – como por sistemas centrais (*central systems*) – conscientes e de domínio geral. Sistemas modulares são, por definição, “mecanismos computacionais de domínio específico” (FODOR, 1983, p. 120), isto é, são individuados pelas funções que realizam. Módulos restringir-se-iam a um domínio específico de inputs sobre o qual executariam computações. De acordo com Jesse Prinz (2006):

Dizer que módulos são de domínio específico é dizer que eles são exclusivamente dedicados “aos seus assuntos”. [...] Considere as colunas de células no córtex visual primário que são utilizadas para detectar arestas. Estas células podem ser dedicadas a esta função apenas. Talvez módulos sejam deste modo (PRINZ, 2006, p. 7).

Além de serem de domínio específico, os módulos são caracterizados por Fodor como apresentando outras 8 propriedades típicas:

- (1) *Mandatory operation*: módulos operam de maneira automática (não podemos evitar que o módulo compute *inputs* da classe em que é especializado, uma vez que o processamento não está sob controle consciente e não pode ser “abortado”)

- (2) Acessibilidade limitada: as computações intermediárias entre o *input* e o *output* do módulo não são (ou pelo menos não são totalmente) acessíveis à consciência
- (3) Rápido processamento: módulos produzem outputs de modo eficiente
- (4) Encapsulamento informacional: as informações à disposição do módulo para computar inputs são limitadas (e.g. mesmo algumas informações que seriam cruciais à computação do módulo e que estão acessíveis à consciência – como crenças possuídas pelo indivíduo – podem estar interditas ao módulo)
- (5) "*Shallow*" outputs: os *outputs* do módulo são muito simples em termos de conteúdo, não sendo resultado de um processamento sofisticado (demonstram às vezes uma certa "estupidez")
- (6) Arquitetura neural fixa: módulos são implementados em regiões neurais específicas
- (7) Padrões específicos de danos: módulos podem ser seletivamente danificados, acarretando déficits dissociados da inteligência geral do indivíduo
- (8) Determinação ontogenética: módulos desenvolvem-se em um ritmo e sequência característicos, através de maturação (e não pela aprendizagem)

Não é por acaso que estas características se encontram agrupadas em módulos. A rapidez de processamento (3), por exemplo, seria consequência do encapsulamento informacional (4): assumindo que a velocidade de processamento varia inversamente com a quantidade de informações que um sistema consulta, o fato de ser encapsulado permite ao módulo ignorar muitas informações, resultando em eficiência computacional. Lesões em regiões associadas a um módulo podem

danificá-lo (7), uma vez que módulos possuem localização fixa (6), comprometendo apenas o específico domínio sob o qual atuam, sem prejudicar capacidades paralelas. Encapsulamento informacional (4) e limitada acessibilidade (2) seriam "dois lados da mesma moeda" (ROBBINS, 2009): ambas as propriedades referem-se ao fluxo de informações entre sistemas. Encapsulamento envolve restrição do fluxo de informações para dentro do módulo, enquanto inacessibilidade tem a ver com a opacidade do que ocorre dentro do módulo à introspecção. Apenas o produto final do módulo seria fenomenologicamente saliente: uma analogia que pode ser feita é comparar o output do módulo com um presidente que não pode frequentar todas as reuniões de seu interesse, podendo, todavia, ser informado das decisões finais, sem precisar tomar conhecimento de todos os detalhes e ideias rejeitadas que antecederam o que lhe foi fornecido (cf. PRINZ, 2006, p. 9). O *output* do módulo seria simples (5) também devido ao encapsulamento informacional (4), que o impediria de acessar "todas as informações relevantes que o organismo possui". Um exemplo clássico de processamento de informações que apresenta características modulares são ilusões de óptica (o exemplo mais utilizado é a *Mueller-Lyer illusion*). Considere a imagem abaixo:

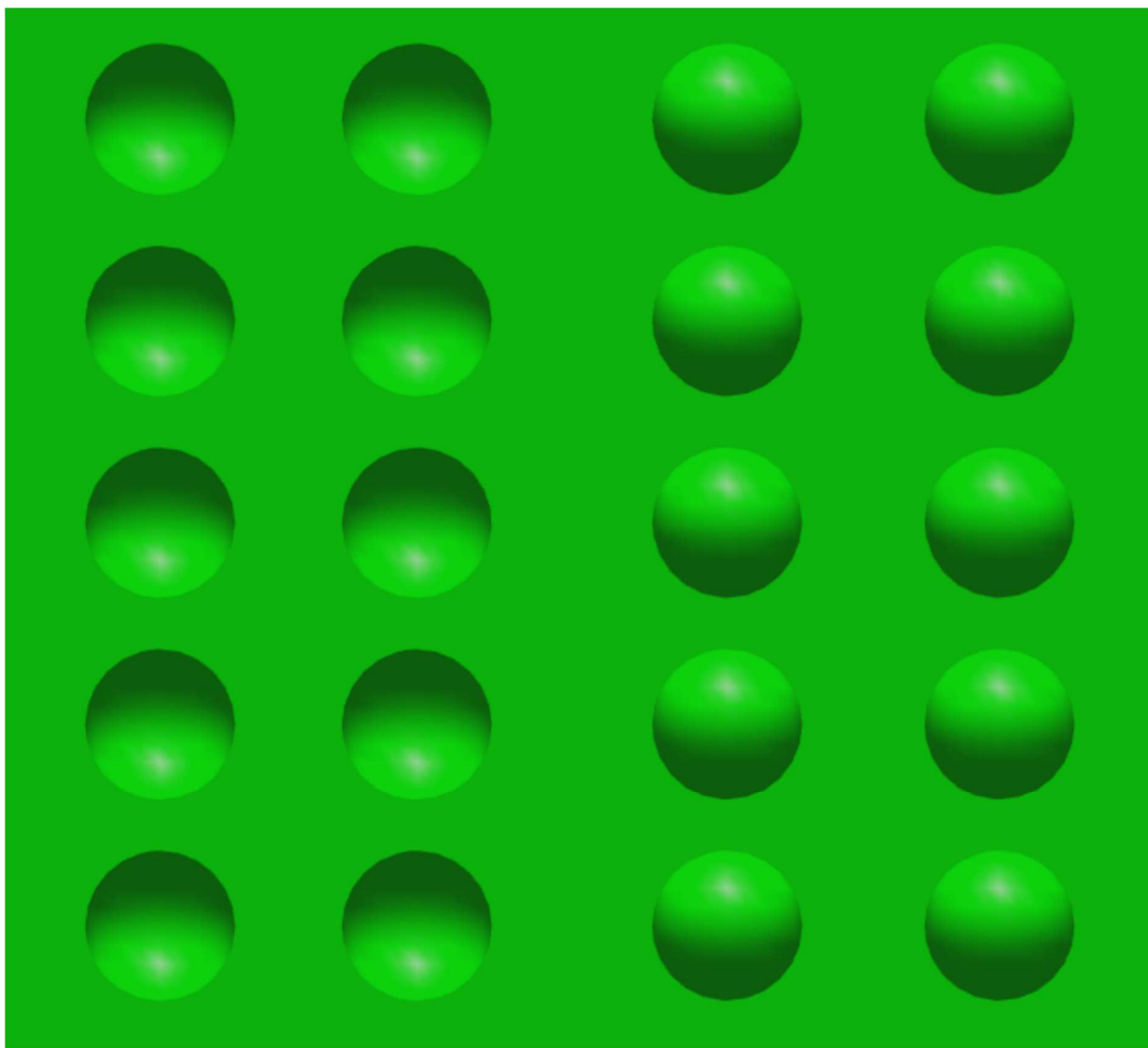


Figura 2 – Representação bidimensional de superfícies côncavas e convexas

Na figura 2, as duas colunas da esquerda parecem ser côncavas e as duas colunas da direita parecem convexas. Estas categorizações são rápidas (3), automáticas (1) e não requerem sofisticação de raciocínio (não é necessário possuir noções de óptica, muito menos imaginar quais seriam os gradientes de luz e sombra formados numa imagem bidimensional a partir de informações sobre o posicionamento da fonte de iluminação e da curvatura de cada superfície – empresa esta que está longe de ser automática). Entretanto, o preço desta eficiência será a falta de flexibilidade: como fora dito, a rapidez é conquistada pelo encapsulamento informacional, e este impede que o módulo utilize informações cruciais das quais o sujeito tem consciência. O *input* da figura 2, por exemplo, é ambíguo, e se formos pensar explicitamente e de modo delongado, chegaremos à conclusão de que é

impossível determinar quais superfícies são côncavas e quais são convexas, uma vez que se ignora a posição da fonte de iluminação (pode estar em cima ou embaixo, sendo a concavidade/convexidade das superfícies função desta variável). O módulo, no entanto, toma como premissa invariável que "a iluminação vem de cima", resultando num *output* fenomenologicamente acessível que nos induz a ver as superfícies da direita como côncavas e as da esquerda como convexas. Mas se a luz vier de baixo, então, se raciocinarmos de modo explícito, concluiremos que o que é o caso é na verdade o oposto do que *parece* ser o caso: as superfícies da direita revelar-se-iam convexas e as da esquerda côncavas. Ainda que estejamos cientes de que a fonte de iluminação está embaixo e não em cima (premissa esta determinante), continuaremos a ter a errônea impressão inicial. O módulo, por ser encapsulado, ignora uma informação que lhe é crucial e da qual estamos conscientes, continuando a assumir como premissa que a luz vem de cima, não admitindo revisar seu conhecimento do mundo, ainda que se acumulem contraevidências.

Módulos entregam seus outputs aos sistemas centrais. Fodor afirma que são estes os sistemas que temos em mente quando falamos, pré-teoreticamente, de processos mentais conscientes como *o pensamento e resolução de problemas*⁵. Se os módulos fornecem respostas apenas a uma classe restrita de inputs (por serem de domínio específico), consultando um estoque limitado de informações (por serem informacionalmente encapsulados), os sistemas centrais, por sua vez, operariam com as propriedades inversas: são de "domínio geral" e "cognitivamente penetrantes". A função típica dos sistemas centrais seria, para Fodor, a "fixação de crenças":

Os sistemas centrais olham para o que lhes é entregue pelos sistemas de input, e eles olham para o que está na memória, usando estas informações para chegar à computação das 'melhores hipóteses' acerca de como é o mundo (FODOR, 1983, p. 104).

Sistemas centrais são, na terminologia de Fodor, isotrópicos (*isotropic*) e "quineanos" (*quinean*). Isotropia se refere ao holismo epistêmico: dizer que um sistema é isotrópico é afirmar que, em princípio, qualquer informação que o organismo possua pode vir a ser relevante para o grau de confirmação de uma hipótese (opõe-se, portanto, ao encapsulamento informacional, em que até mesmo algumas

⁵ Cf. FODOR, 1983, p.103.

informações determinantes e das quais estamos conscientes são ignoradas). “Quineano” (relativo às teorias do filósofo Willard van Orman Quine) é uma característica que se refere ao fato de o grau de confirmação atribuído a uma certa hipótese ser influenciado por propriedades emergentes do sistema inteiro de crenças, tais como simplicidade, plausibilidade e conservadorismo. A posse dessas duas propriedades por um sistema requer deste acesso potencialmente ilimitado à memória (cf. ROBBINS, 2009) Sistemas centrais também são tipicamente lentos – em comparação aos sistemas modulares –, conforme argumenta Fodor:

Uma pessoa pode – e o faz frequentemente – passar horas pensando em um problema de filosofia ou xadrez, embora não haja motivos para supormos que a complexidade computacional destes problemas seja maior que a daqueles problemas que são rotineiramente resolvidos sem esforço por processos perceptuais (FODOR, 1983, p. 63).

Se o custo da eficiência dos sistemas modulares é a restrição de informações disponíveis para consulta, a flexibilidade dos sistemas centrais, em contrapartida, acarreta usualmente maior tempo de processamento. Um modo como este inconveniente pode ser contornado por processos centrais é utilizando heurísticas, que são “métodos para chegar a soluções satisfatórias com uma quantidade modesta de computação” (SIMON, 1990, p. 11). Heurísticas são atalhos mentais que reduzem o tempo e esforço associados a uma tarefa ao utilizar apenas uma subclasse das informações relevantes e/ou processando informações com base em simples regras. É possível ilustrar este ponto utilizando o mesmo exemplo da Figura 2. Se quisermos identificar superfícies côncavas e convexas sem depender da impressão passiva que o output do módulo nos fornece, então teremos que pensar de modo consciente e sequencial, fazendo uso daquilo que aprendemos sobre o mundo (em vez de noções inatas) e que pode ter alguma relevância para a presente tarefa. Todavia não existirá, neste caso, um método único de processamento, mas uma pluralidade de inferências aptas a levar-nos à correta categorização das superfícies: umas serão mais frugais e rápidas e demandarão pouco esforço cognitivo, enquanto outras serão mais complexas, lentas e cognitivamente exigentes. Assim como alguém pode fazer uso de seus conhecimentos de óptica e geometria para inferir que, se a fonte de iluminação estiver embaixo, serão convexas as superfícies que tiverem a região inferior iluminada e côncavas as que tiverem iluminada a região superior, um outro indivíduo pode optar pelo uso de uma regra mais frugal, tal como: “Se uma superfície parecer

côncava/convexa e a fonte de iluminação estiver localizada em cima, então ela é tal como parece ser; se a fonte de iluminação estiver embaixo, então a superfície tem a curvatura oposta à que parece ter”. Esta última, não obstante o fato de não ser de natureza modular e nem mesmo automática, permite-nos chegar ao mesmo resultado da primeira, de um modo computacionalmente mais parco e eficiente. O sistema central, desta forma, não apenas poderia corrigir o que lhe é entregue pelo módulo, como também poderia escolher, dentre muitas alternativas de correção, aquela que julgar a mais apropriada a cada caso (e.g. é razoável utilizarmos heurísticas quando inseridos numa situação que nos demande rápida resposta, mesmo quando possuímos meios mais complexos e ideais – e conseqüentemente mais lentos – de processar a mesma informação).

Para Pylyshyn (1999), que defende uma teoria de teor modularista acerca dos processos perceptuais, uma importante parte da visão, denominada “*early vision system*”, faria seu trabalho sem a intervenção de conhecimentos, crenças ou expectativas, até mesmo quando o uso de conhecimentos pode prevenir o sujeito de cometer erros⁶. Essa visão da percepção como um processo construtivo *bottom-up* considera os *input systems* como mecanismos que capturariam as informações que nos cercam de modo totalmente passivo. Lupyan (2015) descreve como é problemática, considerando-se a percepção um módulo encapsulado, a interferência de sistemas globais:

Assim como velas capturam o vento, nossos olhos capturam fótons, nossos ouvidos vibrações, nossa pele forças mecânicas e térmicas e, após transduzi-los em sinais neurais, o que muitos textos se referem como ‘processamento posterior’, provêm as sensações da visão, audição, calor, etc. Nessa visão, devido ao fato de estados cognitivos serem as conseqüências de causas perceptuais, afirmações de que estados cognitivos podem afetar a percepção são tomados como estando em conflito com a causalidade básica (LUPYAN, 2015, p. 4).

Se o presente framework estiver correto, sinais fisiológicos provenientes de sistemas perceptuais, não podendo processar seus inputs relevantes influenciados por informações presentes em níveis superiores (como crenças), não devem ser modulados pelo conteúdo desses últimos. A alteração no processamento perceptual *low-level* em função das crenças do organismo indicaria interferências *top-down*

⁶ PYLYSHYN, 1999, p. 343.

proibidas pela teoria modularista. No subcapítulo seguinte é exposto um framework teórico rival que parte da premissa oposta: níveis de processamento inferiores são informados previamente pelos processos globais de maneira antecipatória (quando a circunstância de inserção do indivíduo assim o permitir). Desse modo, espera-se que o processamento perceptual low-level seja modulado por crenças – resultado contrário ao previsto por aquelas que consideram os *input systems* módulos encapsulados.

2.2 FRAMEWORK 2 - PREDICTIVE BRAIN

Afastando-se das abordagens tradicionais das ciências cognitivas que consideram a percepção e cognição como processos acumulativos e construtivos de detecção de características de modo *bottom-up* (e.g. Fodor, 1983), o framework emergente denominado Predictive Brain (também referido como *Predictive Processing*, *Predictive Coding* e *Bayesian Brain*) pretende explicar percepção, cognição e até mesmo ação em termos de princípios teóricos e arquiteturas neurocognitivas fundamentais exibindo uma estrutura preditiva *top-down* (SETH, 2016, p. 1). Influenciado pelo trabalho de Helmholtz (2005), que considerava a percepção um processo de inferência probabilística guiado por conhecimentos prévios, o atual modelo proposto por Friston (2005) e defendido no âmbito da filosofia da mente por filósofos como Jakob Hohwy (2015) e Andy Clark (2015, 2016) não é apenas plausível em termos computacionais, mas também neuralmente implementável, tendo em vista que os processos mentais/cerebrais estão constrangidos por princípios físicos mais fundamentais, como a energia livre termodinâmica (*thermodynamic free-energy*). Sendo a mente produto de processos cerebrais e cérebros situados em um meio físico, não pode a cognição operar fora das restrições impostas pela física mas, ao contrário, emergir dentro dessas.

Em sua concepção mais simplificada, o Predictive Brain (PB) assume que a percepção é o resultado do cérebro tentando inferir a causa mais provável dos inputs sensoriais ao minimizar a diferença entre os inputs recebidos e os inputs esperados com base em modelos preditivos continuamente atualizados. O *percept* (resultado mental ou produto da percepção) é concebido de modo semelhante a uma inferência para a melhor explicação (abdução lógica): dos efeitos (no presente caso, impressões sensoriais) busca-se as hipóteses que se afiguram como as causas mais prováveis,

levando em considerações as experiências passadas que resultaram em representações de regularidades encontradas na interface mente-mundo (na linguagem do PB, essas representações são como *probability density functions*⁷). Porém, diferentemente de uma abdução – fundamentada no conceito de explicação – , o PB baseia-se em mecanismos que se ajustam de maneira preditiva, resultando em algo que poderia ser denominado “inferência para a melhor predição”⁸. Andy Clark ressalta a mudança desse modelo de cognição proativa em relação à visão tradicional focada na reação passiva a estímulos:

O processamento preditivo plausivelmente representa o último e mais radical passo nessa retirada da visão passiva – dominada por inputs – acerca do processamento neural. De acordo com essa classe emergente de modelos, sistemas naturalmente inteligentes (humanos e outros animais) não esperam passivamente estimulação sensorial antes de ela chegar. Antes de um ‘input’ chegar na cena, estes sistemas cognitivos proativos já estão ocupados predizendo sua forma mais provável e implicações (Clark, 2015, p. 2).

O "truque" utilizado pelo cérebro para processar estímulos com parcimônia computacional é a utilização de um *forward model*⁹ que gera virtualmente o input esperado antes de sua ocorrência. A diferença entre o sinal esperado e o sinal previsto, denominada “erro de predição”, é o *feedback* que os níveis de processamento inferiores transmitem aos laterais/superiores. Erros de predição sinalizam o quanto o mundo desvia da representação de mundo individual. Enquanto acertos preditivos corroboram as hipóteses sobre as regularidades encontradas no mundo externo (pois são, em termos lógicos, equivalentes afirmações do consequente), não necessitando atualizar o modelo interno, erros de predição forçam a formulação de novos modelos (ou a atualização dos antigos) de modo a evitar erros futuros (porque são negações do consequente, das quais se segue a falsidade do antecedente). Pode-se afirmar que os modelos se autorregulam conforme a seguinte dinâmica: (1) reconhecimento da condição inicial fornecida pela circunstância x, (2) escolha da hipótese/modelo que putativamente melhor prediz o desenvolvimento da

⁷ Uma função que descreve a probabilidade relativa de uma variável aleatória tomar um valor dado.

⁸ Não haveria necessidade de distinguir inferência para a melhor explicação de inferência para a melhor predição se assumido o modelo proposto por Hempel (2001), que considera haver uma simetria entre explicação e predição, sendo apenas modos diferentes de ler um mesmo argumento (a conclusão sendo prevista como consequência lógica das premissas e estas consistindo em explicações do fato referido pela conclusão). Entretanto, o presente trabalho assume que nem sempre argumentos que suportam predições conseguem explicar o que é previsto (cf. SALMON & WESLEY, 2006).

⁹ Um mecanismo que prediz o estado futuro de um sistema (cf. PICKERING & CLARK, 2014).

circunstância x , (3) utilização da hipótese selecionada conjuntamente com a condição inicial da circunstância x para prever e gerar virtualmente o input seguinte em vários níveis de processamento, (4) recebimento do input esperado, (5) comparação entre o input recebido e o input esperado, (6) transmissão aos níveis superiores do erro de predição encontrado, (7) refinamento/alteração da hipótese de modo a prever melhor o sinal, (8) utilização da nova hipótese (ou da mesma hipótese, em caso de acerto) para prever e gerar virtualmente o input seguinte em vários níveis de processamento, repetição da etapa (4) e seguintes. A elaboração de modelos mentais que preveem cada vez melhor os estímulos com base em erros de predição tem como resultado um menor gasto energético, uma vez que o sistema busca minimizar cada vez mais o erro de predição enviado aos níveis superiores. Outra questão importante é o grau de acerto preditivo: uma hipótese pode acertar boa parte do sinal e mesmo assim haver um erro residual inexplicado pela mesma. Manter ou descartar a hipótese dependerá do grau de precisão demandado pela circunstância. Dependendo do contexto, geram-se predições mais ou menos precisas, para as quais detalhes sensoriais que as contrariam podem ser de grande importância ou simplesmente irrelevantes. Deslocar-se pela própria casa no escuro, por exemplo, é uma tarefa que não requer muita atenção aos detalhes sensoriais visuais (pois é dada maior importância ao modelo virtual do que ao input), enquanto que ao guiar um carro em uma estrada desconhecida será dada maior importância aos sentidos.

O mecanismo do Predictive Brain é implementado pelo cérebro de modo hierárquico, sendo que os níveis superiores (mais gerais e abstratos) informam os níveis inferiores (mais espacial e temporalmente precisos) e preveem o input que virá destes, cada nível analisando aspectos distintos dos estímulos. As predições ocorrem na forma de uma cadeia *top-down*, havendo um fluxo de informações que atinge até mesmo os processos perceptuais *low-level*, que desta forma têm acesso às crenças relevantes do sistema muito antes de receberem estímulos (logo, segundo esse modelo, a percepção apresenta penetrabilidade cognitiva, contrariando a visão modularista que a considera informacionalmente encapsulada). Em vez de os inputs serem analisados de modo gradual e acumulativo partindo dos processos perceptuais *low-level* e sendo contextualizados e confrontados com as crenças apenas pós-perceptualmente, o PB assume que a informação que sobe dos níveis inferiores aos superiores é apenas o erro de predição encontrado em cada camada (erros estes que são produto de predições geradas pelos níveis superiores). Não há a necessidade do

nível inferior passar adiante todo o sinal recebido originalmente, mas apenas o erro. Em um sistema cognitivo organizado dessa maneira é possível que um nível receba pouca informação (e por conseguinte economize energia) e ainda assim ter acesso aos mesmos dados que teria caso os tivesse recebido integralmente. Por exemplo, em uma situação ideal, em que o input recebido é idêntico ao input previsto, não há a necessidade de transmitir adiante o sinal, pois a ausência de recebimento de um sinal de erro já informaria a mesma situação com um gasto energético nulo. Em uma circunstância normal, um modelo que acertou boa parte dos inputs só precisa ser informado dos desvios, podendo dar pouca importância a detalhes irrelevantes (ao nos deslocarmos de um ponto a outro em uma cidade conhecida, não prevemos contingências como os carros estacionados ou as cores das roupas das pessoas, mas operamos em um nível que considerará como erro de previsão eventos como ruas bloqueadas ou prédios demolidos que estavam presentes no modelo recrutado).

2.2.1 Implementação neural e evidências provenientes de respostas corticais

De acordo com o framework do PB, as *representation units* codificam as previsões *top-down* supostamente provenientes das células piramidais profundas, enquanto as *error units*, geradas pelas células piramidais superficiais, o fazem com os erros de previsão no sentido *bottom-up*¹⁰. As respostas cerebrais evocadas por estímulos, como os *Event-related potentials* (ERPs) do EEG¹¹, são interpretadas pelo PB como sinalizando graus de erros preditivos: quanto maior a diferença entre o estímulo recebido e a previsão, maior será o erro de previsão e conseqüentemente maior será a resposta evocada¹². Muitos componentes do ERP são de fato modulados em função de desvios em relação ao previsto pelo tipo de processo. O P300

¹⁰ Cf. FRISTON, 2005.

¹¹ O eletroencefalograma mede a atividade eletrofisiológica do cérebro por meio de eletrodos posicionados no escalpo. A técnica de *Event-Related Potentials* (ERP's) consiste em isolar a atividade cerebral evocada pelos estímulos da atividade de fundo (processos dissociados do estímulo crítico e que ocorrem paralelamente a este), permitindo enxergar as áreas do cérebro e os componentes do ERP que são elicitados apenas em função da natureza do estímulo apresentado. As regiões ativadas e as amplitudes dos sinais eletroencefalográficos são dados que auxiliam na compreensão dos processos cognitivos estudados de um modo interdito à tradicional abordagem estímulo-comportamento. As já documentadas correlações entre os componentes do ERP e certos processos cognitivos (e.g. o N400 e o acesso semântico ou a supressão do ritmo mu e a previsão de ações) permitiriam testar e aprimorar hipóteses filosóficas. Para uma explicação detalhada de como funciona a eletroencefalografia e a técnica de ERPs, ver LUCK, 2005.

¹² Cf. BOWMAN et al, 2013.

(positividade amplamente distribuída no escalpo, mais concentrada na região central, que atinge seu pico comumente 300 milissegundos após um evento), é um componente do ERP que apresenta um sinal mais atenuado para estímulos mais prováveis e mais forte para estímulos improváveis (probabilidade subjetiva)¹³. O participante é exposto a repetições de dois tipos de imagens ou letras (e.g. X e O) nas quais um estímulo é mais frequente que o outro (e.g. 20% de “X” e 80% de “O”). Após acostumar-se com aparições mais frequentes de “O”, esperar-se-á que a probabilidade de os itens futuros serem “O” seja maior que “X”. O erro de predição supostamente sinalizado por maiores amplitudes do P300 após “X”, entretanto, parece ser gerado por um nível de processamento que leva em conta aspectos perceptuais mais *low-level*. Já o componente N400 (deflexão negativa, concentrada na região centro-parietal, cuja latência ocorre aproximadamente 400 milissegundos após um estímulo), está principalmente associado à dificuldade de integrar uma palavra – em termos semânticos – ao contexto em que ela aparece. Kutas & 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 nas anomalias semânticas, os autores se depararam, acidentalmente, com o N400. Quanto maior a probabilidade subjetiva de uma determinada palavra P aparecer em uma sentença incompleta (i.e. *cloze probability*), maior a negatividade do N400 quando, ao invés de P, outra palavra aparecer (negatividade 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 “pera” aparecendo no lugar de “maçã”¹⁴). O N400 não se restringe a uma única modalidade de input (é evocado por palavras escritas, faladas e por linguagem de sinais)¹⁵, sendo elicitado também por imagens que não se encaixam no contexto sentencial (GANIS et al, 1996). Apesar de o P300 e o N400 serem ambos componentes que apresentam sinais evocados mais fortes em violações de expectativa em função da probabilidade subjetiva dos itens apresentados – constituindo evidências favoráveis ao PB –, seus níveis de análise são diferentes: o P300 parece sinalizar erros de predição em um nível que considera aspectos mais

¹³ Cf. LUCK, 2005, p. 43.

¹⁴ Cf. FEDERMEIER & KUTAS, 1999.

¹⁵ Cf. KUTAS & VAN PETTEN, 1994.

físicos do estímulo, enquanto o erro de predição do N400 ocorreria em um nível semântico, após a identificação do significado cognitivo do estímulo (e.g. o conceito de cachorro após a aparição do estímulo “cachorro”). A expectativa do P300 provém de uma detecção de curto prazo da frequência de aparição de dois estímulos (adquirida durante o experimento); a expectativa do N400, entretanto, depende de uma rede semântica estabelecida previamente, que considera a plausibilidade do estado de coisas representado.

O componente P600 (positividade centro-parietal cujo pico de atividade ocorre frequentemente aos 600 milissegundos) é tipicamente modulado por propriedades sintáticas das sentenças lidas¹⁶. Erros gramaticais produzem um P600 mais forte do que frases bem formadas (estas desenvolvendo-se em conformidade com as restrições sintáticas impostas pela sentença em evolução). Patel et al (1998), buscando verificar se o P600 era um componente específico da linguagem ou originário de processos mais gerais envolvendo violações de sequências governadas por regras, encontraram em erros de sintaxe musical amplitudes maiores de P600 do que em acertos (quanto maior a distância do acorde recebido em relação ao esperado, maior o P600). Desse modo, assim como o N400, o P600 parece ser um sinal que acompanha um processo pós-perceptual que, embora atue sobre um domínio específico (i.e. sintaxe), recebe informações provenientes de diferentes modalidades sensoriais.

Thornhill e Van Petten (2012), procurando identificar um componente do ERP que fosse modulado por predições léxicas, desenharam um experimento em que os participantes liam frases que permitiam fazer predições precisas quanto à palavra final. Tendo em vista que o N400 de palavras inesperadas que são semanticamente próximas às esperadas é mais atenuado do que o N400 de palavras semanticamente distantes, os autores puderam dissociar predições de palavras de predições de conceitos (significados cognitivos das palavras). Foi observada uma positividade frontal no intervalo de 600 a 900 milissegundos após a palavra final, denominada PNP (*post-N400 positivity*). Assim como o N400, o PNP foi menor para palavras previsíveis do que imprevisíveis. Entretanto, a amplitude do PNP foi insensível à similaridade semântica das palavras inesperadas em relação às esperadas. Essa dissociação indicaria “que a positividade fronto-central reflete um nível de análise distinto do N400,

¹⁶ Cf. GOUVEA et al, 2010.

que é mais sensível a fatores léxicos do que semânticos” (THORNHILL & VAN PETTEN, 2012, pp. 390-391). O sinal do PNP apresenta uma manifestação dicotômica: ou a palavra lida é a palavra esperada (ausência de PNP) ou é uma palavra inesperada (PNP presente).

Uma tarefa cognitiva aparentemente simples, como a compreensão de uma frase, envolve uma série de processamentos de distintos níveis, cada um operando sobre o domínio do qual é especializado. O estímulo “cachorro”, aparecendo no final da frase “Scooby-doo é um...”, primeiramente será processado pelo córtex visual, que analisa propriedades pré-léxicas como formas e aspectos físicos low-level. O reconhecimento de letras e palavras ocorre posteriormente. Após o processamento léxico, que reconheceu no estímulo a realização física de um item léxico abstrato, ocorreriam os processamentos semântico (o significado da palavra e o encaixe semântico do objeto referido pela palavra no estado de coisas descrito) e sintático (reconhecimento da categoria da palavra e o quão adequada é a sua inserção na frase). Numa abordagem *bottom-up* modularista, propriedades sintáticas, léxicas e semânticas não poderiam influenciar o processamento perceptual pré-léxico, uma vez que sistemas perceptuais são informacionalmente encapsulados (podem passar o *output* aos níveis superiores, mas não podem consultar os níveis superiores durante a computação que antecede a formulação do output). Uma arquitetura preditiva como a do Predictive Brain, no entanto, assume que há fluxo de informações entre os diversos níveis de modo *top-down*, podendo os níveis semântico e sintático informarem o nível léxico (o nível semântico informa que o objeto que melhor se encaixa no contexto exemplifica o conceito “cachorro” e sintático informando que a estrutura gramatical da sentença sugere que a palavra seguinte será um substantivo) e este, identificando a palavra “cachorro” como o item léxico mais provável como tendo o conceito cachorro como significado, informar o nível perceptual pré-léxico, traduzindo a palavra recrutada num formato acessível à análise de formas computada pelo córtex visual. Essa hipótese específica foi formulada e testada por Dikker et al. (2010) e Dikker & Pykkänen (2011). Apresentando frases que constroem as expectativas léxicas, semânticas e sintáticas quanto às palavras finais, gerando previsões acerca de diferentes propriedades pós-léxicas dos estímulos vindouros, os autores verificaram efeitos de previsão no córtex visual aparecendo a apenas cerca de ~100 milissegundos após aparecerem as palavras finais. Utilizando magnetoencefalografia, foi possível encontrar maiores amplitudes do componente M100 em estímulos inesperados do que

em estímulos esperados (maiores amplitudes estando correlacionadas a erros de predição). Os resultados encontrados surpreendem pelo fato de um processo *low-level* pré-léxico (e por conseguinte pré-semântico e pré-sintático) apresentar sensibilidade a propriedades que estão fora de seu domínio de análise. Como propriedades semânticas, léxicas e sintáticas podem modular o sinal evocado pelo córtex visual se este apenas analisa formas e aspectos físicos e permanece cego a palavras, conceitos e sintaxe? Com diferentes níveis de processamento colaborando para antecipar seus inputs mais prováveis, é possível que um processo perceptual, mesmo sendo incapaz de processar os aspectos mais abstratos do superior, seja auxiliado por este ao ter acesso prévio a “estimativas de forma” (*form-estimates*) que são traduções de um formato em outro. O córtex visual não precisa consultar os níveis superiores no exato intervalo em que processa seu input, mas apenas fazer uso da informação recebida anteriormente. Considerando que a velocidade de processamento é inversamente proporcional à quantidade de informação que o sistema consulta (esse é o argumento do modularista para explicar a rapidez dos sistemas perceptuais com base no encapsulamento informacional), a estratégia do Predictive Brain (o nível superior informando o nível inferior antes de este receber seu *input*) permite a um sistema possuidor de penetrabilidade cognitiva entregar seu *output* ao nível superior com igual eficiência e parcimônia.

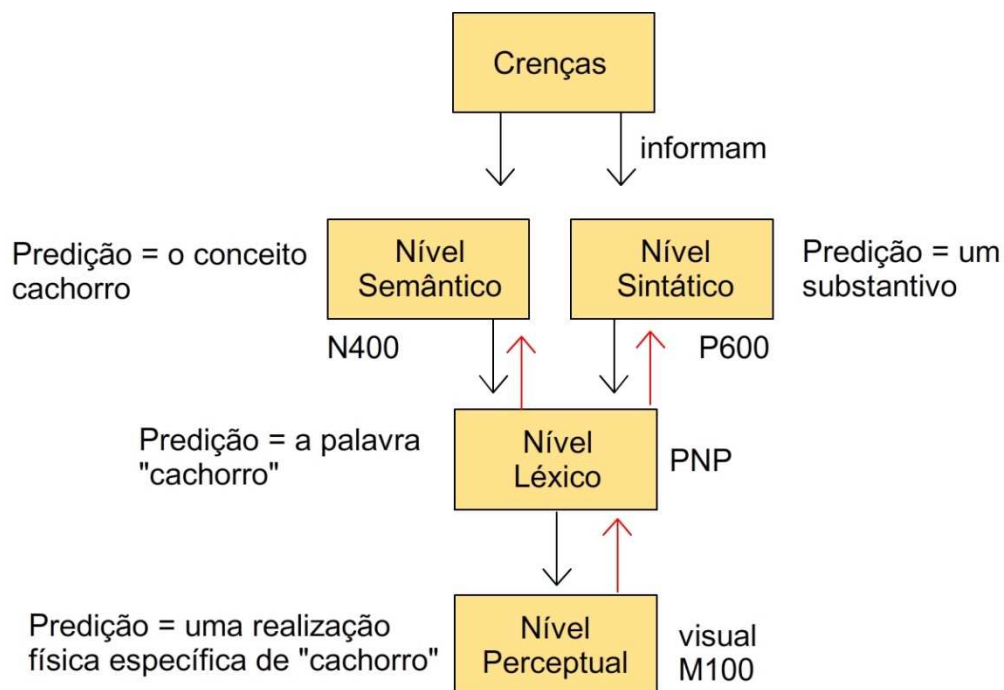


Figura 3 – Esquema representando o fluxo de informação *bottom-up* e *top-down* entre diversos níveis de processamento linguístico.

Outros níveis de processamento de estímulos linguísticos podem ser também analisados isoladamente. O *Recognition Potential* (RP) é um componente cuja negatividade (entre 200 a 250 milissegundos) é mais forte para palavras do que pseudopalavras, e também mais forte nestas do que em uniões aleatórias de letras. Não há consenso se o RP reflete o mero reconhecimento de um estímulo visual ou se está associado também a processamento semântico. O fato do N400 ser elicitado quando temos que integrar uma palavra (ou o objeto referido pela palavra) em uma sentença (ou estado de coisas afigurado pela sentença) indica que o processamento semântico que o N400 reflete é de uma ordem superior ao simples processamento semântico de um termo isolado, que o precederia. Evidências favoráveis à hipótese de que o RP está associado a um processamento semântico anterior ao N400 foram encontradas por Martín-Loeches et al (2001), que verificaram que palavras expressando conceitos concretos evocam um RP mais negativo que palavras que designam conceitos abstratos. A diferença encontrada não poderia ser explicada por outras razões senão às diferenças de significado das palavras, ou ao fato de as concretas evocarem “imagens mentais” mais nítidas do que as abstratas. Martín-Loeches et al (2004) também encontraram RP’s com amplitudes maiores em palavras que se encaixavam no contexto sentencial, em comparação com palavras que

apareciam fora do contexto. Apesar do sinal de erro apresentar amplitude menor que o sinal de acerto (o que pode ser considerado uma anomalia para o PB, visto que erros preditivos deveriam estar correlacionados a amplitudes maiores do sinal), o experimento realizado sugere que o RP não refletiria uma computação passiva da palavra, pois é afetado pelo contexto que a precede.

2.2.2 Tipo de computação implementada e exemplos corroboradores

Ao ser considerado o tipo de computação que o modelo proposto pelo Predictive Brain implementa, a inferência bayesiana desponta como melhor candidata, pois a estratégia do framework realiza algo que se aproxima de maneiras ideais de pesar novas evidências contra conhecimentos prévios, e a busca das causas ocultas que fazem o atual input sensorial o mais provável corresponde à inferência bayesiana (CLARK, pos. 404, 2015). Aplicada ao PB, a inferência bayesiana calcula a probabilidade a posteriori / *posterior probability* (probabilidade da hipótese dado o input/evidência) de uma hipótese ser a causa de um input com base no produto entre a probabilidade a priori / *prior probability* (probabilidade da hipótese antes do input/evidência) e a verossimilhança / *likelihood* (probabilidade do input/evidência dada a hipótese)¹⁷:

$$P(h|i) = P(i|h).P(h)$$

Como dois cenários distintos podem causar o mesmo input ou série de inputs, o cérebro precisa decidir qual deve ser a causa mais provável a fim de determinar o *percept* que será fenomenologicamente acessível. Ao optar por uma hipótese em detrimento de outra, um cérebro bayesiano deverá utilizar como critério aquela que tiver a maior probabilidade a posteriori, o que dependerá do quão provável for a hipótese (independente da nova evidência) e do quão provável é o input dado ela (valendo o mesmo para as hipóteses rivais). O valor dos *priors* é ajustado conforme a experiência, uma vez que a probabilidade a priori de uma hipótese H no instante t1 e a probabilidade de H ser a causa do input recebido em T1 (verossimilhança) podem

¹⁷ Cf. HOHWY, 2013, p. 18.

gerar uma probabilidade a posteriori em t_1 com valor diferente da priori. No instante t_2 , a nova probabilidade a priori será o valor da probabilidade a posteriori resultante em t_1 . Essa dinâmica permite aprender a partir do erro, ajustando os modelos internos de forma probabilística (quanto mais experiência, maior o número da amostra, mais representativo se torna o modelo e menor a geração de erros).

O T-rex illusion¹⁸ é um exemplo de ilusão cujo processamento possui as características de uma implementação bayesiana. Os inputs recebidos são ambíguos (i.e. duas causas possíveis são capazes de os gerar), mas o *percept* não o é. O cérebro precisa decidir qual a hipótese que se afigura a mais provável na presente circunstância – levando em consideração as experiências passadas – para gerar o conteúdo da percepção. A partir de um input inicial estático (i_1t_1), duas hipóteses rivais competem para melhor predizê-lo (ou explicá-lo): a hipótese de que a superfície S está na parte frontal do sólido (h_1) e a hipótese de que a superfície S está na parte traseira do sólido, sendo este vazado (h_2), conforme a figura abaixo:

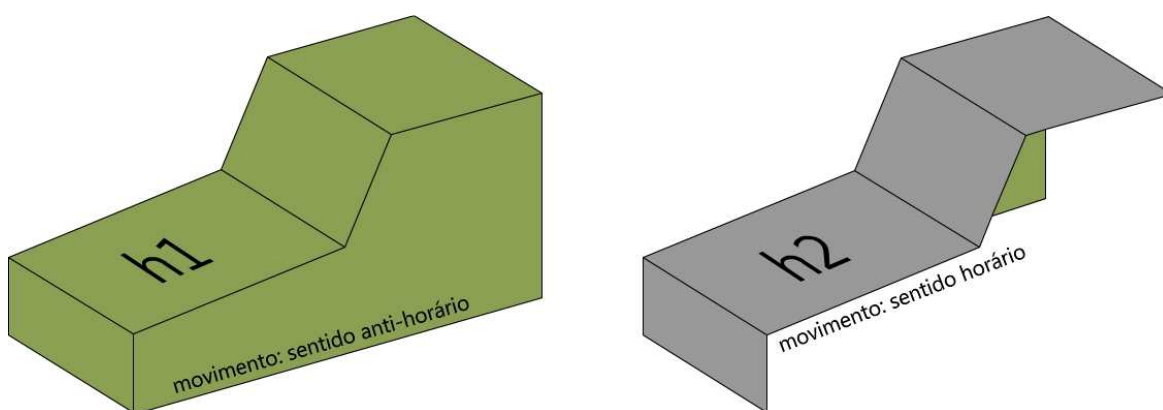


Figura 4 – Hipóteses possíveis acerca da causa dos inputs recebidos.

Uma vez que tanto h_1 como h_2 são causas igualmente possíveis de i_1t_1 , teríamos, caso utilizássemos um método inferencial não-probabilístico, uma indeterminação quanto à hipótese verdadeira, pois i_1t_1 é a afirmação do consequente das seguintes premissas nas quais h_1 e h_2 figuram:

¹⁸ Disponível em <https://www.youtube.com/watch?v=5cJRe9iDDjs>

1. $h_1 \rightarrow i_1t_1$
2. $h_2 \rightarrow i_1t_1$
3. i_1t_1

Com uma inferência bayesiana, no entanto, a probabilidade a posteriori da causa ser h_1 – considerando o estímulo recebido e cada probabilidade a priori e verossimilhança – é superior à probabilidade de ser h_2 :

1. $P(h_1|i_1t_1) = P(i_1t_1|h_1).P(h_1)$
2. $P(h_2|i_1t_1) = P(i_1t_1|h_2).P(h_2)$
3. $P(i_1t_1|h_1) = P(i_1t_1|h_2)$
4. $P(h_1) > P(h_2)$
5. $P(h_1|i_1t_1) > P(h_2|i_1t_1)$

Apesar do valor de verossimilhança (premissa 3) ser igual nas duas hipóteses (pois i_1t_1 é consequência em mesmo grau em ambas), a probabilidade a priori de h_1 é superior à de h_2 (premissa 4), pois sólidos com as características de h_1 são mais frequentes do que superfícies vazadas como em h_2 . Para chegar à conclusão de que a probabilidade a posteriori de h_1 é maior que a de h_2 não é necessário atribuir valores numéricos a cada probabilidade, pois sabendo que $P(h_1)$ é superior a $P(h_2)$ e que $P(i_1t_1|h_1)$ é igual a $P(i_1t_1|h_2)$, deduz-se que $P(i_1t_1|h_1).P(h_1)$ é maior que $P(i_1t_1|h_2).P(h_2)$. O *percept* resultante da implementação de um mecanismo bayesiano será h_1 , pois atingiu a maior probabilidade a posteriori.

Andy Clark afirma que, se o modelo unificador do Predictive Brain estiver correto, “*percepção e ação estão intimamente relacionadas e trabalham conjuntamente para reduzir o erro de predição ao esculpir e selecionar inputs sensoriais*” (CLARK, p. 183, 2013). A ação, no presente paradigma denominada “inferência ativa” (*active inference*), em contraste com a inferência perceptual (*perceptual inference*), não seleciona a hipótese que melhor prediz aquilo que é recebido passivamente, mas testa hipóteses sobre o mundo ao antecipar os inputs que receberia caso elas fossem verdadeiras, criando com isso as condições iniciais desejadas por meio do agir. Assim como uma teoria científica não deve apenas explicar os dados já obtidos (uma vez que teorias rivais também os podem explicar igualmente bem), mas também predizer resultados futuros (tendo como *desideratum*

a realização de um *experimentum crucis* para decidir entre hipóteses), um bom modelo do mundo precisa submeter-se a circunstâncias críticas que o possam corroborar, ajustar ou refutar. No exemplo do T-rex illusion, as hipóteses h1 e h2 podem ser testadas ativamente quando nos movemos, pois elas geram previsões específicas quando à evolução dos inputs de acordo com a posição do sujeito. H1 prediz que o sólido girará no sentido anti-horário; h2 prediz um movimento no sentido oposto. Quando nos movemos ao redor do objeto no sentido anti-horário, tanto h1 quanto h2 predizem inputs idênticos até um determinado ponto (aqui referido como instante t5). No instante seguinte (t6), entretanto, cada hipótese faz uma predição distinta (h1 prediz ixt6 e h2 prediz iyt6), o que consiste em um *experimentum crucis*:

1. $h1 \rightarrow i1t1 \wedge i2t2 \wedge i3t3 \wedge i4t4 \wedge i5t5 \wedge ixt6 \wedge \neg iyt6$
2. $h2 \rightarrow i1t1 \wedge i2t2 \wedge i3t3 \wedge i4t4 \wedge i5t5 \wedge iyt6 \wedge \neg ixt6$

Apesar de h1 continuar recebendo uma maior probabilidade a posteriori até t5, persistindo como a hipótese vencedora que determina a fenomenologia, no instante t6 faz-se visível a parte traseira do objeto, revelando se tratar de uma superfície vazada conforme h2. Embora a sequência de inputs do instante t1 a t5 ter sido ambígua (podendo ter sido causada tanto por h1 como por h2), a sequência completa, que inclui o input iyt6, desfaz a ambiguidade a favor de h2, visto que o input que é recebido agora é incompatível com h1:

1. $P(h1|iyt6) = P(iyt6|h1).P(h1)$
2. $P(h2|iyt6) = P(iyt6|h2).P(h2)$
3. $P(iyt6|h1) = 0$
4. $P(iyt6|h2) = 1$
5. $P(h1) > P(h2)$
6. $P(h1|iyt6) < P(h2|iyt6)$

Não obstante os inputs recebidos do instante t1 a t5 estarem em conformidade com as duas hipóteses, as percepções do sentido do movimento são opostas. Ao considerar o T-rex como uma instância de uma regularidade de observação frequente no mundo (h1), o cérebro bayesiano produz a ilusão de que a cabeça está se movimentando em direção ao observador. Isso evidencia que a percepção envolve

um processo de escolha de hipóteses, no qual o que é percebido é a hipótese que, dado um modelo do mundo, melhor prediz estímulos sensoriais. Espera-se que a ilusão, caso seja adaptada a um experimento com eletroencefalografia, gere respostas corticais maiores no instante t_6 quando o input recebido não for o esperado pela hipótese dominante (h_1), mas o previsto pela hipótese rival (h_2) – o que evidenciaria a escolha do cérebro por h_1 e a sua utilização para antecipar a evolução do cenário ambíguo. Uma maneira de verificar se o cérebro está a considerar a hipótese h_2 mesmo enquanto seleciona h_1 como a melhor explicação (mantendo h_2 como segunda opção em vez de descartá-la completamente) é apresentando um input z em t_6 que viole tanto a expectativa gerada por h_1 (x) como por h_2 (y). Se o modelo defendido estiver correto, espera-se sinais de erro mais fortes (medidos por respostas eletrofisiológicas maiores) em t_6 após o input z (que não é previsto por nenhuma hipótese) do que após o input y (que é previsto pela hipótese h_2), e respostas maiores em y do que em x (previsto por h_1 , a hipótese mais provável).

2.3 ESCOLHA DE FRAMEWORK

Tendo em vista que a presente pesquisa faz uso de eletroencefalografia e, mais especificamente, da técnica de Event-Related Potential's, é preciso levar em consideração qual modelo permite uma melhor interpretação dos dados eletrofisiológicos obtidos. Como os componentes eletrofisiológicos de interesse vêm sendo recentemente interpretados como sinalizando o quanto um estímulo desvia de sua predição, dependendo muitas vezes de informações presentes em níveis superiores (e.g. sinal correlacionado ao processamento de formas sendo modulado por informações sintáticas, léxicas e semânticas), o paradigma do Predictive Brain parece oferecer uma estrutura explicativa mais adequada que a da teoria modular focada na reação passiva a estímulos. Entre as principais razões para a escolha do Predictive Brain podemos citar:

- 1) A possibilidade de correlacionar o grau de probabilidade de um estímulo, dado conhecimentos prévios, com a amplitude do sinal (utilizando inferência bayesiana)
- 2) O crescente número de evidências de que os sinais corticais estão correlacionados com erros de predição

- 3) Evidências de que alguns processos perceptuais apresentam penetrabilidade cognitiva, contrariando o que é assumido pela posição modularista

O Predictive Brain possibilita a formulação de hipóteses específicas sofisticadas, estado ancoradas em um framework teórico mais geral (sendo este, por sua vez, constrangido por princípios físicos mais fundamentais). Modelos probabilísticos permitem não apenas explicações teóricas de processos mentais, como também geram predições observáveis associadas aos sinais de erro medidos pelos instrumentos de aferição da atividade cerebral. Conjectura-se que enquanto frases são lidas, representações de estados de coisas e eventos são mentalmente modeladas de modo gradual e preditivo, e que os estímulos vindouros são antecipados com base naquilo que se afigura “o mais provável” em termos epistêmicos, semânticos, sintáticos, léxicos e perceptuais. Ao lermos a sentença incompleta “O gato está fugindo do...”, por exemplo, nosso cérebro procurará o tipo de entidade que terá a maior probabilidade de estar na relação “x está fugindo de y” tendo “x” como uma instância de gato (tendo como alvo de consulta as representações de regularidades produzidas por experiências passadas). Não necessitamos atribuir valores numéricos à palavra “cachorro” ao considerarmos a mesma como a melhor candidata a “y” entre as demais possibilidades. A palavra “cachorro” será mais provável que “elefante”; esta, por sua vez, será mais provável que “elétron” ([P(“cachorro”|”O gato está fugindo do...”)]> [P(“elefante”|”O gato está fugindo do...”)]> [P(“elétron”|”O gato está fugindo do...”)]). O modelo probabilístico, caso correto, produzirá sinais cerebrais mais salientes após a apresentação do estímulo “elétron” do que após “elefante”, assim como maior em “elefante” do que em “cachorro”. Deste modo constrói-se uma ligação entre modelos de processamento léxico, semântico, sintático e epistêmico com padrões de atividade cerebral, sendo que os dados empíricos permitem-nos reformular as teorias de modo a melhor predizê-los.

3 SOBRE A CAPACIDADE DE ATRIBUIR ESTADOS MENTAIS REPRESENTACIONAIS

Das entidades que encontramos no mundo com as quais a interação é imprescindível, outros seres humanos despontam como aquelas que em maior grau se esquivam de serem subsumidas por leis gerais e padrões precisos. Não obstante a suposta complexidade daquilo que se passa em cérebros alheios e que está causalmente relacionado com as ações que observamos, nossa cognição parece estar munida, pré-experencialmente, das ferramentas que nos permitem antecipar comportamentos satisfatoriamente com base na atribuição de entidades mentais teoréticas. “Teoria da Mente” (ToM) é o que se convencionou chamar a capacidade de atribuir estados mentais representacionais a outras pessoas, tais como crenças, intenções e desejos, e de explicar/predizer comportamentos levando em consideração o modo como estes estados mentais tipicamente interagem com estímulos e comportamentos (i.e. seus papéis causais). É comum estas regularidades serem expressas na forma de leis “*Ceteris Paribus*” não-estritas, como “geralmente, se alguém acredita que a ação p produz a consequência q e deseja que q se torne o caso, então, na ausência de meios alternativos a p e desejos incompatíveis com q , esta pessoa faz a ação p ”. De modo semelhante a uma teoria científica, as entidades teoréticas postuladas pela ToM (i.e. estados mentais inobserváveis) e as regularidades expressas por suas leis permitiriam ao “*mindreader*” inferir estados mentais, predizer e explicar comportamentos, como se estes fossem “deduzidos” de premissas contendo leis gerais e condições iniciais.

3.1 FALSE BELIEF TASK VERBAL

O False Belief Task, criado por Wimmer e Perner (1983), é um teste originalmente concebido para examinar a habilidade de uma pessoa em atribuir estados mentais que envolvam uma representação da realidade. O teste consiste em apresentar uma sequência de ações cujo estímulo permite a atribuição de uma crença em conflito com a realidade. Fora assumido por muitos pesquisadores que seria necessário para passar no teste levar em consideração a crença falsa de um dos personagens, requerendo igualmente uma apreciação de seu papel causal. De acordo

com esta visão, seria correto inferir que se um ser humano ou animal passar no teste, ele possui uma Teoria da Mente.

O False Belief Task verbal exposto na Figura 5 apresenta um cenário com duas personagens, Sally (de vermelho) e Anne (de azul), duas caixas (L e R) e um objeto (uma bola) de posse de Sally (1). Sally coloca sua bola na caixa L (2) e em seguida sai da sala (3). Anne, na ausência de Sally, caminha em direção à caixa L, pega a bola e a transfere para a caixa R, caminhando em seguida ao local em que estivera previamente (4-7). Sally retorna e encontra Anne no mesmo local em que estava quando Sally havia saído (8). O teste informa que Sally retornou à sala para pegar sua bola, e em seguida pergunta em direção a qual local (L ou R) será o próximo movimento de Sally. A resposta correta será “em direção à caixa L”, pois Sally não viu a bola ser trocada de lugar, possuindo a crença falsa de que a bola permanece em seu local original. Crianças neurotípicas passam no FBT tradicional por volta dos 4 anos (WELLMAN et al., 2001), fato que supostamente indicaria a apreciação de crenças falsas e seus respectivos papéis causais, seguindo-se disto, a posse de uma ToM.

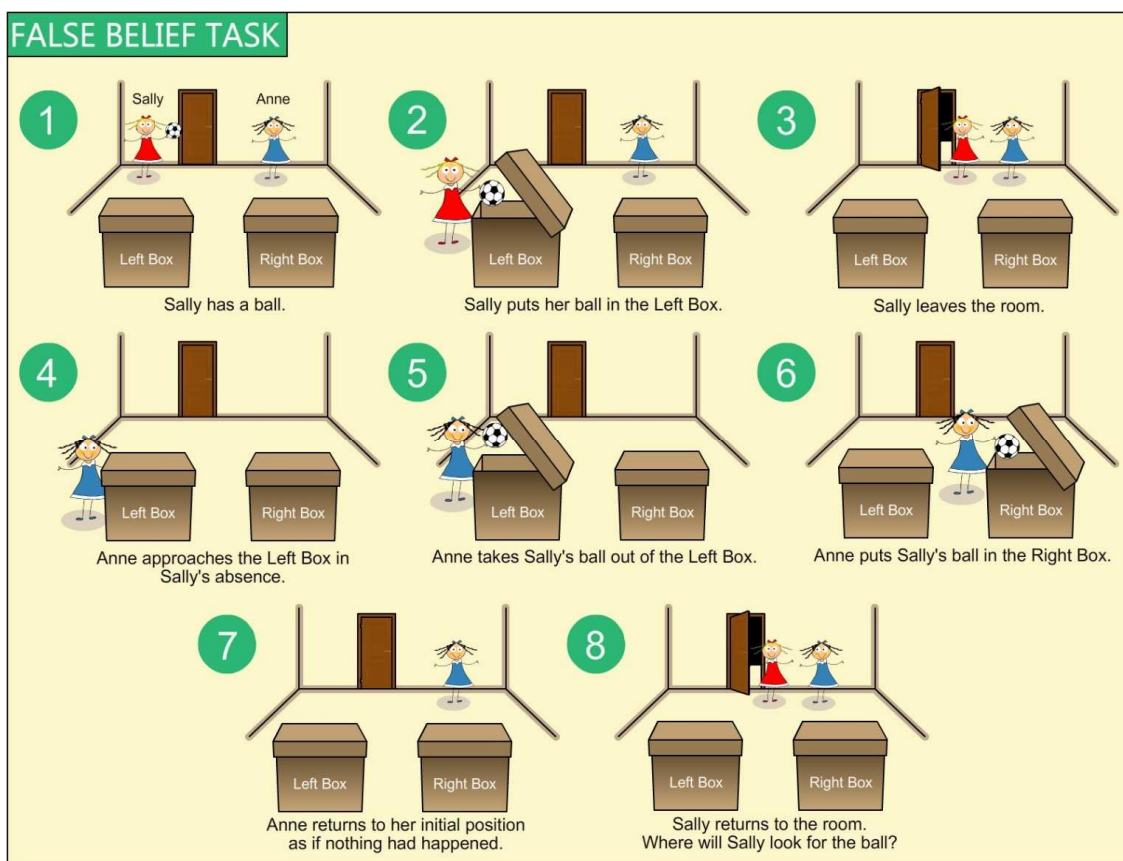


Figura 5 – Sequência de inputs do False Belief Task verbal.

3.2 O FALSE BELIEF TASK NÃO-VERBAL

Clements e Perner (1994) desenharam uma versão não-verbal de FBT que testaria a Teoria da Mente implícita das crianças. Não obstante o fato de crianças com 3 anos tipicamente errarem ao responder verbalmente onde o personagem procurará seu objeto, elas antecipam sua ação olhando para o local em direção ao qual elas acreditam que o mesmo irá se deslocar, sem precisarem ser questionadas. Rebecca Saxe (2012) detalha quais são as premissas assumidas nestes testes não-verbais:

Antes que crianças possam falar ou seguir instruções, seus movimentos oculares já revelam suas expectativas acerca do mundo. Crianças olham por mais tempo para objetos ou eventos que as surpreendem e irão apresentar olhar antecipatório em direção ao local em que esperam que algo interessante ocorra (SAXE, 2012, p. 108).

Clements e Perner verificaram que o olhar antecipatório das crianças com 3 anos era direcionado à caixa vazia – local em que Sally acredita estar o objeto que procura. Ruffman et al. (2001) também constataram que as mesmas crianças cujo olhar antecipatório lhes permite passar no FBT não-verbal “teimam” na resposta incorreta quando explicitamente indagadas no FBT verbal. Este resultado sugeriria, de acordo com os autores, que as crianças de 3 anos não teriam acesso consciente ao conhecimento que guiaria seus olhares (RUFFMAN et al., 2001, p. 201). Haveria uma dissociação entre a computação inconsciente que guiaria o olhar antecipatório das crianças e o raciocínio explícito requerido para responder verbalmente. Recentemente, o uso de novos métodos e equipamentos permitiu reduzir ainda mais a idade das crianças que apresentam sucesso preditivo no FBT não-verbal. Southgate et al. (2007), utilizando um aparelho de eye tracking, conseguiram verificar que crianças de 2 anos já apresentam olhar antecipatório compatível com a atribuição de uma crença falsa. Onishi e Baillargeon (2005), utilizando o método da “violação da expectativa”, também encontraram evidências que, segundo eles, seriam favoráveis à hipótese de que crianças de 15 meses já possuíam uma Teoria da Mente representacional (ONISHI & BAILLARGEON, 2005, p. 257). Assumindo-se que crianças olham por mais tempo para eventos que contrariam suas expectativas, as mesmas, caso já possuam uma ToM, apresentarão olhar prolongado apenas quando

o personagem procurar seu objeto no exato local em que ele se encontrar, olhando relativamente menos quando ele procurar onde ele erroneamente acredita que ele se encontra (pois se as crianças antecipam corretamente a ação, elas se espantarão apenas quando o comportamento contrariar suas expectativas, que será a procura no local onde o objeto de fato estiver).

Ainda mais recente e surpreendente é o experimento de Southgate e Verneti (2014) que, adaptando eletroencefalografia e eye-tracking a uma versão modificada de False Belief Task (no qual há uma única caixa), obtiveram evidências de que crianças com apenas 6 meses seriam capazes de antecipar comportamentos com base em crenças falsas (ao menos de um modo inconsciente). Tomando como premissa que a supressão sensório-motor da frequência alfa (sensorimotor alpha suppression) indica predição de ações, a antecipação de comportamentos na versão modificada de FBT produziria esta atividade, não se verificando supressão da frequência alfa quando o realizador do FBT não possui nenhuma expectativa relativa ao comportamento da personagem. Southgate e Verneti verificaram que, no experimento com crianças de 6 meses, as mesmas antecipavam ações/inações em função da crença falsa da personagem (i.e. por meio da consideração de um estado de coisas não-factual tal como representado pela personagem), e não em função do estado de coisas factual acerca do qual as crianças estavam cientes. Estes resultados não indicariam apenas que crianças de 6 meses seriam capazes de levar em consideração os estados mentais representacionais alheios (até mesmo nos casos em que crença e mundo estão em conflito, como em crenças falsas), mas que elas, em acréscimo a esta capacidade precoce, já estariam cientes das consequências comportamentais típicas de crenças e das relações destas com estímulos do ambiente e outros estados mentais.

Senju et al. (2009), aplicando o mesmo teste de Southgate et al. (2007) em adultos com síndrome de Asperger com QI elevado (média de 115.6), encontraram evidências de que esse grupo, apesar de apresentar bom desempenho no FBT explícito, fracassa no FBT implícito. Frith (2004) considera que o sucesso preditivo de indivíduos com síndrome de Asperger em FBT's explícitos deve-se ao fato de sua inteligência permitir-lhes utilizar inferências lógicas:

Por exemplo, eles [os indivíduos com síndrome de Asperger] podem apresentar sucesso ao entender cenários complexos, como rastrear

onde Mary acredita que John acha que ele pode comprar um casaco. Entretanto, apresentar uma teoria da mente explícita não necessariamente implica uma habilidade intuitiva de leitura de mentes (FRITH, U., 2004, p. 678).

A figura 6 resume os resultados dos experimentos referidos (simplificados na forma dicotômica “sucesso ou fracasso”):

	FBT Implícito	FBT Tradicional
(Southgate & Vernetti, 2014)	6 meses	6 meses
	✓	✗ (*)
(Onishi & Baillargeon, 2005)	15 meses	15 meses
	✓	✗ (*)
(Wimmer & Perner, 1983)	3 anos	3 anos
	✓ Teimosia (Perner, 1994)	✗
	5 anos	5 anos
	✓ (*)	✓
(Baron-Cohen, Leslie & Frith, 1985)	10 anos - Síndrome de Down (mean QI = 64)	10 anos - Síndrome de Down (mean QI = 64)
	✓ (*)	✓
	11 anos - Autistas (mean QI = 82)	11 anos - Autistas (mean QI = 82)
	✗ (*)	✗
(Senju, Southgate, 2012)	Adultos com Síndrome de Asperger (mean QI = 115)	Adultos com Síndrome de Asperger (mean QI = 115)
	✗	✓

Figura 6 – Resumo dos resultados dos principais experimentos que utilizaram versões do FBT. Os resultados marcados com asterisco não correspondem a nenhum experimento realizado, consistindo apenas em estimativas inferidas dos demais experimentos.

Como crianças de 6 meses ao mesmo tempo já possuiriam conceitos tão abstratos como crenças e desejos e estariam cientes de seus papéis causais se elas carecem dos recursos cognitivos necessários para adquiri-los explicitamente e das experiências necessárias para que tal aprendizagem se efetive? E como adultos com síndrome de Asperger, cuja flexível cognição lhes permite destaque em variadas áreas do conhecimento (e.g. ganhar a Medalha Fields, prêmio máximo da Matemática), fracassam nos mesmos testes em que crianças com apenas 6 meses conseguem passar? Os resultados obtidos por meio do uso de diversas versões de

False Belief Tasks constituem forte evidência de que a capacidade conjunta de atribuir estados intencionais e considerar o modo como estas entidades interagem com estímulos e comportamentos possui uma base inata de domínio específico passível de dano seletivo. Entretanto, do fato de seres humanos neurotípicos possuírem um sistema cognitivo destinado à atribuição de estados mentais possuindo forte semelhança com uma teoria científica — devido ao fato de postular entidades teóricas inobserváveis e conectá-las umas com as outras e com inputs e outputs observáveis por meio de leis — não se segue que a teoria utilizada seja verdadeira, ainda que permita grande sucesso preditivo. As crenças e desejos atribuídos pela ToM podem ser meras ficções úteis das quais nossa espécie faz uso para interagir satisfatoriamente com entidades tão imprevisíveis como outros seres humanos, sem qualquer correspondência com o que se passa realmente em cérebros alheios. É necessário que se faça a distinção entre as entidades mentais de ontologia duvidosa que comumente atribuímos às outras pessoas e julgamos como a razão de seus comportamentos e aquilo que de fato há em seus cérebros e que está causalmente relacionado com suas ações — tema do capítulo seguinte.

4 SOBRE A ONTOLOGIA DE ESTADOS MENTAIS REPRESENTACIONAIS

Embora as pesquisas empíricas em Teoria da Mente tenham sido dirigidas principalmente por não-filósofos, o arcabouço teórico desenvolvido por filósofos da mente e da linguagem permanece sendo utilizado como o modelo normativo subjacente à capacidade de *mindreading*¹⁹. Um modelo muito utilizado é a concepção de entidades mentais como estados possuidores de intencionalidade, usualmente formalizados em atitudes proposicionais. *Intencionalidade* é a propriedade de certos estados mentais de serem direcionados ou “acerca de” objetos e estados de coisas do mundo (cf. SEARLE, 1983, p. 1). Um estado como uma crença, uma intenção ou desejo envolve uma relação específica entre um sujeito e um estado de coisas ou objeto (e.g. na atitude proposicional “Ricardo acredita que Paris é a capital da França”, “Ricardo” é o sujeito, “acredita” é a relação ou modo psicológico e “Paris é a capital da França” o estado de coisas tomado por Ricardo como correspondendo ao mundo). Atitudes proposicionais também apresentam opacidade referencial, não permitindo generalizações existenciais ou substituição de termos correferenciais *salva veritate*²⁰: de “Ricardo acredita que Belerofonte tinha um cavalo alado” não se segue que “Belerofonte tinha um cavalo alado”, assim como não se segue de “Ricardo acredita que Aristóteles era um filósofo” a veracidade da proposição “Ricardo acredita que o preceptor de Alexandre era um filósofo”, ainda que o nome próprio “Aristóteles” e a descrição definida “o preceptor de Alexandre” tenham um mesmo referente.

A concepção de estados mentais representacionais como atitudes subjetivas específicas em relação a estados de coisas ou objetos representados, entretanto, é problemática por restringir-se a um nível tão-somente psicológico, mantendo indefinida sua correspondência. Apesar de ser relativamente simples explicar e prever comportamentos por meio da atribuição de atitudes proposicionais e leis que as conectam com estímulos e comportamentos, não se pode verificar objetivamente em cérebros as entidades postuladas pela nossa linguagem ordinária, muito menos utilizá-las — assim concebidas — em teorias científicas. Uma primeira tentativa de analisar o significado exato dos termos mentalistas de maneira científica — voltada ao observável e objetivo — foi o behaviorismo analítico proposto por Gilbert Ryle

¹⁹ Cf. APPERLY, 2010, p. 5.

²⁰ Cf. CRANE, 2001, p. 11.

(2000), que argumentou que se o uso de termos e predicados mentais ordinários como crenças, competências e traços característicos pessoais são supostos como possuidores de significados, não poderiam ser estes eventos inobserváveis, muito menos propriedades possuídas por uma substância imaterial cartesiana (mito que Ryle pretende eliminar). Alternativamente a isto, o significado das sentenças que invocam tais entidades, após a devida análise, traduzir-se-iam, sem nenhum resquício, a dados observáveis (ou potencialmente observáveis) de comportamentos ou disposições para se comportar sob várias circunstâncias de inputs. Acreditar em uma proposição “P” nada mais é que possuir uma ou várias disposições comportamentais tocantes a “P”. Achar que termos mentais designam outra coisa senão comportamentos e disposições seria um erro categorial de supor a mente uma coisa/substância à parte, resultado de confusões e do uso abusivo da linguagem ordinária. Por serem circunstâncias de estímulos e comportamentos eventos objetivos e observáveis, uma linguagem que reduz estados mentais a disposições input-output, ainda que não faça menção a nenhuma entidade misteriosa intermediária, já consegue reduzir o que era à primeira vista subjetivo e privado a algo mensurável e, conseqüentemente, cientificamente tratável.

David Lewis, em sua teoria exposta nos artigos *Argument for the Identity Theory* (1966), *Psychophysical and Theoretical Identifications* (1972), *How to Define Theoretical Terms* (1979) e *Mad pain and Martian pain* (1980), propõe, em acréscimo aos padrões *input-output* behavioristas, a consideração de entidades intermediárias na definição do mental. Lewis é um funcionalista na medida em que considera que o que é essencial para que algo seja caracterizado como um estado mental é a função que ele realiza. O conceito de um estado mental, portanto, assemelhar-se-ia aos conceitos de artefatos, uma vez que para algo ser uma instância de um tipo de artefato bastaria realizar a função que o define, prescindindo, deste modo, de um substrato físico específico. De modo semelhante a David Armstrong e Hilary Putnam — outros defensores do funcionalismo —, Lewis propõe que consideremos um estado mental como algo que é causado por certos inputs sensoriais e que, pela interação com outros estados mentais, causa certos outputs comportamentais. O papel causal do estado mental da dor, por exemplo, seria o de tipicamente ser causado por inputs cuja função é detectar danos ao corpo e, por sua vez, causar a emissão de ruídos altos e o comportamento de repelir o objeto nocivo. Porém, é necessário que os conceitos funcionalistas especifiquem uma classe finita de causas típicas e efeitos típicos em

várias circunstâncias, podendo ser formalizado através de disjunções de inputs e outputs que tipicamente relacionam-se ao estado mental que se intenciona definir.

David Lewis considera a folk psychology (isto é, o modo como nos referimos a estados mentais alheios e através destes explicamos e predizemos o comportamento) uma teoria empírica que pretende explicar a regularidade entre estímulos e respostas encontradas no comportamento humano. Termos mentais – definidos pela função que realizam – seriam termos teóricos (*T-terms*) introduzidos pela teoria em questão e que, não obstante o fato de serem inobserváveis, adquirem significado por meio de relações que possuem com outros estados mentais, estímulos sensoriais e comportamentos (estes dois últimos termos observáveis, ou *O-terms*). Entidades teóricas são definidas como “os ocupantes dos papéis causais especificados pela teoria, como as entidades, quaisquer que sejam, que possuem certas relações causais umas com as outras e com os referentes dos *O-terms*” (LEWIS, 1972, p. 211). Lewis pretende com isso demonstrar que, se a teoria psicológica for verdadeira, o que for que existir que tenha relações causais com os *O-terms* de fato existe, mesmo que ainda não sejamos capazes de identificar com exatidão as entidades que “realizam a função” em nossos cérebros (pois só possuiríamos descrições definidas/definições funcionais delas). Deste modo, Lewis pretende “fixar” a referência dos termos mentais, tornando seus significados parasitários dos significados dos termos observáveis. Podemos concluir deste argumento que, se a teoria funcionalista for verdadeira, termos mentais são como descrições definidas que possuem um referente (como “o homem mais alto do mundo”), ainda que permaneça incógnito para o usuário da expressão de qual entidade se trata. Se a teoria for falsa, termos mentais são descrições definidas que nada referem (como Pégasus, “o cavalo alado de Belerofonte”). De acordo com Ned Block (1982), os termos mentais para o funcionalista são totalmente caracterizados em termos que envolvem apenas uma linguagem lógico-matemática e termos para sinais de inputs e outputs comportamentais: “Portanto, o funcionalismo satisfaz um dos ‘desiderata’ do behaviorismo, ao caracterizar o mental numa linguagem inteiramente não-mental” (BLOCK, 1982, p. 522). Em comparação com o behaviorismo, o funcionalismo aceitaria em sua definição de estado mental as entidades inobserváveis que o behaviorista pretendia eliminar de seu discurso, desde que o inobservável figure na proposição como uma expressão quantificada constituída inteiramente por objetos observáveis: “Se eu estiver certo, *T-*

terms são elimináveis – nós sempre podemos substituí-los pelo seus definientia” (LEWIS, 2002, p. 92).

Outra importante contribuição de David Lewis para a discussão envolvendo a ontologia dos termos mentais é a sua tentativa de conciliar o funcionalismo com a teoria da identidade tipo-tipo. Os teóricos da identidade tipo-tipo (e.g. PLACE, 1956; SMART, 1959 e FEIGL, 1958) afirmam que estados mentais são numericamente idênticos a estados cerebrais, ou seja, não são outra coisa senão estados cerebrais (o exemplo mais famoso é a afirmação “dor = fibras-C disparando”). Já a teoria híbrida que o Lewis propõe não abandona a visão funcionalista de que o que define um estado mental é a sua relação com inputs, outputs e outros estados mentais – daí se seguindo a possibilidade de múltiplas instanciações. Substratos físicos vários poderiam realizar em princípio o papel causal que a “dor funcionalista” desempenha em cérebros humanos e animais, falseando a tese da identidade (um marciano poderia ter um estado hidráulico funcionalmente equivalente ao papel causal da dor humana, que é realizado por fibras-C). Entretanto, Lewis demonstra que a proposição defendida pelo teórico da identidade pode ser concluída de um argumento que se baseia na definição funcionalista e em um fato contingente do mundo. O argumento é formalizado do seguinte modo²¹:

Premissa 1: Estado mental M = o ocupante do papel causal R (pela definição de M)

Premissa 2: Estado neural N = o ocupante do papel causal R (pela teoria fisiológica)

Logo, o estado mental M = estado neural N (pela transitividade de =)

Este argumento quer demonstrar que, embora possa haver (logicamente) uma espécie ou sistema que possua um estado não-neural como ocupante do papel causal R, é um fato contingente do mundo em que vivemos que em humanos e em outros animais os ocupantes sejam estados neurais. O ocupante de um papel causal R pode variar de espécie para espécie, como é o caso da ausência de fibras-C em certos animais (como cachorros), que não os torna devido a isto insensíveis à dor. Assim como um estado mental qualitativo como a dor pode ser definido e referido pelo modo como interagem com inputs e outputs, um estado mental representacional como “a crença de que irá chover” pode ser definido como “o estado, qualquer que seja, que é

²¹ Cf. LEWIS, 2002, p. 88.

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 entidade mental por meio de apenas entidades observáveis que, ao defini-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 (i.e. as demais entidades mentais que fazem parte da definição também serão definidas pelas relações com eventos observáveis).

Paul Churchland (1981) opõe-se às tentativas behavioristas e funcionalistas de traduzir a linguagem mentalista de primeira pessoa a um vocabulário objetivo de terceira pessoa. Crenças, desejos e intenções constituiriam, para Churchland, "entidades fantasmagóricas", resíduos dualistas a serem eliminados pelas explicações das neurociências com o avanço destas. O materialismo eliminativista (*eliminative materialism*) que defende considera que as teorias acerca da mente adotadas pelo senso comum consistem em explicações radicalmente inadequadas de nossas atividades cerebrais, "*defeituosas demais para sobreviver à redução interteórica*" (CHURCHLAND, 1981, p. 72), devendo, portanto, ser simplesmente eliminadas (e não suavemente reduzidas) por uma neurociência completa, do mesmo modo como o desenvolvimento da ciência eliminou entidades que nada designavam como "o flogisto" e "os espíritos vitais". Não apenas a ontologia das entidades teóricas postuladas pela Folk Psychology é vista como problemática pelo materialismo eliminativista, como também as "leis" que as unem para permitir explicações e predições de comportamentos. De acordo com Churchland, "*os condicionais reconstruídos que pretendem ser as 'leis' da folk psychology são estritamente falsos, ou limitam-se ao trivial por serem qualificados por várias cláusulas ceteris paribus*" (CHURCHLAND, 1989, p. 230). Não existindo leis estritas conectando estímulos, estados mentais e comportamentos — conforme a famosa argumentação de Donald Davidson²² —, resta ao defensor da folk psychology utilizar regularidades que se verificam na maioria dos casos, que permitem mais acertos do que erros preditivos no uso cotidiano, mas que estão longe de serem consideradas leis sérias pela ciência.

²² DAVIDSON, 2001.

Ao atacar a folk psychology pela sua suposta falsidade enquanto teoria, o eliminativista considera inexistentes suas entidades teóricas e fracas as generalizações que as subsumem. De fato, parece implausível que crenças venham a se revelar, com o advento de uma neurobiologia completa, “atitudes proposicionais instanciadas numa linguagem do pensamento”. Mas uma noção mais geral de crença, como “algum tipo de informação contida em cérebros que é tomada pelo organismo como correspondendo ao mundo externo”, não deverá se distanciar muito daquilo que descobriremos futuramente como “as causas reais do comportamento”. Se o funcionalista estiver certo, há algo que ocupa um determinado papel causal e, embora não saibamos identificar atualmente qual substrato material realiza crenças em nossos cérebros, no momento em que o desvelarmos poderemos denominá-lo “crença”, bastando que possua as propriedades relacionais por meio das quais nos referíamos a ele por descrição. Uma forma de procurar o correlato neural de uma crença, caso ele realmente existir, é buscando padrões de atividade cerebral correlacionados com a posse ou recrutamento de crenças específicas, comparando a resposta cerebral de indivíduos que possuam uma crença (ou conjunto de crenças) com a de indivíduos que não a possuam. Se crentes e não-crentes realizam uma mesma tarefa cognitiva e — todo o resto sendo constante — apresentam sinais cerebrais significativamente distintos, então devemos suspeitar que o que está a causar a diferença é “aquele x que ocupa um determinado papel causal”. No capítulo seguinte é apresentada uma tentativa de se chegar aos correlatos neurais de crenças por meio de experimentos com eletroencefalografia.

5 ACESSANDO INFORMAÇÕES SOBRE CRENÇAS POR MEIO DE RESPOSTAS ELETROFISIOLÓGICAS

5.1 PROJETO DO EXPERIMENTO

Kutas & 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 (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, mais de mil artigos foram escritos utilizando 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 (cf. KUTAS & FEDERMEIER, 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 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”. A latência do componente elicitado pelas frases falsas medida pelos pesquisadores 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 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).

Partindo destes achados empíricos, conjectura-se que se o conhecimento de um fato produz um N400 em um indivíduo que o possui e que lê algo que frustra sua

expectativa semântica relativa a este mesmo conhecimento, 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. 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 este fato, 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/salgada”, uma vez que as propriedades “ser homem” e “ser salgado” não são predicáveis de entidades como casas.

Quanto maior a probabilidade subjetiva de uma determinada palavra P aparecer em uma sentença incompleta (*cloze probability*), maior a negatividade do N400 quando, ao invés de P, outra palavra aparecer (negatividade 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”²³). Levando isso em consideração, no projeto aqui exposto, procurou-se selecionar sentenças relativas à filosofia que provoquem no indivíduo que as conheça (e.g. um filósofo) uma expectativa muito forte quanto à palavra final. As sentenças escolhidas não apenas têm de constituir um “denominador comum” que idealmente inclua somente aquilo que todos os participantes da categoria dos filósofos conheçam, mas também devem aparecer de tal forma que apenas uma única palavra esperada ao final da sentença seja capaz de tornar a sentença verdadeira. “Teeteto é um diálogo escrito por Platão”

²³ Cf. FEDERMEIER & KUTAS, 1999.

é uma sentença deste tipo, pois apenas a palavra “Platão” torna a sentença verdadeira, e apenas “Platão” será esperada por aqueles que deste fato estão cientes. O modo que é aqui utilizado para formalizar e selecionar exemplares deste tipo de sentença consiste em “quebrar” a frase em duas partes, a primeira delas sendo a sentença incompleta (A) e a segunda o nome que aparece em seu final (B), que completa a sentença, resultando em “A...B”. Para expressar que nenhum outro nome além de B torna a sentença incompleta verdadeira, adicionou-se uma cláusula de unicidade: $A...B \wedge \neg(\exists\theta) A...\theta \wedge \neg(\theta=B)$ ²⁴. A Figura 7 detalha a estrutura da formalização:

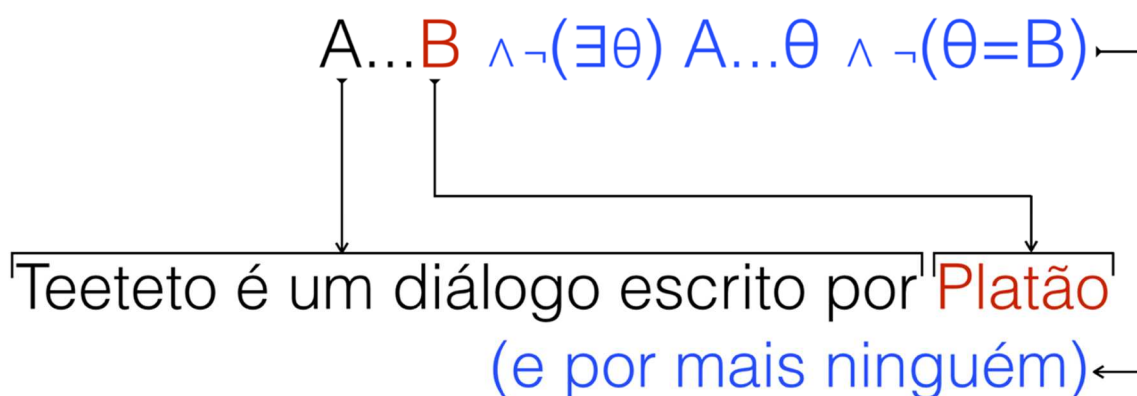


Figura 7 - Exemplo de sentença do tipo $A...B \wedge \neg(\exists\theta) A...\theta \wedge \neg(\theta=B)$

Já “Platão escreveu o diálogo Teeteto”, apesar de ser acerca do mesmo estado de coisas da sentença anterior, não constrange consideravelmente a expectativa do filósofo, uma vez que muitas outras palavras além de “Teeteto”, aparecendo no lugar desta, tornam a sentença verdadeira, visto que Platão escreveu vários diálogos. Deste modo, não é uma sentença do tipo $A...B \wedge \neg(\exists\theta) A...\theta \wedge \neg(\theta=B)$, mas do tipo $A...B \wedge (\exists\theta) A...\theta \wedge \neg(\theta=B)$:

²⁴ 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.

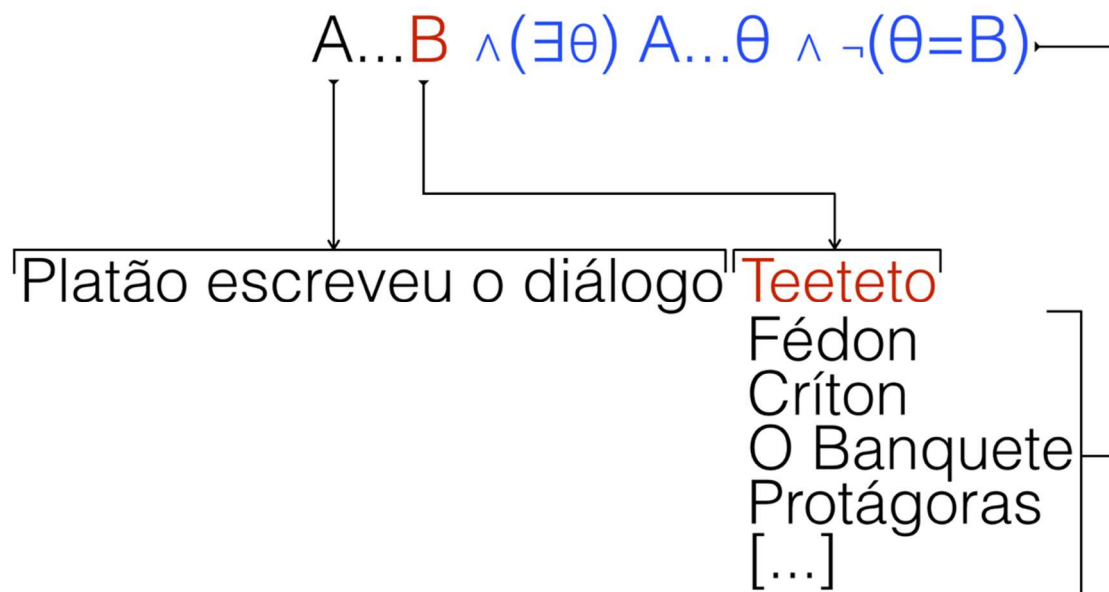


Figura 8 – Exemplo de sentença do tipo $A...B \wedge (\exists \theta) A... \theta \wedge \neg(\theta=B)$

As 80 sentenças relativas à filosofia que serão apresentadas em um monitor, palavra por palavra, aos participantes, enquanto estes estiverem com eletrodos em suas cabeças, são sentenças do tipo $A...B \wedge \neg(\exists \theta) A... \theta \wedge \neg(\theta=B)$ que praticamente todo filósofo sabe serem verdadeiras. Entretanto, metade delas acabará 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 $A...B \wedge \neg(\exists \theta) A... \theta \wedge \neg(\theta=B)$ – o conjunto Q:

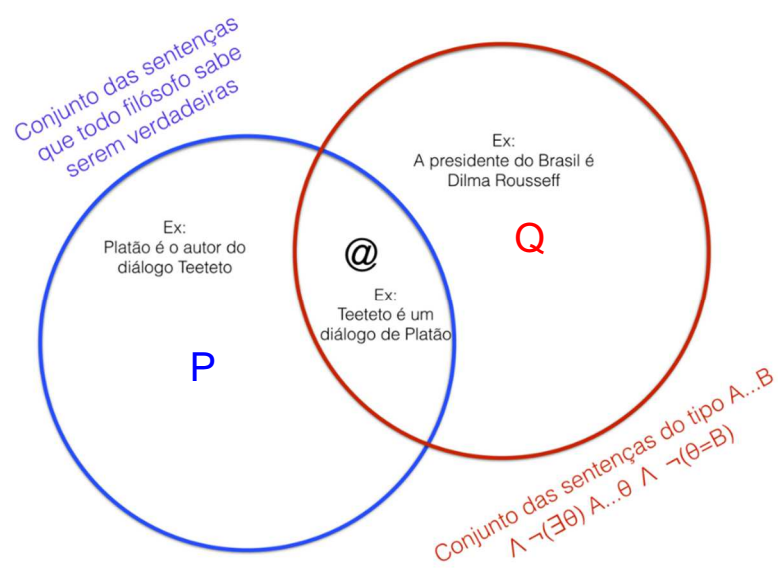


Figura 9 - Conjunto @ ($P \cap Q$): o conjunto das sentenças do tipo $A...B \wedge \neg(\exists\theta) A... \theta \wedge \neg(\theta=B)$ que todo filósofo sabe serem verdadeiras

O experimento pode considerar dois grupos dicotômicos, filósofos e não-filósofos, que idealmente conheceriam e ignorariam todos os fatos descritos pelas sentenças do conjunto @, respectivamente. Outra alternativa seria comparar os sinais dos ERP's com o desempenho individual dos participantes em um questionário escrito que os avaliaria quanto ao grau de conhecimento de fatos filosóficos. Deste modo 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 que seria um resultado esperado com participantes com um conhecimento parcial do conjunto @:

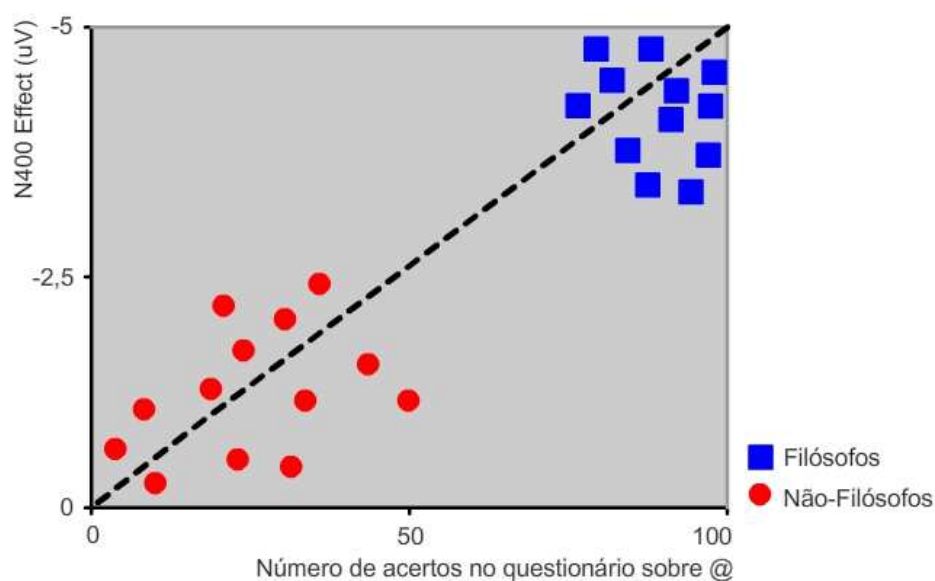


Figura 10 – Correlação esperada entre medição de conhecimento filosófico em questionário escrito e N400 Effects.

Prediz-se esses resultados por se supor que apenas quem tiver 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 filósofo, de altíssima precisão. É possível que uma antecipação deste grau não se limite à expectativa semântica, mas que esta também influencie a expectativas léxica: não apenas penso em Platão, o filósofo que escreveu o referido diálogo, mas espero a palavra “Platão” escrita na tela, com determinadas letras e comprimento, e nenhuma outra coisa exceto isso. Um tão grande sucesso preditivo elicitaria, no filósofo, uma resposta bastante atenuada ao ler as palavras esperadas (tanto no P300, que aqui se conjectura estar associado à expectativa léxica, quanto no N400). Já nas palavras que tornam as sentenças falsas, espera-se que o filósofo apresente um N400 bastante negativo e, possivelmente, um P300 mais positivo²⁵:

²⁵ 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 constriam tanto a expectativa do leitor, não gerando expectativa léxica, como “Descartes foi um filósofo medieval/contemporâneo”. O preço desta abordagem que envolve menor precisão é a obtenção de um *N400 Effect* mais fraco, requerendo um maior número de participantes para se obter dados significativos.

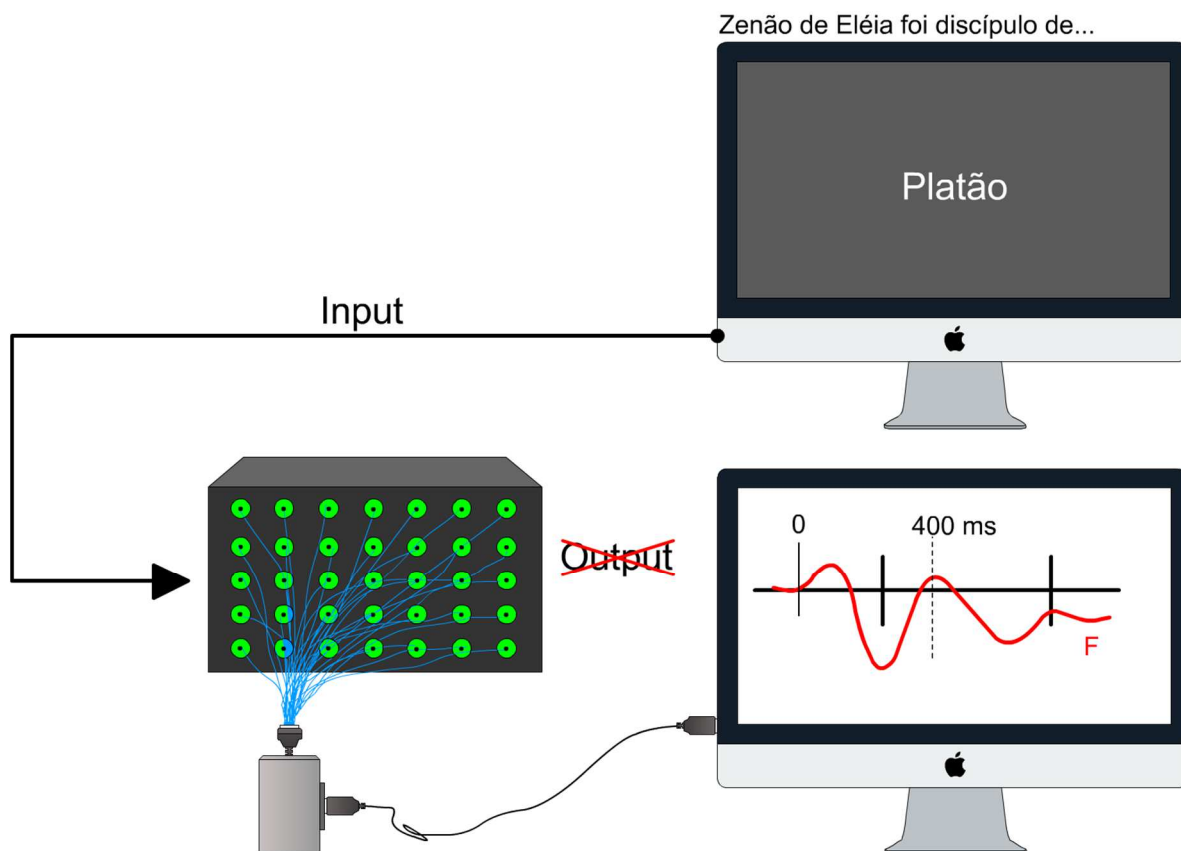


Figura 11 – Representação do sinal esperado pelos participantes que tiverem conhecimento das sentenças do conjunto @, ao lerem a palavra incorreta: um N400 forte.

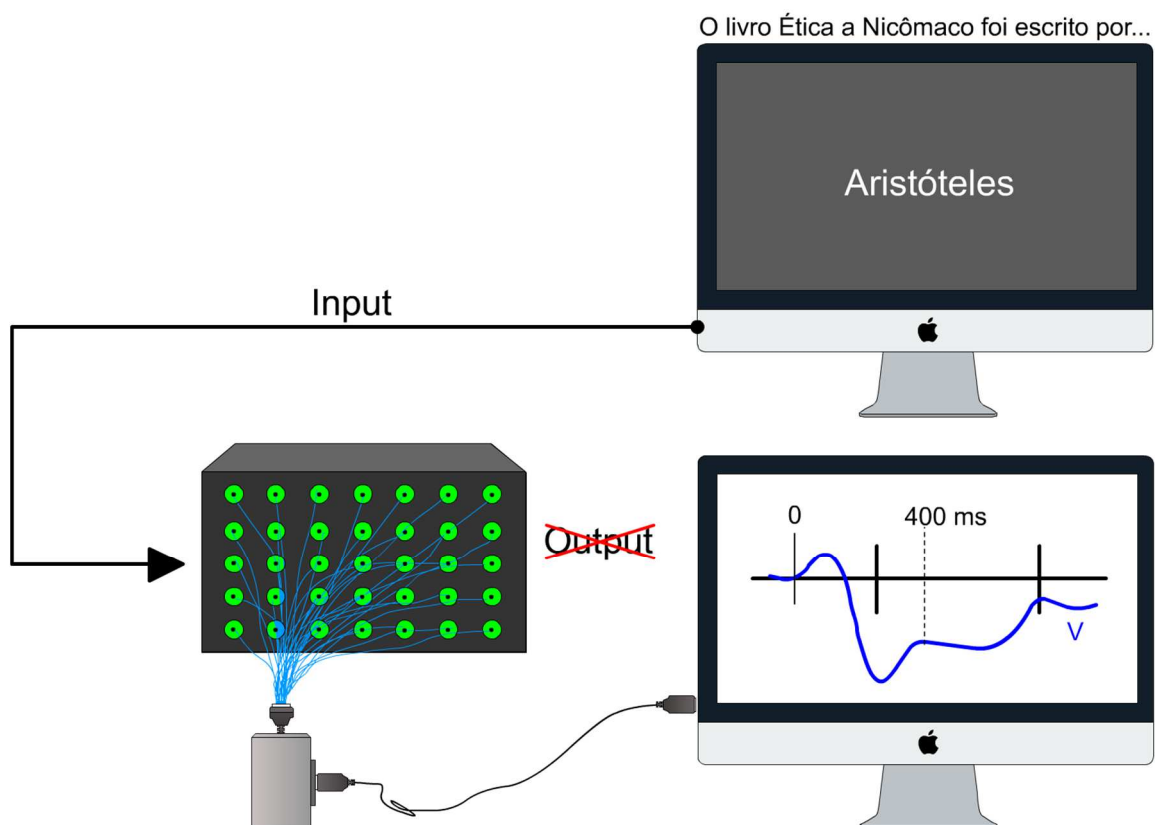


Figura 12 - Representação do sinal esperado pelos participantes que tiverem conhecimento das sentenças do conjunto @, ao lerem a palavra correta: um N400 fraco.

Um 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 13 ilustra os resultados esperados nas duas condições (Verdadeiro e Falso) nas duas categorias (Filósofos e Não-filósofos). Já a Figura 14 apresenta os *N400 Effects* esperados nas categorias. Para análise estatística será calculada a amplitude média dos *N400 Effects* de cada indivíduo do intervalo entre 300 e 500 milissegundos após a apresentação do estímulo crítico²⁶.

²⁶ As etapas de análise dos dados são as seguintes: a) filtrar o sinal do EEG bruto de 0.4 a 8Hz; 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 a 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 palavras 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.

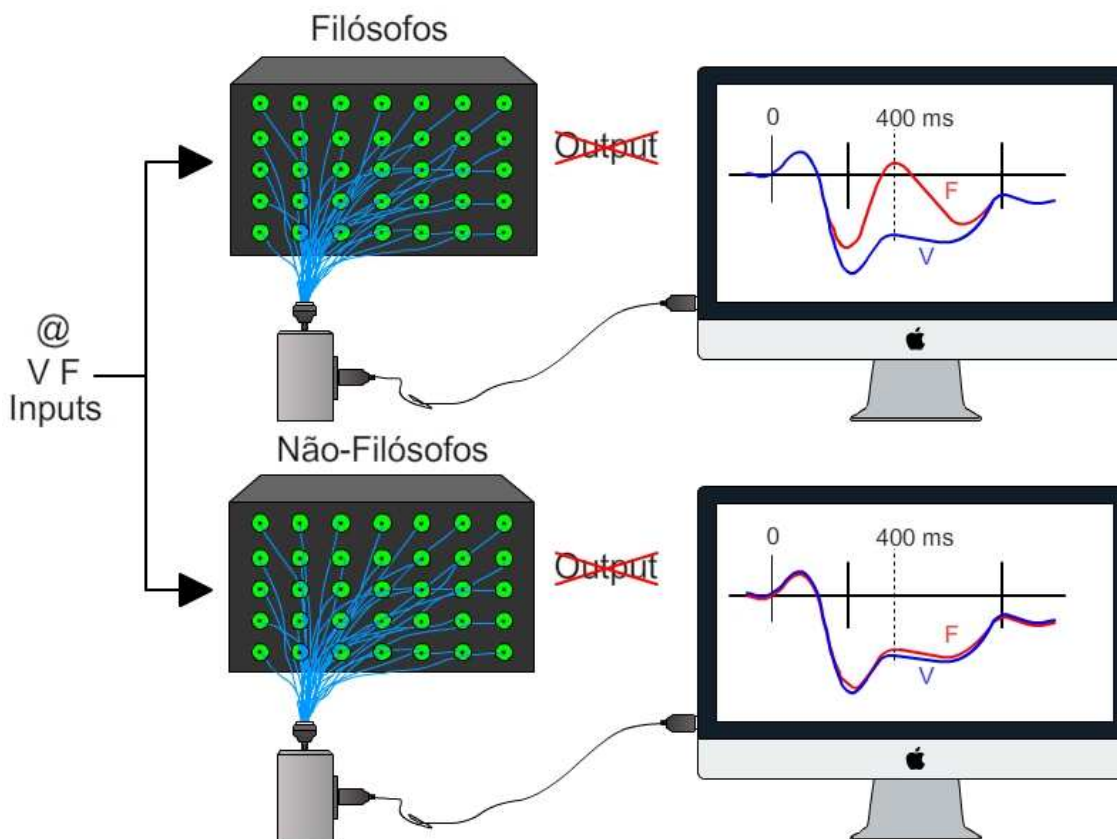


Figura 13 – Comparação das médias esperadas dos sinais dos filósofos e não-filósofos, nas sentenças verdadeiras e falsas.

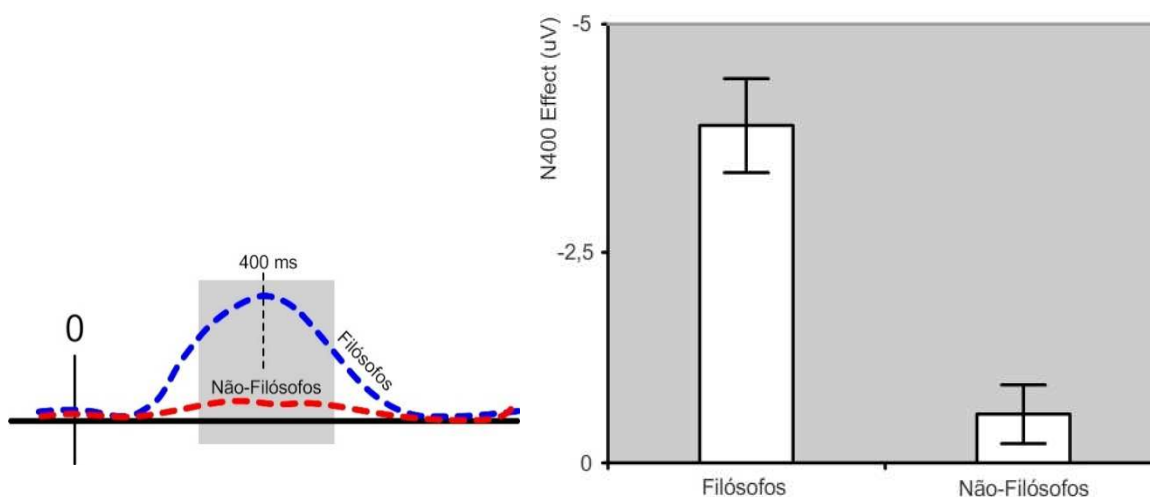


Figura 14 – Intervalo de análise das amplitudes médias (na figura esquerda, em cinza) e barras de erro esperadas dos N400 Effects dos Filósofos e Não-Filósofos (direita).

O atual experimento não extrairia propriamente dos participantes informações acerca de seus estados mentais que já não eram pressupostas: assume-se que os filósofos possuam conhecimento filosófico, enquanto os participantes da

categoria dos não-filósofos serão selecionados apenas se afirmarem antecipadamente ignorar a maioria dos assuntos filosóficos, e se espera encontrar *N400 Effects* significativamente diferentes entre os que sabem e os que ignoram, dadas as razões supramencionadas. Mas se o previsto ocorrer, então será a princípio possível inverter a situação, indo um passo além: em vez de escolhermos 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 último caso teríamos uma extração de informação não-suposta acerca da mentalidade alheia, permitindo uma “leitura de mentes” que prescindia da observação de outputs comportamentais dados certos inputs. O participante recebe inputs sem ser instruído a realizar nenhuma tarefa a não ser a leitura passiva do que lhe é exposto, e dependendo do modo como seu cérebro responde a estes inputs conseguir-se-ia inferir as crenças que estão a modular as respostas eletrofisiológicas²⁷.

Como fora exposto nos capítulos anteriores, teorias filosóficas sobre o significado dos termos que se referem a estados mentais intencionais (como crenças e desejos) costumam variar de eliminativismos a realismos ingênuos, 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. 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

²⁷ Aqui foi utilizado o termo “inferir” num sentido fraco e pragmático. Na verdade, a inferência reversa constitui uma afirmação do conseqüente – 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 conseqüente, conforme é demonstrado no argumento da Figura 17).

eletrofisiológicos provenientes do interior da *black box*, como até aqui demonstrada, 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* para prever 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 incorreta²⁸”. 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).

²⁸ 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”.

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, prever 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 15 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 @):

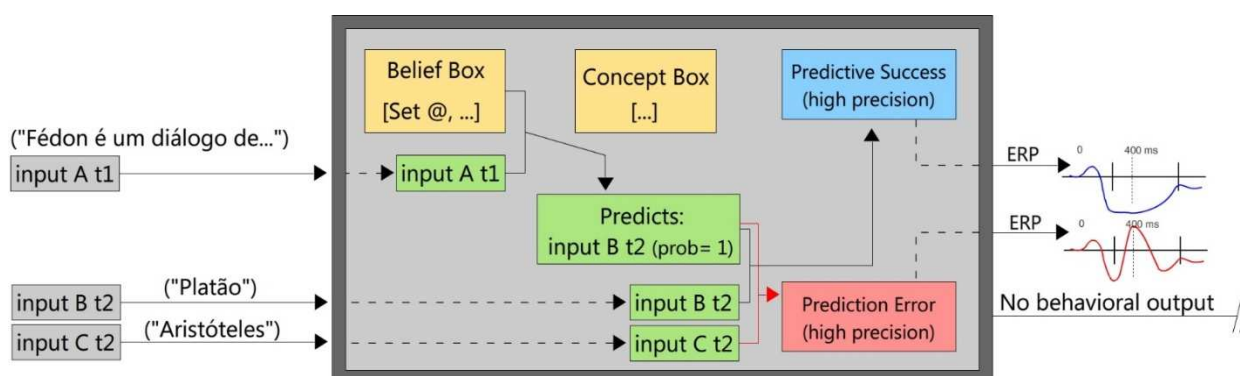


Figura 15 – Modelo de processo cognitivo que modularia os ERP's dos filósofos em uma condição preditiva de alta precisão.

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):

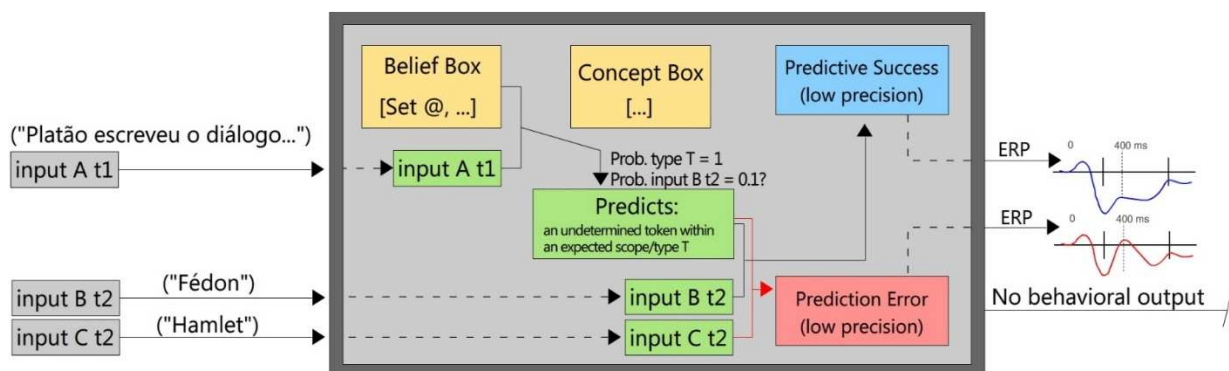


Figura 16 – Modelo de processo cognitivo que modularia os ERP's dos filósofos em uma condição preditiva de baixa precisão.

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...”)²⁹. 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 16 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 surpresa 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 elicitando ERP's diferenciados quando uma palavra incorreta – porém a ele plausível – aparecer:

²⁹ Cf. LAU, E.F., HOLCOMB, P.J. & KUPERBERG, GR., 2013.

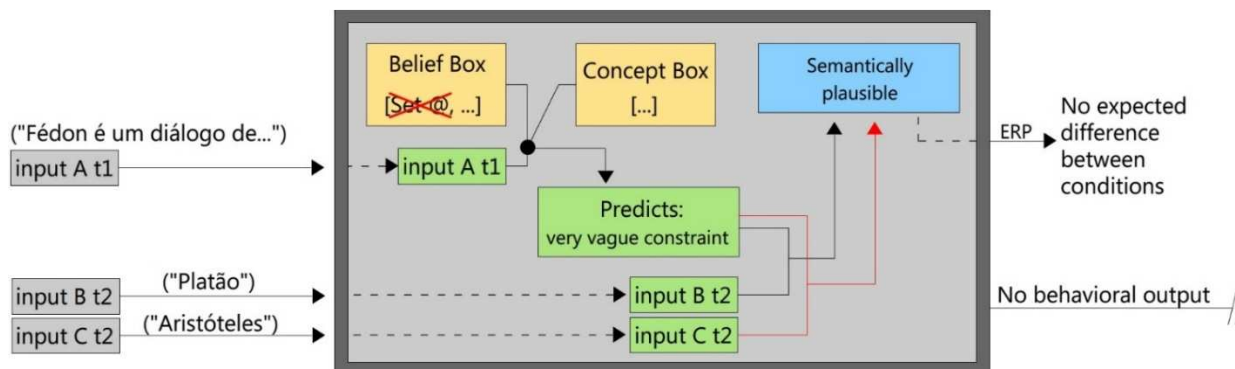


Figura 17 – Modelo de processo cognitivo que modularia os ERP's dos não-filósofos em uma condição preditiva de alta precisão.

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 expectativa semântica será violada:

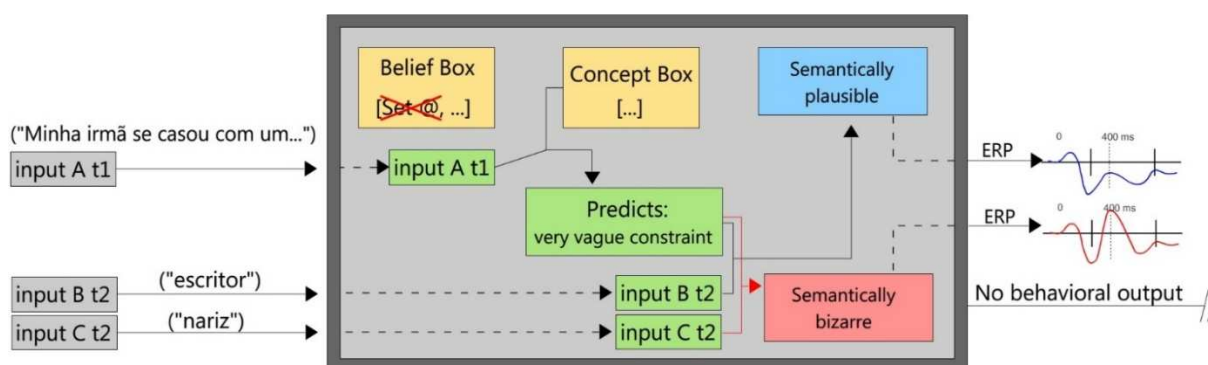


Figura 18 – Modelo de processo cognitivo que modularia os ERP's de sentenças semanticamente congruentes/incongruentes.

Os quatro modelos esboçados acima possuem em comum o fato de previsões satisfeitas elicitarem um N400 mais atenuado que o de previsões frustradas: haveria uma conjunção constante entre frustrações de expectativas (aqui descritas como eventos mentais, subjetivos) e N400 acentuados (aqui descritos como eventos

eletrofisiológicos, objetivos)³⁰, assim como entre expectativas satisfeitas e N400 atenuados. Nas violações de conhecimento, a diferença entre as condições de alta e baixa precisão seria o grau de incerteza dos inputs esperados, o que viria a refletir nos ERP's³¹. 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 (representações de categorias, propriedades, abstrações)³².

5.2 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:

³⁰ 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 ($\forall x$) (x está com dor) \rightarrow (as fibras-C de x disparam)).

³¹ 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).

³² Certamente conceitos são os “*building blocks*” que compõem as crenças, porém, crenças são representações de um nível superior ao de conceitos, envolvendo proposições que são ou verdadeiras ou falsas, sendo verdade e falsidade nunca predicáveis de conceitos (e.g. não há um gato verdadeiro). Em acréscimo a isso, crenças envolvem um tipo específico de atitude em relação a uma proposição (posso pensar em um estado de coisas sem me comprometer com sua correspondência com o mundo, o que não ocorre quando possuo uma crença).

Premissas Gerais	1	$(\forall x)(\forall \omega)(\forall \varphi) x \text{ é um filósofo } \wedge \text{"}\omega... \varphi\text{" é uma sentença do conjunto } @ \rightarrow x \text{ acredita que } [\omega... \varphi \wedge \neg(\exists \theta) \omega... \theta \wedge \neg(\theta = \varphi)]$	P
	2	$(\forall x)(\forall \omega)(\forall \varphi)(\forall \beta)(\forall t) x \text{ acredita que } [\omega... \varphi \wedge \neg(\exists \theta) \omega... \theta \wedge \neg(\theta = \varphi)] \wedge \text{"}\beta\text{" é apresentado a } x \text{ no instante } t \text{ após } x \text{ ter lido "}\omega\text{" } \wedge \neg(\beta = \varphi) \rightarrow x \text{ tem a sua expectativa semântica frustrada ao ler a palavra "}\beta\text{" que foi apresentada no instante } t$	P
	3	$(\forall x)(\forall \omega)(\forall \varphi)(\forall t) x \text{ acredita que } [\omega... \varphi \wedge \neg(\exists \theta) \omega... \theta \wedge \neg(\theta = \varphi)] \wedge \text{"}\varphi\text{" é apresentado a } x \text{ no instante } t \text{ após } x \text{ ter lido "}\omega\text{" } \wedge \rightarrow x \text{ tem a sua expectativa semântica satisfeita ao ler a palavra "}\varphi\text{" que foi apresentada no instante } t$	P
Hipóteses	4	$(\forall x)(\forall \varphi)(\forall t) x \text{ tem a sua expectativa semântica frustrada ao ler a palavra "}\varphi\text{" que foi apresentado no instante } t \wedge \text{ os sinais eletroencefalográficos de } x \text{ estão sendo medidos a partir do instante } t \rightarrow \text{ os sinais dos eletrodos de } x \text{ localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } t$	H
	5	$(\forall x)(\forall \varphi)(\forall t) x \text{ tem a sua expectativa semântica satisfeita ao ler a palavra "}\varphi\text{" que foi apresentado no instante } t \wedge \text{ os sinais eletroencefalográficos de } x \text{ estão sendo medidos a partir do instante } t \rightarrow \neg(\text{os sinais dos eletrodos de } x \text{ localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } t)$	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	$\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})$	P
	10	"Fédon é um diálogo de Platão" é uma sentença do conjunto @	P
Consequências lógicas das premissas	11	$\text{Pedro é um filósofo } \wedge \text{"Fédon é um diálogo de Platão" é uma sentença do conjunto } @$	CONJ. 6,10
	12	$\text{Pedro é um filósofo } \wedge \text{"Fédon é um diálogo de Platão" é uma sentença do conjunto } @ \rightarrow \text{Pedro acredita que } [\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$	EU1
	13	$\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$	MP 11,12
	14	$\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})] \wedge \text{"Platão" é apresentado a Pedro no instante } 0 \text{ após Pedro ter lido "Fédon é um diálogo de" } \rightarrow \text{Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante } 0$	EU3
	15	$\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})] \wedge \text{"Platão" é apresentado a Pedro no instante } 0 \text{ após Pedro ter lido "Fédon é um diálogo de"}$	CONJ. 7,13
	16	$\text{Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante } 0$	MP 14,15
	17	$\text{Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante } 0 \wedge \text{ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante } 0$	CONJ. 8,16
Consequências lógicas das hipóteses 4 e 5	18	$\text{Pedro tem a sua expectativa semântica satisfeita ao ler a palavra "Platão" que foi apresentada no instante } 0 \wedge \text{ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante } 0 \rightarrow \neg(\text{os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } 0)$	EU5
	19	$\neg(\text{Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } 0)$	MP 17,18
Sinal do EEG de Pedro	20	$\neg(\text{Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } 0)$	P

Figura 19 – Formalização das consequências lógicas da hipótese 5 em filósofos.

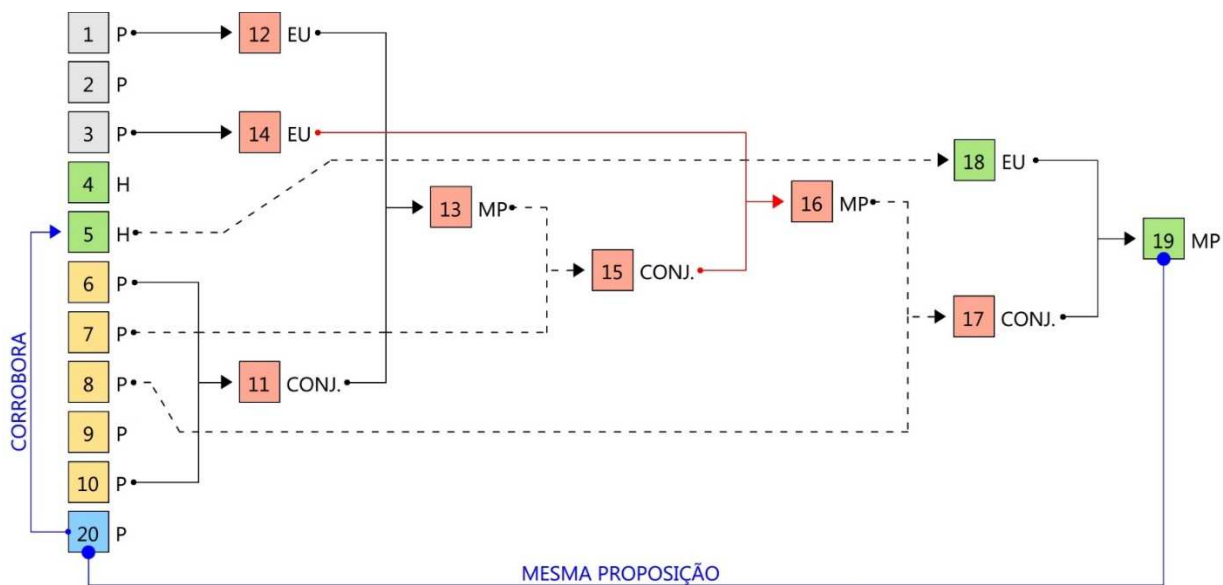


Figura 20 – Esquema simplificado da estrutura do argumento exposto na Figura 19.

Premissas Gerais	1	$(\forall x)(\forall \omega)(\forall \varphi) x \text{ é um filósofo} \wedge " \omega \dots \varphi " \text{ é uma sentença do conjunto } @ \rightarrow x \text{ acredita que } [\omega \dots \varphi \wedge \neg(\exists \theta) \omega \dots \theta \wedge \neg(\theta = \varphi)]$	P
	2	$(\forall x)(\forall \omega)(\forall \varphi)(\forall \beta)(\forall t) x \text{ acredita que } [\omega \dots \varphi \wedge \neg(\exists \theta) \omega \dots \theta \wedge \neg(\theta = \varphi)] \wedge " \beta " \text{ é apresentado a } x \text{ no instante } t \text{ após } x \text{ ter lido " } \omega " \wedge \neg(\beta = \varphi) \rightarrow x \text{ tem a sua expectativa semântica frustrada ao ler a palavra " } \beta " \text{ que foi apresentada no instante } t$	P
	3	$(\forall x)(\forall \omega)(\forall \varphi)(\forall t) x \text{ acredita que } [\omega \dots \varphi \wedge \neg(\exists \theta) \omega \dots \theta \wedge \neg(\theta = \varphi)] \wedge " \varphi " \text{ é apresentado a } x \text{ no instante } t \text{ após } x \text{ ter lido " } \omega " \wedge \rightarrow x \text{ tem a sua expectativa semântica satisfeita ao ler a palavra " } \varphi " \text{ que foi apresentada no instante } t$	P
Hipóteses	4	$(\forall x)(\forall \varphi)(\forall t) x \text{ tem a sua expectativa semântica frustrada ao ler a palavra " } \varphi " \text{ que foi apresentado no instante } t \wedge \text{ os sinais eletroencefalográficos de } x \text{ estão sendo medidos a partir do instante } t \rightarrow \text{ os sinais dos eletrodos de } x \text{ localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } t$	H
	5	$(\forall x)(\forall \varphi)(\forall t) x \text{ tem a sua expectativa semântica satisfeita ao ler a palavra " } \varphi " \text{ que foi apresentado no instante } t \wedge \text{ os sinais eletroencefalográficos de } x \text{ estão sendo medidos a partir do instante } t \rightarrow \neg(\text{os sinais dos eletrodos de } x \text{ localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } t)$	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	$\text{Fédon é um diálogo de Platão} \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})$	P
	10	Fédon é um diálogo de Platão é uma sentença do conjunto @	P
	11	$\neg(\text{Kant} = \text{Platão})$	P
Consequências lógicas das premissas	12	$\text{Pedro é um filósofo} \wedge \text{Fédon é um diálogo de Platão é uma sentença do conjunto } @$	CONJ. 6,10
	13	$\text{Pedro é um filósofo} \wedge " \text{Fédon é um diálogo de Platão} " \text{ é uma sentença do conjunto } @ \rightarrow \text{ Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$	EU1
	14	$\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$	MP 12,13
	15	$\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})] \wedge " \text{Kant} " \text{ é apresentado a Pedro no instante } 0 \text{ após Pedro ter lido "Fédon é um diálogo de" } \wedge \neg(\text{Kant} = \text{Platão}) \rightarrow \text{ Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante } 0$	EU2
	16	$\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})] \wedge " \text{Kant} " \text{ é apresentado a Pedro no instante } 0 \text{ após Pedro ter lido "Fédon é um diálogo de" } \wedge \neg(\text{Kant} = \text{Platão})$	CONJ. 7,11,14
	17	$\text{Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante } 0$	MP 15,16
	18	$\text{Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante } 0 \wedge \text{ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante } 0$	CONJ. 8,17
Consequências lógicas das hipóteses 4 e 5	19	$\text{Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante } 0 \wedge \text{ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante } 0 \rightarrow \text{ os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } 0$	EU4
	20	$\text{Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } 0$	MP 18,19
Sinal do EEG de Pedro	21	$\text{Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } 0$	P

Figura 21 - Formalização das consequências lógicas da hipótese 4 em filósofos.

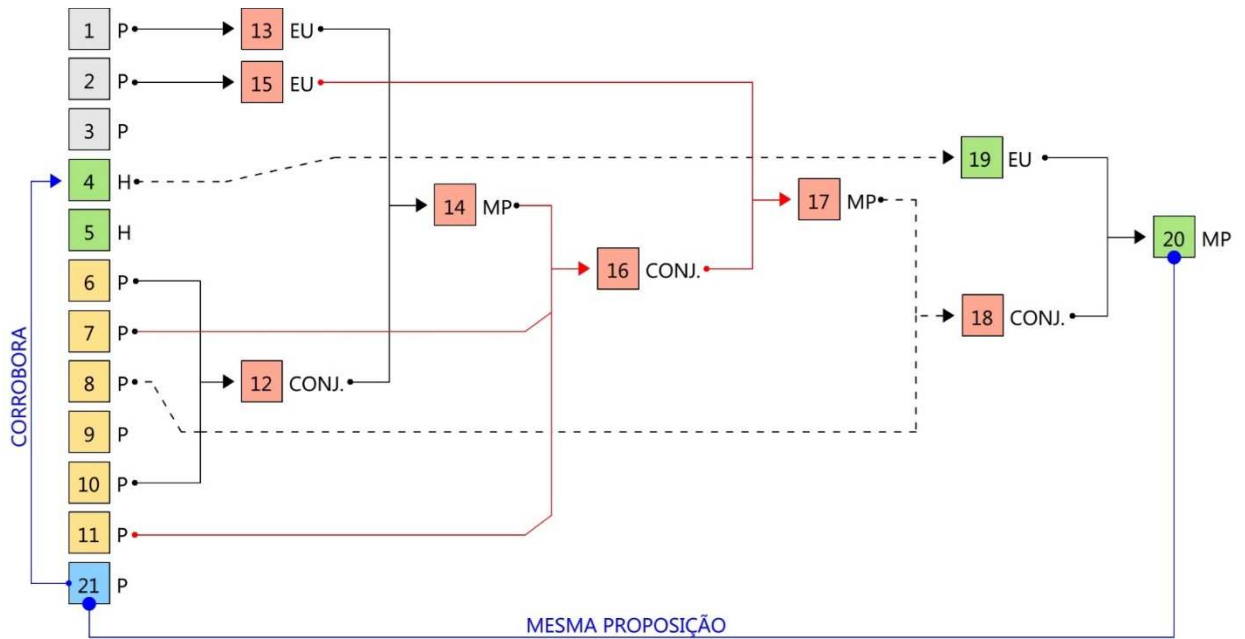


Figura 22 - Esquema simplificado da estrutura do argumento exposto na Figura 21.

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 Figura 23 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	$(\forall x)(\forall \omega)(\forall \varphi) x \text{ é um filósofo } \wedge \text{"}\omega... \varphi\text{" é uma sentença do conjunto } @ \rightarrow x \text{ acredita que } [\omega... \varphi \wedge \neg(\exists \theta) \omega... \theta \wedge \neg(\theta = \varphi)]$	P
	2	$(\forall x)(\forall \omega)(\forall \varphi)(\forall \beta)(\forall t) x \text{ acredita que } [\omega... \varphi \wedge \neg(\exists \theta) \omega... \theta \wedge \neg(\theta = \varphi)] \wedge \text{"}\beta\text{" é apresentado a } x \text{ no instante } t \text{ após } x \text{ ter lido "}\omega\text{" } \wedge \neg(\beta = \varphi) \rightarrow x \text{ tem a sua expectativa semântica frustrada ao ler a palavra "}\beta\text{" que foi apresentada no instante } t$	P
	3	$(\forall x)(\forall \omega)(\forall \varphi)(\forall t) x \text{ acredita que } [\omega... \varphi \wedge \neg(\exists \theta) \omega... \theta \wedge \neg(\theta = \varphi)] \wedge \text{"}\varphi\text{" é apresentado a } x \text{ no instante } t \text{ após } x \text{ ter lido "}\omega\text{" } \wedge \rightarrow x \text{ tem a sua expectativa semântica satisfeita ao ler a palavra "}\varphi\text{" que foi apresentada no instante } t$	P
	4	$(\forall x)(\forall \varphi)(\forall t) x \text{ tem a sua expectativa semântica frustrada ao ler a palavra "}\varphi\text{" que foi apresentado no instante } t \wedge \text{ os sinais eletroencefalográficos de } x \text{ estão sendo medidos a partir do instante } t \rightarrow \text{ os sinais dos eletrodos de } x \text{ localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } t$	P
	5	$(\forall x)(\forall \varphi)(\forall t) x \text{ tem a sua expectativa semântica satisfeita ao ler a palavra "}\varphi\text{" que foi apresentado no instante } t \wedge \text{ os sinais eletroencefalográficos de } x \text{ estão sendo medidos a partir do instante } t \rightarrow \neg(\text{os sinais dos eletrodos de } x \text{ localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após } t)$	P
	6	$(\forall \pi) \neg(x \text{ acredita que } \pi) \wedge \pi \rightarrow x \text{ ignora que } \pi$	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 $\wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})$	P
	11	"Fédon é um diálogo de Platão" é uma sentença do conjunto @	P
	12	$\neg(\text{Kant} = \text{Platão})$	P
Consequências lógicas das premissas	13	Pedro é um filósofo \wedge "Fédon é um diálogo de Platão" é uma sentença do conjunto @ \rightarrow Pedro acredita que $[\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$	EU1
	14	Pedro acredita que $[\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$ \wedge "Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" $\wedge \neg(\text{Kant} = \text{Platão}) \rightarrow$ 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 \wedge Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0 \rightarrow 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
Consequências lógicas das hipóteses 7	16	Pedro é um filósofo \wedge "Fédon é um diálogo de Platão" é uma sentença do conjunto @	CONJ. 7,11
	17	Pedro acredita que $[\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$	MP 13,16
	18	Pedro acredita que $[\text{Fédon é um diálogo de Platão } \wedge \neg(\exists \theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta = \text{Platão})]$ \wedge "Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" $\wedge \neg(\text{Kant} = \text{Platão})$	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 \wedge 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	$\neg(\text{Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0})$	P
Contradição deduzida da hipótese 7	23	$(\text{Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0}) \wedge \neg(\text{Os sinais dos eletrodos de Pedro localizados nas regiões central e parietal apresentam uma negatividade de aproximadamente 3 microvolts com latência de aproximadamente 400 ms após 0})$	C 21,22
Conclusões	24	$\neg(\text{Pedro é um filósofo})$	RAA 7-23
	25	$\neg(\text{Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0 } \wedge \text{ Os sinais eletroencefalográficos de Pedro estão sendo medidos a partir do instante 0})$	MT 15,22
	26	$\neg(\text{Pedro tem a sua expectativa semântica frustrada ao ler a palavra "Kant" que foi apresentada no instante 0})$	9,25

Conclusões	27	$\neg\{\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão})] \wedge \text{"Kant" é apresentado a Pedro no instante 0 após Pedro ter lido "Fédon é um diálogo de" } \wedge \neg(\text{Kant}=\text{Platão})\}$	MT 14,26
	28	$\neg\{\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão})]\}$	8,12,27
	29	$\neg\{\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão})]\} \wedge \text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão}) \rightarrow \text{Pedro ignora que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão})]$	EU6
	30	$\neg\{\text{Pedro acredita que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão})]\} \wedge \text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão})]$	CONJ. 10,28
	31	$\text{Pedro ignora que } [\text{Fédon é um diálogo de Platão} \wedge \neg(\exists\theta) \text{ Fédon é um diálogo de } \theta \wedge \neg(\theta=\text{Platão})]$	MP 29,30

Figura 23 – 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.

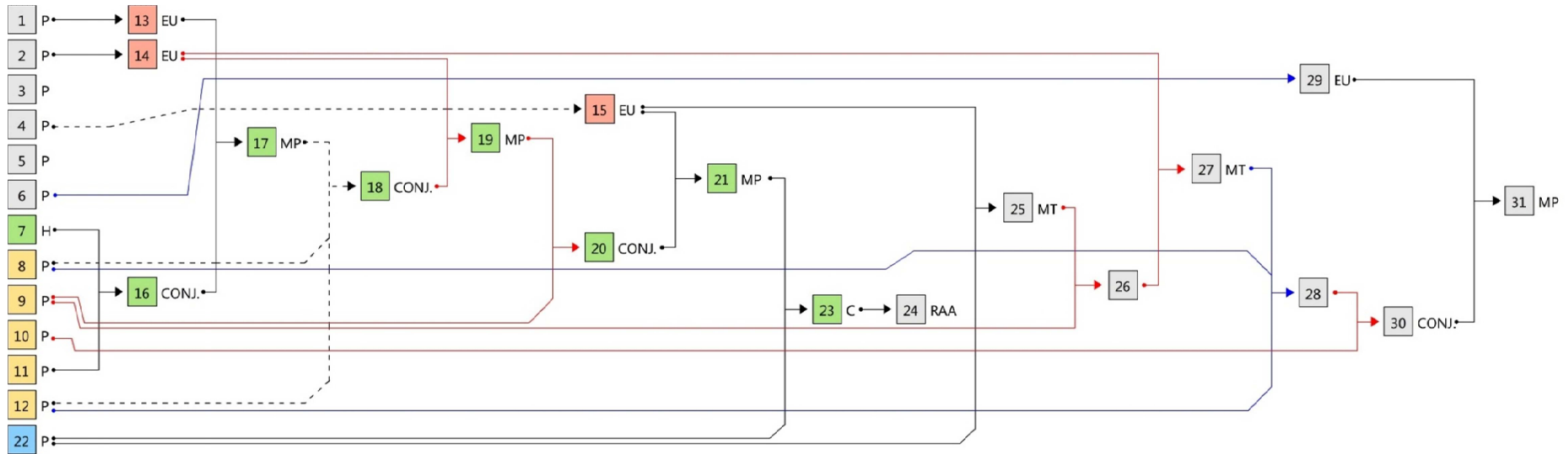


Figura 24 - Esquema simplificado da estrutura do argumento exposto na figura 23.

Para explicitar a estrutura [supostamente] válida 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. 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”³³. 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 70%.

³³ 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).

5.3 RESULTADOS

5.3.1 N400

Por se tratar de um componente com ampla distribuição no escalpo, procurou-se analisar o N400 como um padrão distribuído em 52 eletrodos posicionados nas regiões central, parietal e temporal direita nas quais o N400 costuma se destacar (eletrodos 4, 5, 6, 7, 31, 37, 42, 47, 52, 53, 54, 55, 59, 60, 61, 62, 65, 66, 67, 70, 71, 72, 75, 76, 77, 78, 79, 80, 83, 84, 85, 86, 87, 90, 91, 92, 93, 96, 97, 98, 102, 103, 104, 105, 106, 109, 110, 111, 112, 117, 118 e 124). O equipamento utilizado foi um EEG de 128 eletrodos *Hydrocel Geodesic Sensor Net* (Electrical Geodesic, Eugene, USA) e a referência *online* foi o eletrodo Cz. As posições dos eletrodos estão representadas na Figura 27. A referência *offline* utilizada foi a média de todos os eletrodos (*average-reference*). O sinal do EEG bruto foi filtrado de 0.4 a 8Hz. Foram gerados 80 segmentos de 700 milissegundos (-100 a 600) para cada participante, recortados do sinal filtrado (o instante 0 é o da apresentação da palavra crítica), gerando 40 segmentos de cada condição. Os segmentos contaminados com artefatos e canais ruins foram eliminados. As médias das condições verdadeiro e falso foram calculadas a partir dos segmentos restantes. Os 100 milissegundos que precedem o estímulo foram utilizados como *baseline*. A *differential wave* resultante da subtração do ERP das palavras corretas do ERP das palavras incorretas (=N400 Effect) foi gerada para cada indivíduo. A média aritmética de cada grupo (*grand average*) foi computada a partir dos *N400 Effects* individuais. Calculou-se, para análise estatística, a média da amplitude do *N400 Effect* de cada participante no intervalo de 300 a 500 milissegundos (em todos os eletrodos). Os valores foram corrigidos utilizando *Mean Correction*³⁴. Considerou-se como variável dependente a média dos valores dos 52 eletrodos selecionados.

A média do *N400 Effect* dos 52 eletrodos no grupo dos Filósofos (N=26; média de idade = 28,15; desvio padrão = 11,03) foi de -0,3478 μV , enquanto a dos Não-Filósofos (N=9; média de idade = 28,1; desvio padrão = 7,32) foi de 0,0811 μV . Utilizando One-Way ANOVA, verificou-se que a diferença das médias foi significativa [$F(1,33) = 11.073, p=0.002$], conforme as barras de erro e cálculo expostos Figura 30.

³⁴ O *Mean Correction* é utilizado para medir o quanto a amplitude varia em relação à média do segmento inteiro, subtraindo-se a média da amplitude do segmento do valor da amplitude de cada instante.

Separando do grupo dos Filósofos os 12 alunos do primeiro semestre de Filosofia – criando o grupo dos Calouros (N=12; média de idade = 23,41; desvio padrão = 10,51) –, pôde-se observar *N400 Effects* significativamente mais negativos nos Filósofos (N=14; média de idade = 32,21; desvio padrão = 10,1) do que nos Calouros [$F(1,24) = 4.725, p = 0.04$]. A diferença entre a média dos Calouros e dos Não-filósofos não atingiu o nível de significância de 0.05, ainda que tenha ficado muito próximo de ser significativo [$F(1,19) = 4.238, p = 0.054$]³⁵. Os cálculos ANOVA e as barras de erro estão expostos na Figura 31.

Os dados obtidos estão em conformidade com o que era previsto pela hipótese defendida (*N400 Effects* significativamente mais negativos no grupo dos Filósofos em relação ao grupo dos Não-Filósofos). Os resultados apresentaram o mesmo padrão encontrado na réplica do experimento de Kutas (1980) acerca de anomalias semânticas (Figura 26). Foi possível observar não apenas diferenças eletrofisiológicas significativas entre conhecedores e ignorantes de um conjunto de fatos, mas também entre graus de conhecimento, como entre alunos iniciantes do curso de Filosofia e alunos veteranos. A Figura 28 apresenta os ERP's obtidos, a Figura 29 os mapas topográficos. No presente experimento parte-se da premissa de que o grau de conhecimento dos alunos veteranos seja superior ao dos calouros, e o dos calouros superior ao dos não-filósofos, porém não se mediu explicitamente o conhecimento de cada voluntário (o que permitiria enxergar diferenças detalhadas dentro dos grupos, como a presença no grupo dos não-filósofos de um sujeito que tenha conhecimento filosófico superior ao de um aluno iniciante do curso de Filosofia). Uma possibilidade futura é submeter cada participante a um questionário, possibilitando encontrar correlações negativas entre o grau de conhecimento explicitamente medidos e os *N400 Effects*. Apesar do atual experimento classificar os voluntários em apenas três níveis de conhecimento, já é possível perceber de modo preliminar que o *N400 Effect* fica mais negativo em função do grau de conhecimento, conforme a Figura 25:

³⁵ Espera-se, entretanto, que a diferença se torne significativa futuramente com o aumento da amostra.

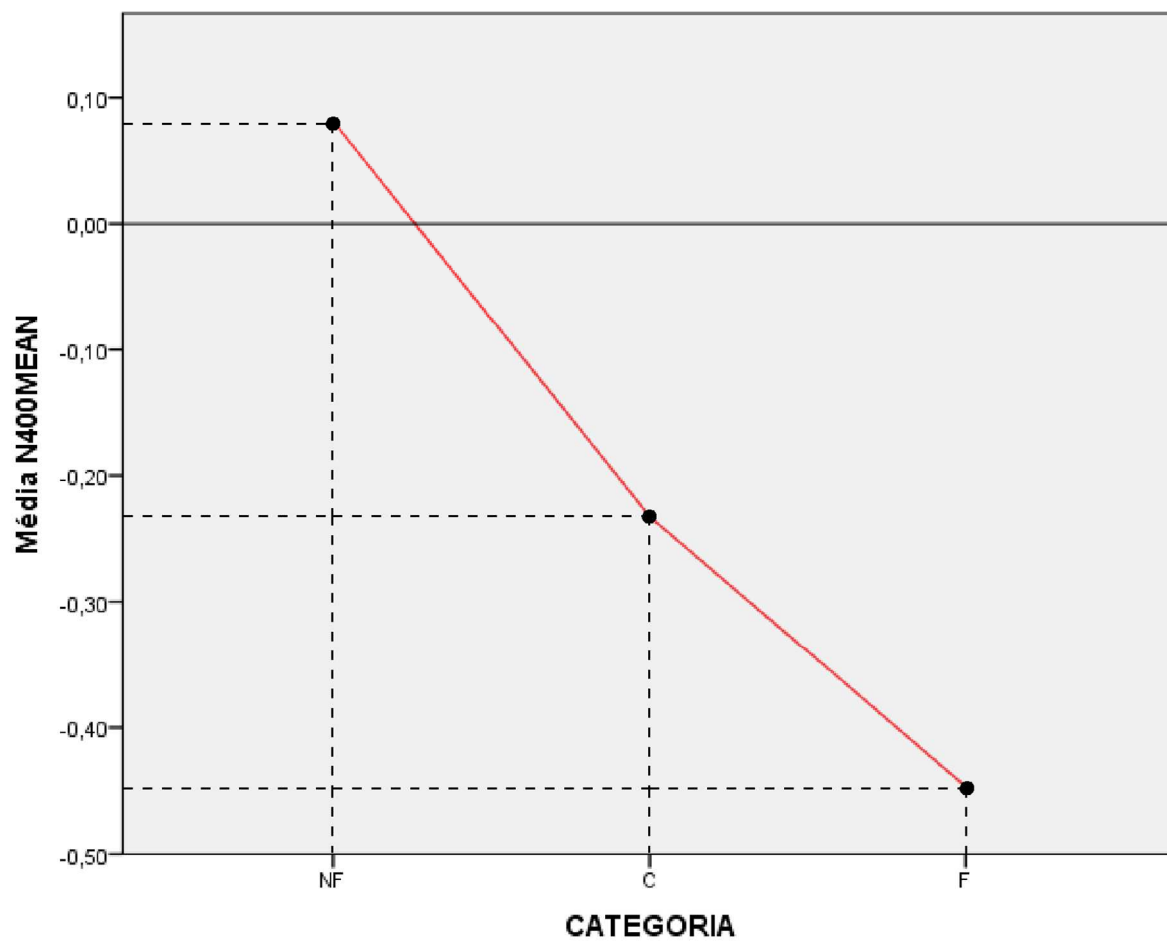


Figura 25 – Variação da amplitude do *N400 effect* em função do grau de conhecimento da cada grupo.

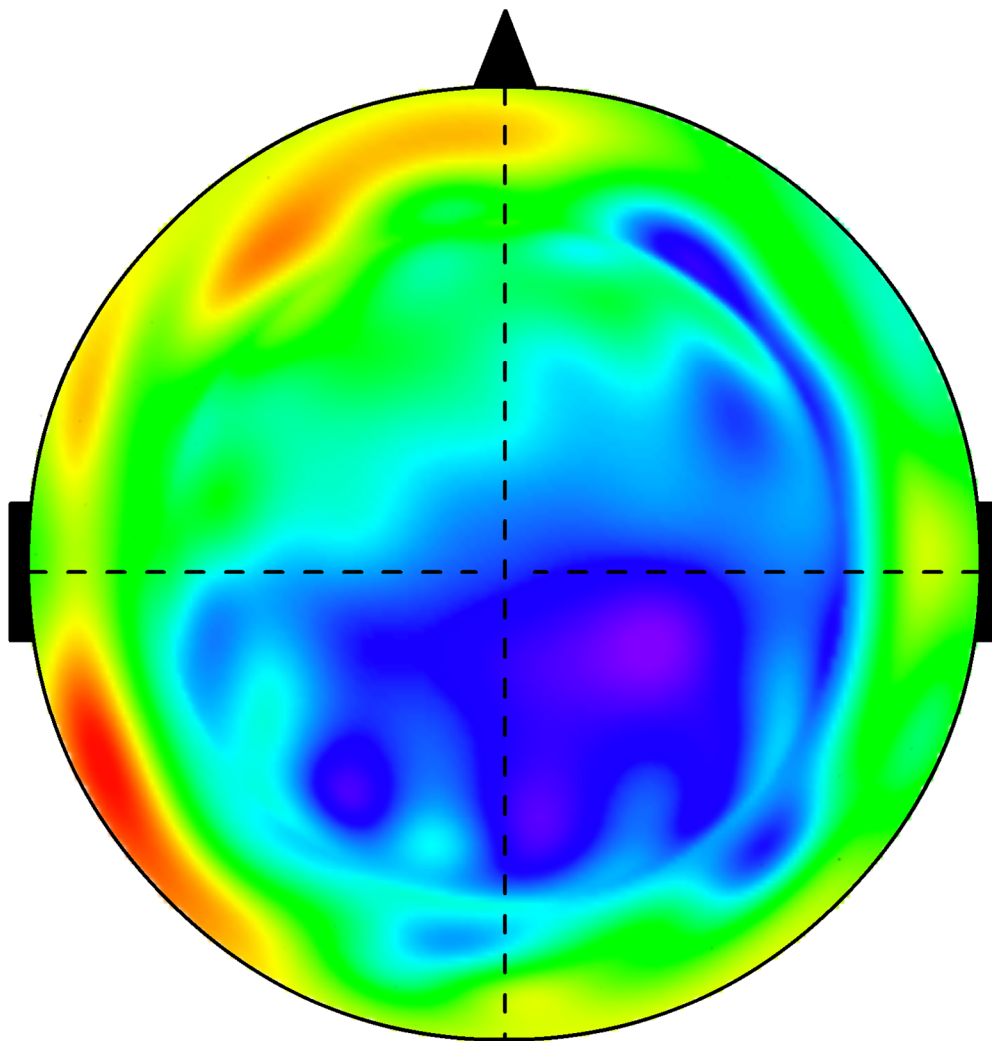


Figura 26 - Mapa topográfico do *N400 Effect* do experimento envolvendo anomalias semânticas, replicando metodologia de KUTAS et al. (1980).

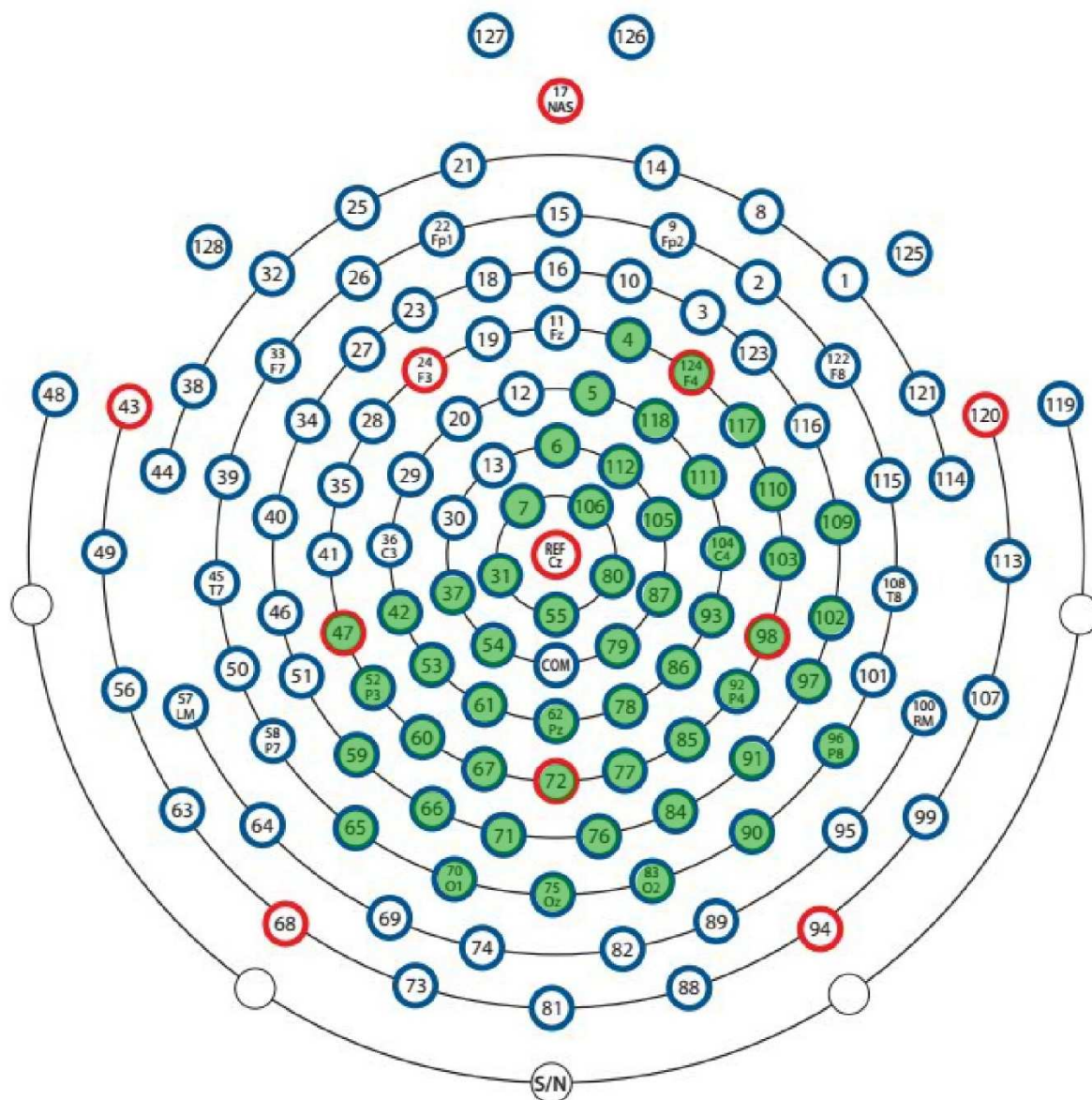


Figura 27 – Eletrodos selecionados para análise estatística do N400.

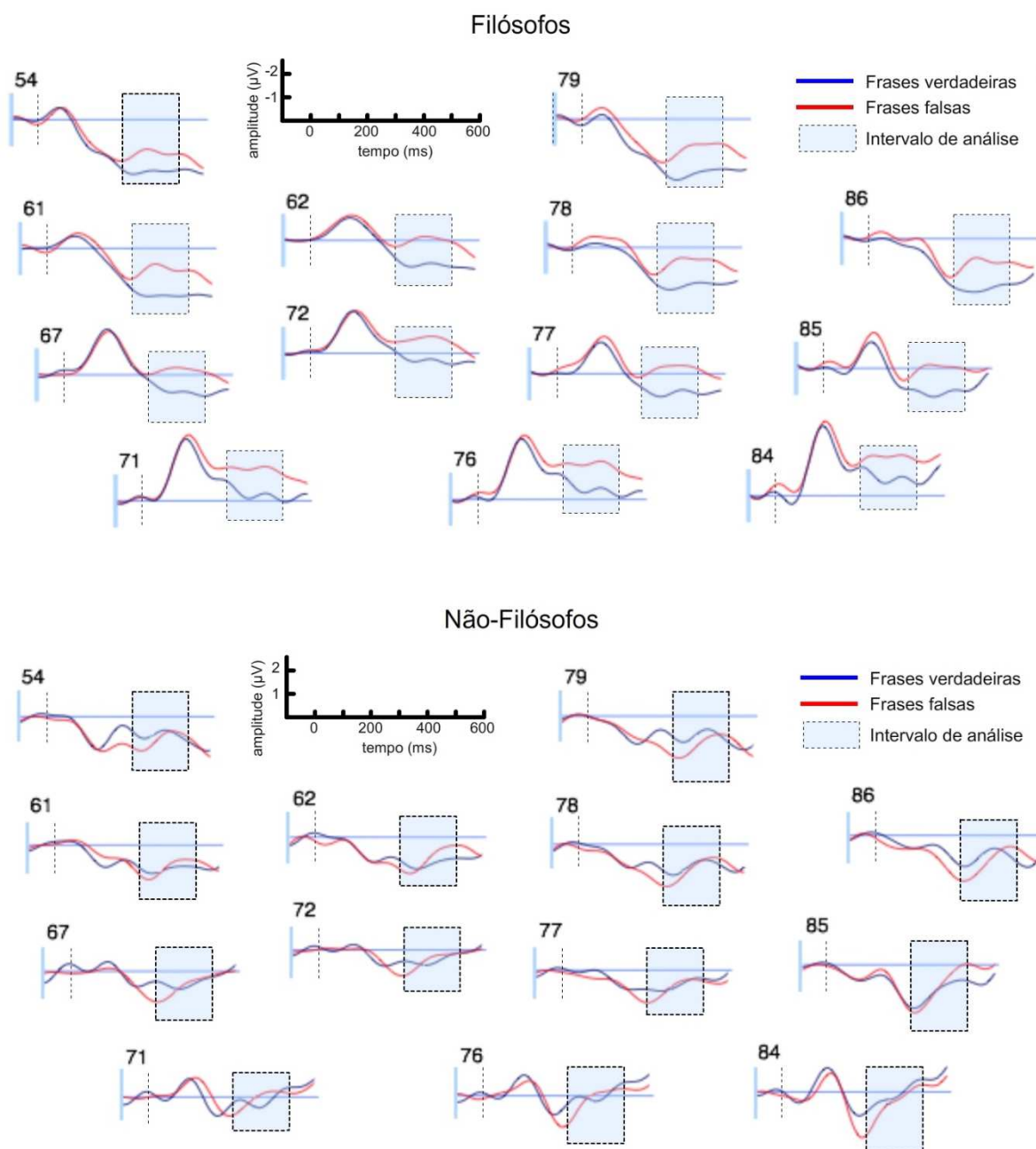


Figura 28 – Diferença nos ERP's do grupo dos filósofos e dos não-filósofos em 13 eletrodos posicionados na região centro-parietal, nas frases verdadeiras e falsas.

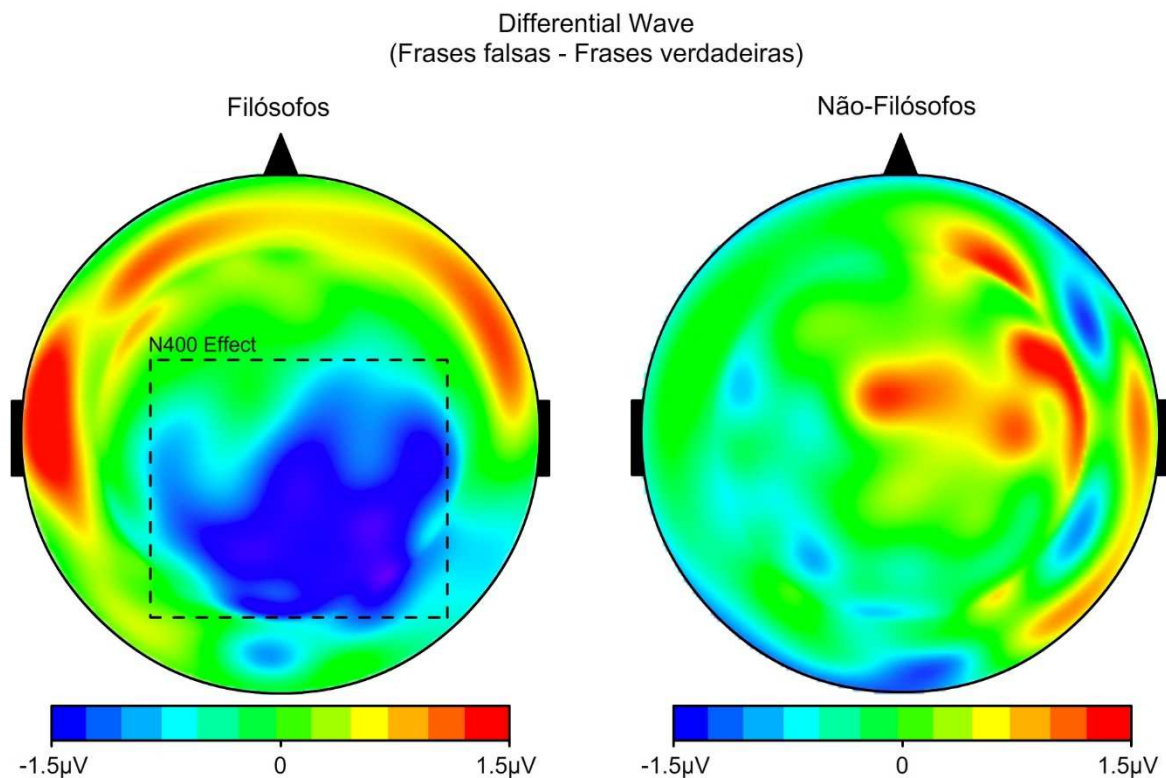


Figura 29 – Mapa topográfico dos *N400 Effects* dos filósofos e não-filósofos aos 400 ms.

Descritivos

N400 EFFECT (MEAN CORRECTION)

	N	Média	Desvio padrão	Modelo padrão	Intervalo de confiança de 95% para média		Mínimo	Máximo
					Limite inferior	Limite superior		
F	26	-,3478	,26890	,05274	-,4564	-,2392	-1,04	,15
NF	9	,0811	,48196	,16065	-,2894	,4516	-,38	1,18
Total	35	-,2375	,37948	,06414	-,3679	-,1072	-1,04	1,18

ANOVA

N400 EFFECT (MEAN CORRECTION)

	Soma dos Quadrados	df	Quadrado Médio	F	Sig.
Entre Grupos	1,230	1	1,230	11,073	,002
Nos grupos	3,666	33	,111		
Total	4,896	34			

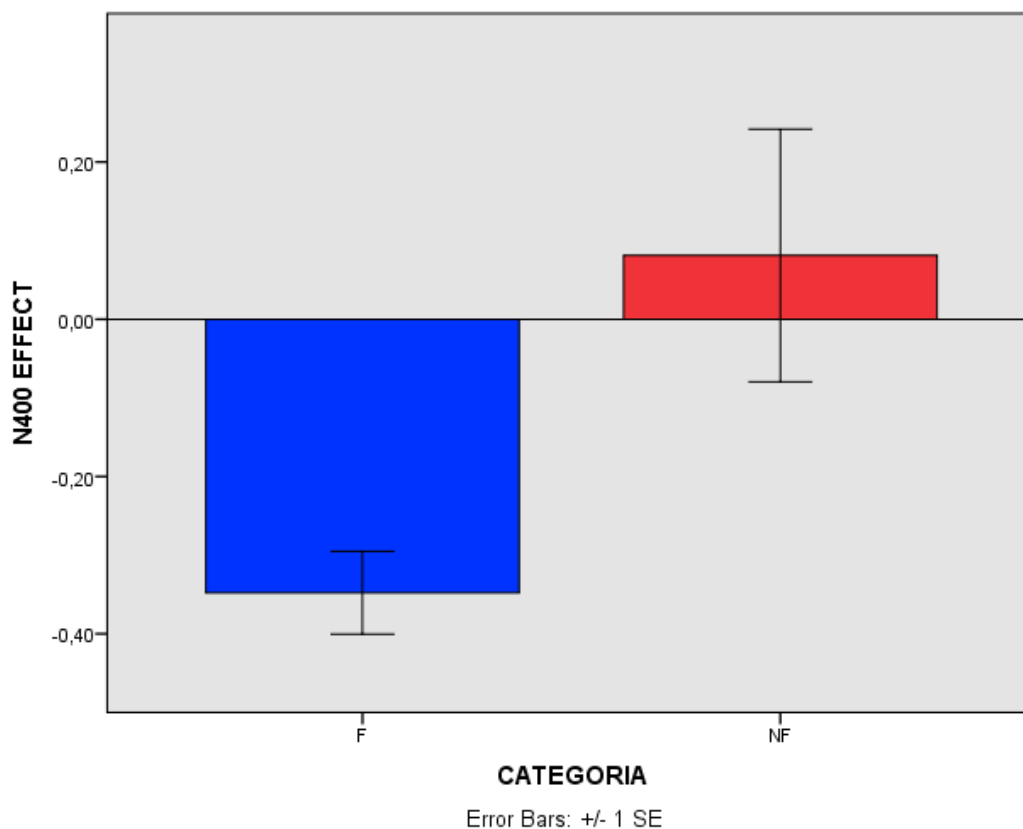


Figura 30 – Tabelas ANOVA, médias e erros padrão dos N400 Effects dos filósofos e não filósofos.

N400 EFFECT (MEAN CORRECTION)

	N	Média	Desvio padrão	Modelo padrão	Intervalo de confiança de 95% para média		Mínimo	Máximo
					Limite inferior	Limite superior		
					NF	9		
C	12	-,2323	,19234	,05552	-,3545	-,1101	-,51	,15
Total	21	-,0980	,37219	,08122	-,2674	,0714	-,51	1,18

ANOVA

N400 EFFECT (MEAN CORRECTION)

	Soma dos Quadrados	df	Quadrado Médio	F	Sig.
Entre Grupos	,505	1	,505	4,238	,054
Nos grupos	2,265	19	,119		
Total	2,770	20			

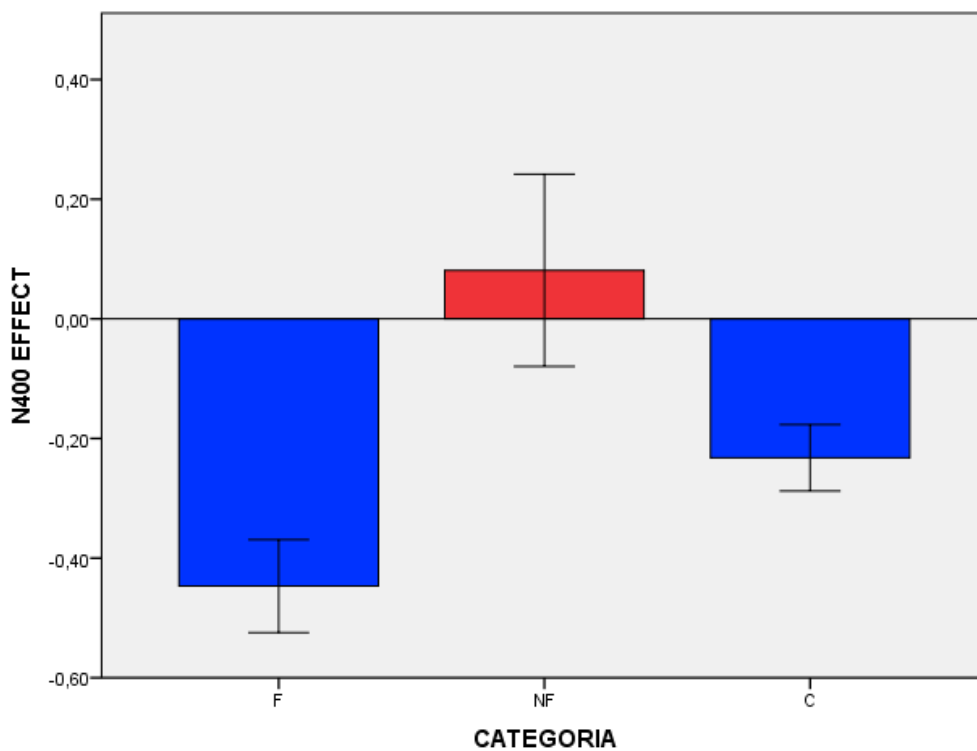
N400 EFFECT (MEAN CORRECTION)

	N	Média	Desvio padrão	Modelo padrão	Intervalo de confiança de 95% para média		Mínimo	Máximo
					Limite inferior	Limite superior		
					F	14		
C	12	-,2323	,19234	,05552	-,3545	-,1101	-,51	,15
Total	26	-,3478	,26890	,05274	-,4564	-,2392	-1,04	,15

ANOVA

N400 EFFECT (MEAN CORRECTION)

	Soma dos Quadrados	df	Quadrado Médio	F	Sig.
Entre Grupos	,297	1	,297	4,725	,040
Nos grupos	1,510	24	,063		
Total	1,808	25			



Error Bars: +/- 1 SE

Figura 31 – Tabelas ANOVA, médias e erros padrão dos *N400 Effects* dos filósofos com maior tempo de estudo, não-filósofos e calouros.

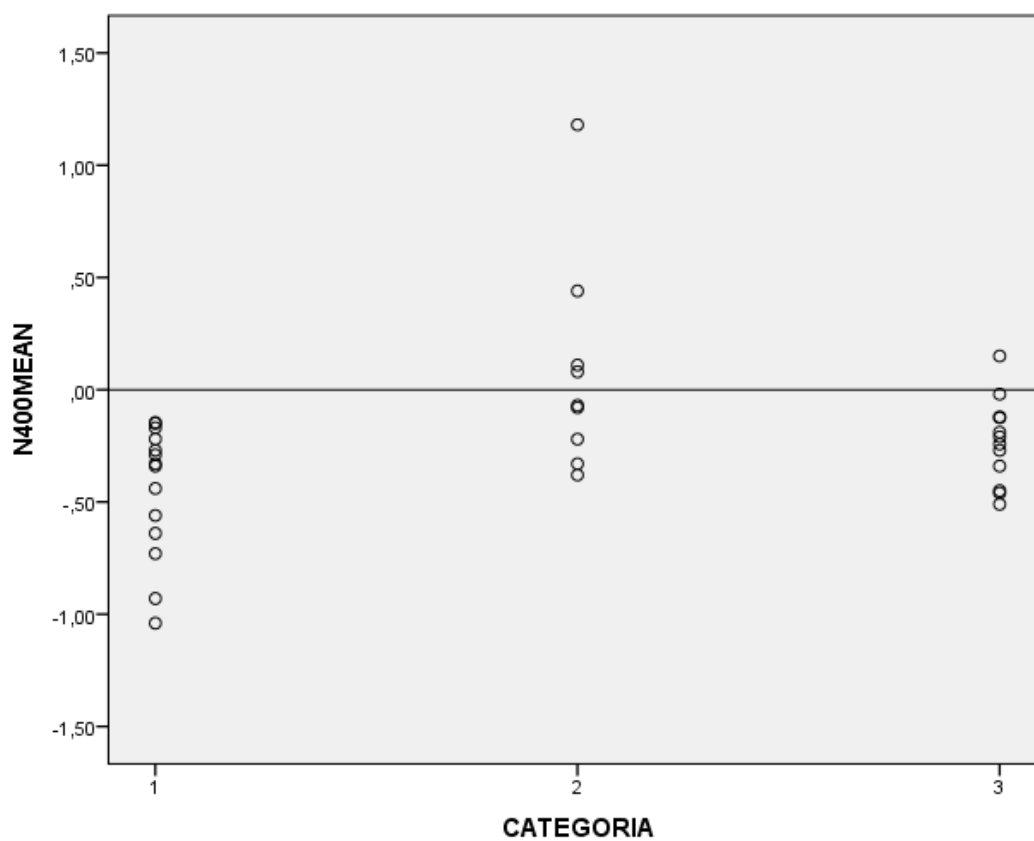


Figura 32 – Valores das médias individuais dos filósofos, não-filósofos e calouros.

5.3.2 Recognition Potential

Diferentemente do N400, o *Recognition Potential* (RP) apresenta uma distribuição mais discreta, costumando se manifestar nos eletrodos próximos à *Visual Word-Form Area* (VWFA), situada no lobo temporal esquerdo. Foram selecionados os eletrodos 50, 51, 58, 59 e 65 (localização na Figura 33), pela proximidade com a suposta origem do componente. Utilizou-se as mesmas *differential waves* geradas para analisar o componente N400. Calculou-se, para análise estatística, a média da amplitude do *RP Effect* de cada participante no intervalo de 100 a 250 milissegundos (em todos os eletrodos). Os valores também foram corrigidos utilizando *Mean Correction*. Considerou-se como variável dependente a média dos valores dos 5 eletrodos selecionados.

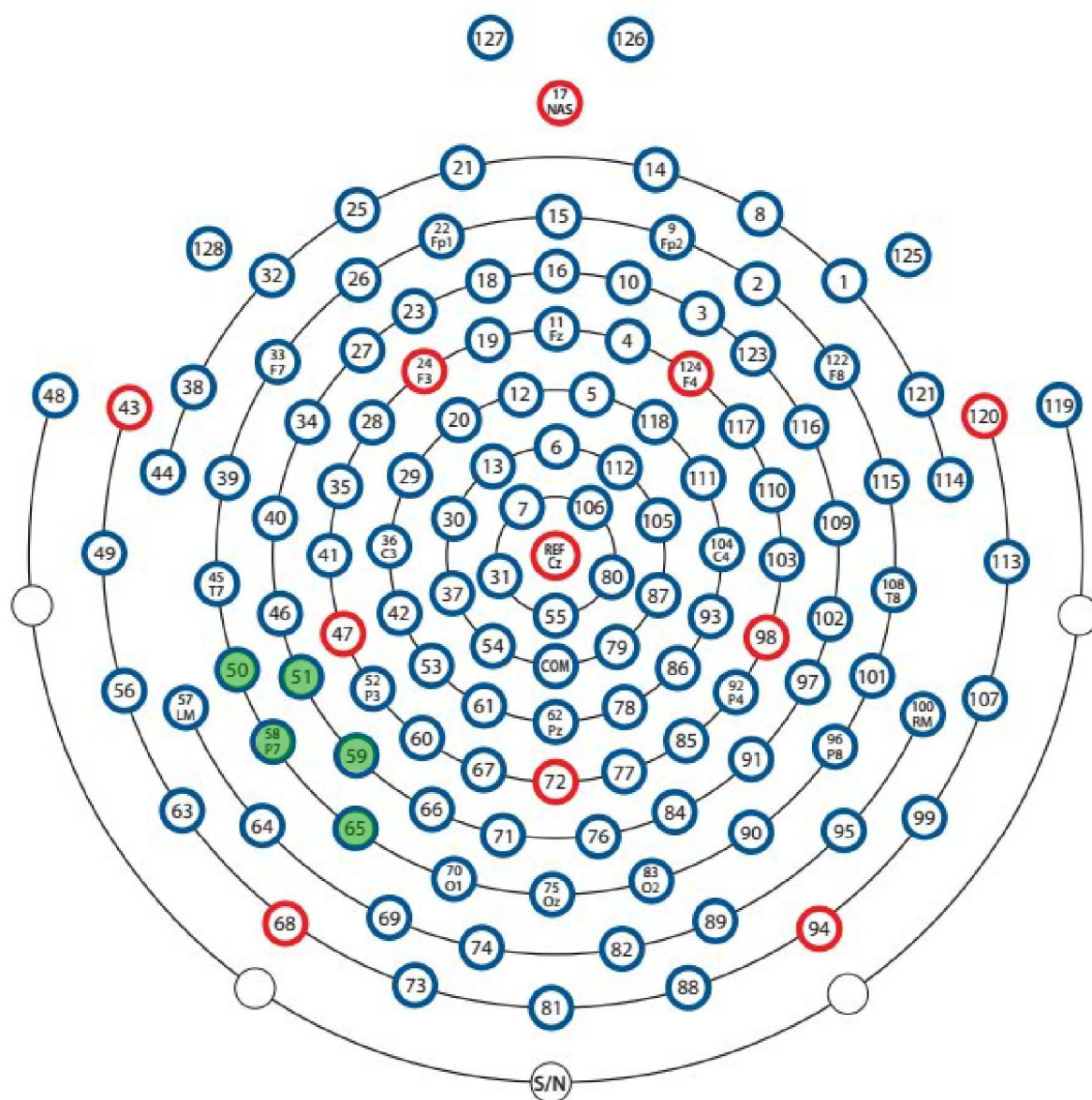


Figura 33 - Eletrodos selecionados para análise estatística do Recognition Potential.

A média do *RP Effect* dos 5 eletrodos próximos à VWFA no grupo dos Filósofos (N=26) foi de $0,32 \mu\text{V}$, enquanto a dos Não-Filósofos (N=9) foi de $-0,2695 \mu\text{V}$. Utilizando One-Way ANOVA, verificou-se que a diferença entre as médias foi significativa [$F(1,33) = 7.504, p=0.01$], conforme a Figura 34.

Descritivos

RP EFFECT

	N	Média	Desvio padrão	Modelo padrão	Intervalo de confiança de 95% para média		Mínimo	Máximo
					Limite inferior	Limite superior		
F	26	,3200	,60193	,11805	,0769	,5631	-,83	1,86
NF	9	-,2695	,38096	,12699	-,5624	,0233	-,90	,32
Total	35	,1684	,60738	,10267	-,0402	,3771	-,90	1,86

ANOVA

RP EFFECT

	Soma dos Quadrados	df	Quadrado Médio	F	Sig.
Entre Grupos	2,324	1	2,324	7,504	,010
Nos grupos	10,219	33	,310		
Total	12,543	34			

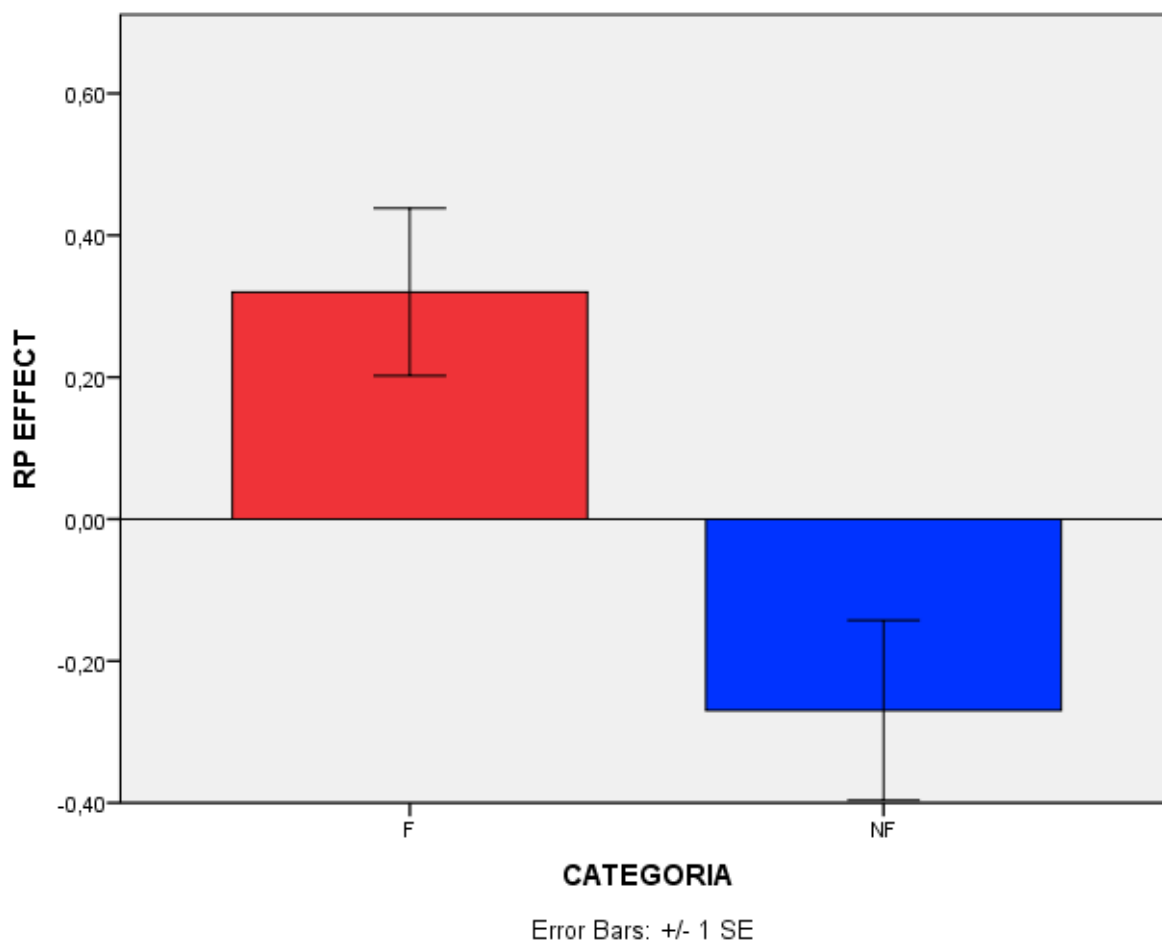


Figura 34 – Tabelas ANOVA, médias e erros padrão dos *RP Effects* dos filósofos e não-filósofos.

5.3.3 Correlação entre o N400 e o *Recognition Potential*

A mesma hipótese (de que conhecimentos específicos geram predições cujas violações/satisfações modulam respostas eletrofisiológicas) prediz dois resultados inversos: *N400 Effects* mais negativos em filósofos do que em não-filósofos e *RP Effects* mais positivos em filósofos do que em não filósofos. As duas predições foram observadas e atingiram níveis de significância satisfatórios, conforme exposto nas seções anteriores. Os dados obtidos, entretanto, permitiram a constatação de algo que não havia sido explicitamente previsto, porém está de acordo com a hipótese defendida: uma correlação negativa fraca significativa entre os *N400 Effects* e os *RP Effects*, de modo que quanto mais negativo o *N400 Effect*, mais positivo é o *RP Effect* ($r = -0.393$, $n = 35$, $p = 0.019$), de acordo com a Figura 35:

Correlações

		N400 Effect	RP effect
N400MEAN	Correlação de Pearson	1	-,393*
	Sig. (2 extremidades)		,019
	N	35	35
RP5MEAN	Correlação de Pearson	-,393*	1
	Sig. (2 extremidades)	,019	
	N	35	35

*. A correlação é significativa no nível 0,05 (2 extremidades).

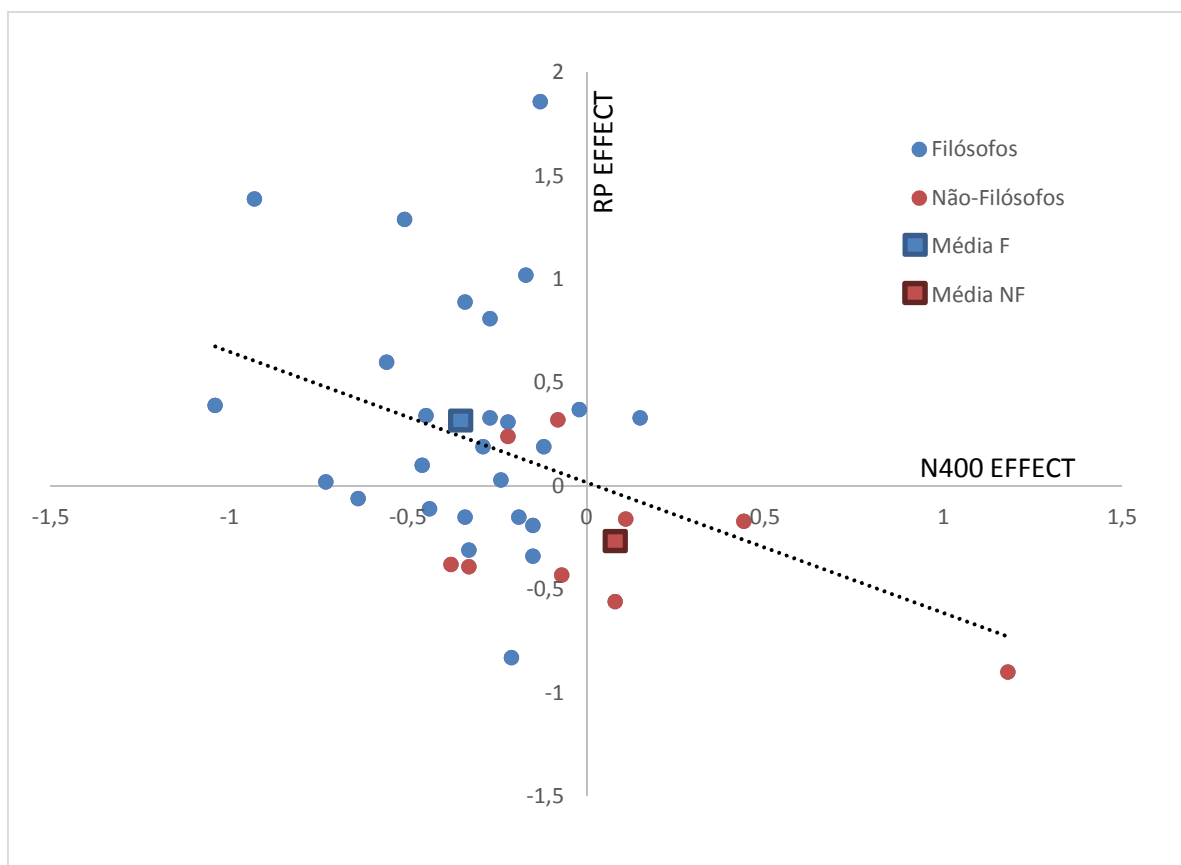


Figura 35 – Correlação de Pearson entre os *RP Effects* e os *N400 Effects* de todos os grupos.

5.4 DIREÇÕES FUTURAS: RECRUTAMENTO DE CRENÇAS EM CONTEXTOS NÃO-PREDITIVOS

As crenças do experimento realizado eram recrutadas, nos filósofos, antes da aparição da palavra final e são utilizadas para predizer diferentes características do estímulo crítico (i.e. a palavra, o significado da palavra e a aparência física do estímulo), de modo que a detecção da veracidade/falsidade da sentença depende de uma confirmação do que fora previsto pela crença, uma vez que o contexto é preditivo. A crença relevante não precisa ser acessada após a leitura da palavra final para que se detecte o valor de verdade da sentença. Ao serem apresentadas sentenças inseridas em contextos não-preditivos será necessário acessar as crenças relevantes apenas após a aparição da palavra final a fim de determinar seus valores de verdade, conforme exposto na Figura 36. A leitura de uma sentença incompleta como "Hanesh é a capital da..." não permite ao participante acessar uma crença específica acerca do país que tem "Hanesh" como capital (pois "Hanesh" é uma palavra inventada) para

predizer o estímulo final. Para que seja detectado que a palavra "França", ao aparecer no final da frase, resultou numa proposição falsa, será necessário acessar a crença relevante que a contraria (i.e. de que apenas Paris é a capital da França). Esse passo inferencial adicional, não encontrado no experimento realizado, pode refletir em diferenças na latência do componente N400. Uma maior demora na deflexão do N400 pode vir a informar o quão rápido uma crença fora de contexto é acessada.

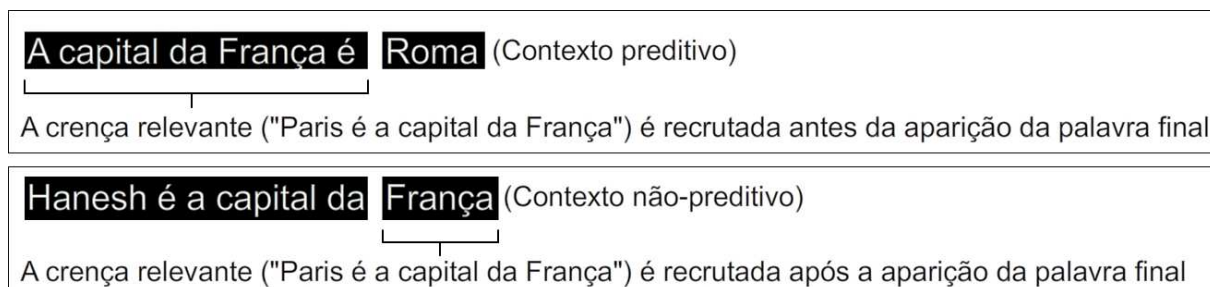


Figura 36 – Diferenças entre contextos preditivos e não-preditivos.

Embora uma tarefa de leitura utilizando contextos não-preditivos pareça não envolver um processo de predição de inputs (pois se afigura inicialmente como um processo passivo em que o significado da palavra final é integrado ao contexto sentencial), há a possibilidade de recrutarmos um corpo de crenças de background que limite as possibilidades. Por exemplo, se a frase incompleta "Platão escreveu o diálogo" for apresentada a um 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).

CONSIDERAÇÕES FINAIS

A lacuna explicativa existente entre o vocabulário mentalista de primeira pessoa (privado e subjetivo) e o discurso científico acerca de dados públicos e objetivamente analisáveis (i.e. estímulos sensoriais, atividade cerebral e/ou comportamentos) é um problema epistemológico de difícil resolução. Não obstante a facilidade com que atribuímos crenças, intenções e desejos a outras pessoas e utilizamos essas entidades para prever comportamentos com considerável sucesso, há fortes razões para suspeitarmos de que essa prática é insuficiente enquanto *explicação* das ações humanas. Assim como chimpanzés são capazes de formar conceitos relacionados a regularidades comportamentais e os utilizar para antecipar as ações futuras de seus congêneres (consistindo em “explicações” tão defeituosas quanto a generalização indutiva “o sol nasceu esta manhã *porque* nasceu em todas as manhãs anteriores”), crianças de apenas 6 meses seriam capazes de considerar tanto a perspectiva epistêmica de um agente (como a posse de uma crença falsa) quanto o modo como ela é causada por estímulos e causa comportamentos (as leis da *folk psychology*). Os experimentos expostos no capítulo 3, que utilizaram diversas versões de *False Belief Tasks*, constituem evidência de que a capacidade conjunta de atribuir estados intencionais e considerar o modo como essas entidades interagem com estímulos e comportamentos possui uma base inata de domínio específico passível de dano seletivo. Argumentou-se que do fato de seres humanos neurotípicos possuírem um sistema cognitivo destinado à atribuição de estados mentais possuindo forte semelhança com uma teoria científica não se segue que a teoria utilizada seja verdadeira, ainda que permita grande sucesso preditivo (pois as entidades da ToM podem ser meras simplificações ou ficções das quais nossa espécie faz uso para interagir com entidades tão imprevisíveis como outros seres humanos, não havendo correspondência com as causas reais — ainda desconhecidas pela ciência — das ações humanas).

No capítulo 4 foram apresentadas teorias filosóficas desenvolvidas na segunda metade do século XX que procuraram se afastar de resquícios dualistas que consideravam a mente (e o conjunto das “entidades mentais” próprias de seu repertório) como possuindo um estatuto ontológico radicalmente distinto das

entidades encontradas no mundo físico. Behavioristas como Gilbert Ryle procuraram, utilizando ferramentas conceituais desenvolvidas por filósofos da linguagem ordinária, analisar os significados de termos como “crenças”, “desejos” e “intenções” de modo que em seus respectivos *definiens* figurassem apenas entidades ou eventos objetivos e públicos. Uma crença, desta forma, traduzir-se-ia em uma classe finita de disposições comportamentais “*input-output*” possuindo a forma lógica de condicionais “se, então”. Funcionalistas como Lewis, Armstrong e Putnam, partindo do mesmo *desiderata* ambicionado pelos behavioristas (i.e. eliminar entidades de ontologia duvidosa do vocabulário mentalista), ao contrário destes, formularam teorias que admitiam referência a estados inobserváveis. Sendo estados mentais definidos pela função que realizam (i.e. quais estímulos os causam, como interagem com outros estados mentais e que comportamentos tipicamente produzem), a proposta funcionalista prescinde de um substrato material específico (permitindo múltipla instanciação). Embora na definição do funcionalista o “estado mental” também seja constituído apenas por componentes objetivos, uma estrutura lógica semelhante à empregada por Bertrand Russell em sua teoria das descrições definidas permite ao funcionalista se referir a uma entidade inobservável por ela supostamente apresentar propriedades relacionais com eventos observáveis (e.g. o estado mental da dor pode ser definido como “o x, tal que x é causado por inputs que sinalizam danos ao organismo & x causa o output comportamental de evitar a causa do dano”). Ainda que não saibamos prontamente quais as entidades materiais que estão sendo causadas por determinados inputs e por sua vez causando comportamentos, podemos saber que *há algo* que está a ocupar este papel causal.

Paul Churchland, principal defensor do materialismo eliminativista, desconsidera a pretensão dos demais filósofos que querem reduzir a fala acerca do mental a um vocabulário científico. Para Churchland, o destino de termos como “crenças” e “desejos” não é o de serem *traduzidos* a outras coisas, mas de serem completamente *eliminados* por uma neurobiologia completa. De fato, há uma grande lacuna entre a linguagem intencional, como entre a atitude proposicional “Édipo acredita que Jocasta será uma excelente esposa” e aquilo que ocorre no cérebro de Édipo e que está causalmente relacionado com o comportamento de se casar com a própria mãe. A atribuição do estado mental exemplificado, por ser uma atitude proposicional, envolve uma relação específica entre um sujeito (i.e. Édipo) e um estado de coisas por ele representado (e este, por sua vez, decompõe-se em

conceitos e representações particulares aspectuais/intensionais, resultando em opacidade referencial). Apesar da linguagem psicológica consistir em simplificações grosseiras possuindo significados vagos, ela abarca análises de grande complexidade e que parecem intratáveis por uma teoria tão-somente neurobiológica.

A presente pesquisa, estando situada no contexto filosófico supracitado, procurou acrescentar novos dados empíricos à discussão de modo a atualizar teorias e ajudar a decidir entre posições rivais. Indo à procura de crenças em cérebros pelo modo como supostamente são recrutadas em processos preditivos subjacentes a uma tarefa de leitura (utilizando, portanto, uma postura de teor funcionalista), encontrou-se, no experimento exposto no capítulo 5, dados que “contêm informação” acerca da existência/inexistência de crenças particulares. Verificou-se que indivíduos aos quais nossa *folk psychology* atribui determinadas crenças apresentam respostas eletrofisiológicas significativamente diferentes de indivíduos aos quais não as atribuímos. Apesar da atribuição de crenças e os potenciais pós-sinápticos mensurados pertencerem a lados opostos da “lacuna entre o mental e o cerebral”, foi possível encontrar correlações entre os “dois mundos”. Diferenças existentes em um lado da lacuna interferem naquilo que aparece do outro lado, embora a tão almejada conexão ou ponte permaneça ela própria fora de nosso alcance epistêmico. A conjunção constante entre o “possuir um conjunto de crenças” e o “apresentar *N400 Effects* e *RP Effects salientes*,” entretanto, permite-nos, mesmo na ignorância do elo misterioso, formular leis fracas e probabilísticas que nos possibilitem inferir a obtenção de *N400 effects* e *RP effects salientes* da posse de crenças específicas (seja na forma de uma implicação material *ceteris paribus* “CP, para todo o x, se x possui o conjunto de crenças C & x realiza o experimento E, então x apresenta *N400 effects* e *RP effects salientes*” ou na forma probabilística “a probabilidade de x apresentar *N400 effects* e *RP effects salientes*, dado que x possui o conjunto de crenças C & x realiza o experimento E é igual a 0.85”). Ainda que seja problemática a utilização de uma inferência reversa (inferir um estado mental de um estado cerebral), por ter a forma lógica de uma afirmação do conseqüente, experimentos futuros podem vir a demonstrar a viabilidade de uma equivalência material (i.e. quem possui as crenças apresenta *N400 effects* e *RP effects salientes* no experimento e quem apresenta *N400 effects* e *RP effects salientes* no experimento possui as crenças). Uma possível maneira de chegar a tal resultado seria categorizar cegamente os participantes apenas por meio dos sinais, verificando a porcentagem de categorizações corretas.

Estando firmado um método filosófico-experimental confiável que indique o recrutamento de crenças por cérebros, faz-se possível avançar na investigação filosófica de uma maneira interdita à abordagem filosófica tradicional. Pode-se, deste modo, acompanhar como cérebros, a partir de estímulos, processam formas, reconhecem palavras, acessam o significado das palavras, modelam estados de coisas e utilizam crenças prévias para predizer os inputs seguintes (cada nível predizendo e analisando aspectos distintos dos estímulos, por serem de domínio-específico). A utilização de um framework teórico como a do Predictive Brain, escolhido, no capítulo 2, como uma arquitetura neurocognitiva mais adequada para o atual experimento do que a teoria da mente modular, permitiu interpretar os resultados com base em previsões geradas por processos globais e erros de predição por eles modulados. Se a única diferença entre os dois grupos de participantes que apresentaram respostas eletrofisiológicas distintas era a conjecturada posse ou ausência de determinadas crenças, então a diferença medida deve ser interpretada como evidência de que *há algo* em quem crê que não se encontra em quem não crê. Dizer que este “x” misterioso que está modulando o sinal é o correlato neural da crença pode ser uma conclusão precipitada. “X” pode vir a revelar-se, com os avanços vindouros da presente técnica, muito distinto de nossa concepção tradicional de crença (e.g. podem ser conjuntos de informação na forma de “*probability-density functions*” sem qualquer semelhança com atitudes proposicionais). Ou talvez seja uma questão meramente conceitual redefinir “crença” de modo altamente técnico ou simplesmente descartá-la e introduzir um novo nome àquilo que há em nossas cabeças e que utilizamos para minimizar as contingências e promover uma melhor adaptação em nossa interação com o mundo.

REFERÊNCIAS

- APPERLY, Ian A.; ROBINSON, Elizabeth J. When can children handle referential opacity? Evidence for systematic variation in 5-and 6-year-old children's reasoning about beliefs and belief reports. **Journal of Experimental Child Psychology**, v. 85, n. 4, p. 297-311, 2003.
- APPERLY, Ian. **Mindreaders: the cognitive basis of " theory of mind"**. Psychology Press, 2010.
- ARMSTRONG, David M. **The Causal Theory of the Mind**. In Philosophy of Mind: Classical and Contemporary Readings, ed. By David Chalmers. New York: Oxford University Press, 2002, 80-87.
- AUSTIN, John. **Other Minds**. Aristotelian Society Supplementary. Volume 20:148-87. 1946.
- BARON-COHEN, A, LESLIE, A, & FRITH, U. **Does the autistic child have a "theory of mind"?** Cognition 21: 37-46, 1985.
- BLOCK, Ned J. Functionalism. **Studies in Logic and the Foundations of Mathematics**, v. 104, p. 519-539, 1982.
- BLOCK, Ned. **Troubles with functionalism**. Minnesota Studies in the Philosophy of Science 9: 261-325, 1978.
- BOAZ, Timothy L. et al. Detection of guilty knowledge with event-related potentials. **Journal of Applied Psychology**, v. 76, n. 6, p. 788, 1991.
- BOWMAN, Howard et al. Attention is more than prediction precision. **Behavioral and Brain Sciences**, v. 36, n. 03, p. 206-208, 2013.
- BRENTANO, Franz. **Psychology from an empirical standpoint**. Routledge, 2014.
- CALL, J., TOMASELLO, M. **Does the chimpanzee have a Theory of Mind? 30 years later**. Trends in Cognitive Sciences, May: 12(5):187-92, 2008.
- CHURCHLAND, Patricia S. **Can Neurobiology Teach us Something About Consciousness?** Proceedings and Addresses of the American Philosophical Association 67 (4): 23-40, 1994.
- CHURCHLAND, Paul M. **Eliminative Materialism and the Propositional Attitudes**. Journal of Philosophy 78: 67-90, 1981.
- CHURCHLAND, Paul M. **Folk psychology and the explanation of human behavior**. Proceedings of the Aristotelian Society 62:209-21, 1988.
- CLARK, A, CHALMERS, D. **The Extended Mind**. Analysis 58 (1): 7-19, 1998.

CLARK, Andy. **From Folk Psychology to Naïve Psychology**. *Cognitive Science* 11: 139-54, 1987

CLARK, Andy. Whatever next? Predictive brains, situated agents, and the future of cognitive science. **Behavioral and Brain Sciences**, v. 36, n. 03, p. 181-204, 2013.

CLARK, Andy. **Surfing uncertainty: Prediction, action, and the embodied mind**. Oxford University Press, 2015.

CLEMENTS, W. A., PERNER, J. **Implicit Understanding of Belief**. *Cognitive Development*, 9, 377-395, 1994.

CRANE, Tim. **Elements of Mind: An Introduction to the Philosophy of Mind**. New York: Oxford University Press, 2001.

DAVIDSON, Donald. **Actions, Reasons, and Causes**. In *Essays on Actions and Events*. Oxford: Clarendon Press, 2nd edn, 2001.

DAVIDSON, Donald. **How is Weakness of the Will Possible?** In *Essays on Actions and Events*. Oxford: Clarendon Press, 2nd edn, 2001.

DAVIDSON, Donald. **Mental Events** In *Essays on Actions and Events*. Oxford: Clarendon Press, 2nd edn, 2001.

DAVIDSON, Donald. **Paradoxos da Irrracionalidade**. Trad. Marco Antonio Frangiotti. Disponível em: <<http://www.cfh.ufsc.br/~wfil/davidson2.htm>>. Acesso em: 7 set. 2014.

DAVIDSON, Donald. **Problems in the Explanation of Action**. In *Problems of Rationality*. Oxford: Clarendon Press, 2004.

DAVIDSON, Donald. **Psychology as Philosophy**. In *Essays on Actions and Events*. Oxford: Clarendon Press, 2nd edn, 2001.

DENNETT, Daniel C. **Brainstorms: Ensaios Filosóficos sobre a Mente e a Psicologia**. São Paulo: Editora UNESP, 2006.

DENNETT, Daniel C. **Intentional Systems Theory**. In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *The Oxford Handbook of Philosophy of Mind*. Oup Oxford, 2011.

DENNETT, Daniel C. **Intentionality**. In Richard L. Gregory (ed.), *Oxford Companion to the Mind*. Oxford University Press. 139-143, 2001.

DENNETT, Daniel C. **Real Patterns**. *Journal of Philosophy* 88 (1):27-51, 1991.

DENNETT, Daniel C. **Tipos de mentes: rumo a uma compreensão da consciência**. Trad. Alexandre Tort. Rio de Janeiro: Rocco, 1997.

DENNETT, Daniel C. **The Part of Cognitive Science That Is Philosophy**. *Topics in Cognitive Science* 1 (2):231-236, 2009.

DENNETT, Daniel C. **True Believers: The Intentional Strategy and Why It Works**. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. By David Chalmers. New York: Oxford University Press, 2002, 556-568.

DESCARTES, René. **Meditações**. Tradução: Fausto Castilho. Campinas: IFCH/UNICAMP, 1993.

DESCARTES, René. **Discurso do Método**. Tradução: Ciro Mioranza. São Paulo, SP: Editora Escala, 2006.

DIKKER, Suzanne et al. Early occipital sensitivity to syntactic category is based on form typicality. **Psychological Science**, v. 21, n. 5, p. 629-634, 2010.

DIKKER, Suzanne; PYLKKANEN, Liina. Before the N400: Effects of lexical–semantic violations in visual cortex. **Brain and Language**, v. 118, n. 1, p. 23-28, 2011.

EVANS, J.St.B.T. **In two minds: dual-process accounts of reasoning**. *Trends in Cognitive Sciences*, Oct;7(10):454-9, 2003.

FEDERMEIER, Kara D.; KUTAS, Marta. 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. **A Theory of the Child's Theory of Mind**. *Cognition* 44: 283-296, 1992.

FODOR, Jerry A. **Propositional Attitudes**. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. By David Chalmers. New York: Oxford University Press, 2002, 542-55.

FODOR, Jerry A. **The Modularity of Mind**. Cambridge, MA: MIT Press, 1983

FRISTON, Karl. A theory of cortical responses. **Philosophical transactions of the Royal Society B: Biological sciences**, v. 360, n. 1456, p. 815-836, 2005

GANIS, Giorgio; KUTAS, Marta; SERENO, Martin I. The search for “common sense”: an electrophysiological study of the comprehension of words and pictures in reading. **Journal of Cognitive Neuroscience**, v. 8, n. 2, p. 89-106, 1996.

GIERE, Ronald N. **How models are used to represent reality**. *Philosophy of Science* 71 (5): 742-752, 2004.

GODFREY-SMITH, Peter. **Folk Psychology as a Model**. *Philosophers's Imprint* 5 (6): 1-16, 2005.

GODFREY-SMITH, Peter. **On Folk Psychology and Mental Representation**. In *Representation in Mind*, ed. Hugh Clapin. Elsevier, 2004.

GOLD, Natalie. **What is a Character Trait?** Disponível em: <<http://www.unc.edu/~gsmunc/EthicsGroup/WhatIsACharacterTrait.pdf>>. Acesso em: 7 set. 2014.

GOLDMAN, Alvin I. **Theory of Mind**. In Oxford Handbook of Philosophy and Cognitive Science. Edited by Eric Margolis, Richard Samuels, and Stephen Stich, Oxford University Press, 402-424, 2012.

GOPNIK, A & WELLMAN, H. M. **Why the child's theory of mind really is a theory**. Mind and Language 7 (1-2):145-71, 1992.

GOPNIK, Alison. **Theories and Modules; Creation Myths, Developmental Realities, and Neurath's Boat**. In Theories of Theories of Mind. Peter Carruthers & Peter K. Smith (eds.), Cambridge University Press, 169-183, 1996.

GORDON, R. **Folk Psychology as Simulation**. Mind and Language, 1: 158–17, 1986.

GOUVEA, Ana C. et al. The linguistic processes underlying the P600. **Language and Cognitive Processes**, v. 25, n. 2, p. 149-188, 2010.

GRIFFIN, Richard., BARON-COHEN, Simon. **The Intentional Stance: Developmental and Neurocognitive Perspectives**. In Daniel Dennett: Contemporary Philosophy in Focus. A. Brook & D. Ross (Eds.), UK: Cambridge University Press, pp. 83-116, 2002.

HAGOORT, Peter et al. Integration of word meaning and world knowledge in language comprehension. **science**, v. 304, n. 5669, p. 438-441, 2004.

HEMPEL, Carl, G. **Explanation and Prediction by Covering Laws**. In The Philosophy of Carl G. Hempel: Studies in Science, Explanation, and Rationality, ed. James H. Fetzer. New York: Oxford University Press, 2001.

HEMPEL, Carl, G. **Explanation in Science and in History**. In The Philosophy of Carl G. Hempel: Studies in Science, Explanation, and Rationality, ed. James H. Fetzer. New York: Oxford University Press, 2001.

HEMPEL, Carl, G. **Rational Action**. In The Philosophy of Carl G. Hempel: Studies in Science, Explanation, and Rationality, ed. James H. Fetzer. New York: Oxford University Press, 2001.

HOHWY, Jakob. **The predictive mind**. Oxford University Press, 2013.

HUME, D. **Investigação acerca do entendimento humano**. In: Os Pensadores. São Paulo, Nova Cultural, 1989.

HUSSERL, Edmund. **Logical Investigations**. Translated by JN Findlay. Humanities Press, 1970.

JACKSON, Frank. **What Mary Didn't Know**. Journal of Philosophy (83): 291-295, 1986.

JACOB, Pierre. **Intentionality**. Stanford Encyclopedia of Philosophy, 31 ago. 2010. Disponível em: <<http://plato.stanford.edu/entries/intentionality/>>. Acesso em: 7 set. 2014.

JOU, Graciela Inchausti de., SPERB, Tania Mara. **Teoria da Mente: diferentes abordagens**. *Psicologia: Reflexão e Crítica*, vol. 12, núm. 2, p. 0, 1999.

KNOBE, J, PRINZ, J. **Intuitions About Consciousness: Experimental Studies**. *Phenomenology and the Cognitive Sciences* 7 (1): 67-83, 2008.

KUTAS, Marta et al. Reading senseless sentences: Brain potentials reflect semantic incongruity. **Science**, v. 207, n. 4427, p. 203-205, 1980.

KUTAS, Marta; FEDERMEIER, Kara 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, 2011.

KUTAS, Marta; VAN PETTEN, C. Psycholinguistics electrified. **Handbook of psycholinguistics**, p. 83-143, 1994.

LEVINE, Joseph. Materialism and qualia: The explanatory gap. **Pacific philosophical quarterly**, v. 64, n. 4, p. 354-361, 1983.

LEWIS, David. **An Argument for the Identity Theory**. *Journal of Philosophy* 63 (2): 17-25, 1970.

LEWIS, David. **How to Define Theoretical Terms**. *Journal of Philosophy* 67 (13): 427-446, 1970

LEWIS, David. **Mad pain and Martian pain**. In *Readings in the Philosophy of Psychology*, ed. by Ned Block. Cambridge: Harvard University Press. Vol. 1, pp. 216-222, 1980.

LEWIS, David. **Psychophysical and Theoretical Identifications**. *Australasian Journal of Philosophy* 50: 249-58, 1972.

LOCKE, J. **Ensaio acerca do entendimento humano**. 5. ed. Tradução de Anair Aiex. São Paulo: Nova Cultural, 1991.

LOW, J., PERNER, J. **Implicit and explicit theory of mind: State of the art**. *British Journal of Developmental Psychology*, 30, 1-13, 2012.

LUCK, Steven J. **An introduction to the event-related potential technique**. MIT press, 2014.

LUPYAN, Gary. Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. **Review of philosophy and psychology**, v. 6, n. 4, p. 547-569, 2015.

MARRAFA, Massimo. **Theory of Mind**. Internet Encyclopedia of Philosophy, 11 nov. 2011. Disponível em: <<http://www.iep.utm.edu/theomind/>>. Acesso em: 7 set. 2014.

MARTIN-LOECHES, Manuel et al. Electrophysiological evidence of an early effect of sentence context in reading. **Biological Psychology**, v. 65, p. 265-280, 2004.

MARTÍN-LOECHES, Manuel et al. Functional differences in the semantic processing of concrete and abstract words. **Neuropsychologia**, v. 39, n. 10, p. 1086-1096, 2001.

MATA, R. **How the central system works? It uses fast and frugal heuristics.** Proceedings of the 26th Cognitive Science Society Meeting, Chicago, Illinois, U.S.A, 2004. Disponível em: <<http://www.cogsci.northwestern.edu/cogsci2004/ma/ma129.pdf>>. Acesso em: 7 set. 2014.

MCGINN, Colin. **Can We Solve the Mind-Body Problem?** *Mind*, 98, 349-66, 1989.

NAGEL, Thomas. What is It Like to Be a Bat? **Philosophical Review** 83 (October):435-50, 1974.

ONISH, K, BAILLARGEON, R. Do 15-Month-Old infants Understand False Beliefs? **Science** 308, 255-257, 2005.

PATEL, Aniruddh D. et al. Processing syntactic relations in language and music: An event-related potential study. **Journal of cognitive neuroscience**, v. 10, n. 6, p. 717-733, 1998.

PATEL, Aniruddh D. et al. Processing syntactic relations in language and music: An event-related potential study. **Journal of cognitive neuroscience**, v. 10, n. 6, p. 717-733, 1998.

PERNER, J., RUFFMAN, T. **Infant's insight into the mind: How deep?** *Science*, 308, 214-216, 2005.

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

PINKER, Steven. **Como a Mente Funciona.** Tradução: Laura Teixeira Motta. São Paulo: Companhia das Letras, 1998.

POVINELLI, D. J., VONK, J. **Chimpanzee minds: suspiciously human?** *Trends in Cognitive Sciences*, Vol. 7, No. 4, 157-160, 2003.

PREMACK, D., & WOODRUFF, G. **Does the chimpanzee have a theory of mind?** *Behavioural and Brain Sciences*, 1, 515-526, 1978.

PRINZ, Jesse J. **Furnishing the Mind: Concepts and Their Perceptual Basis.** London: MIT Press, 2002.

PRINZ, Jesse J. **Is the Mind Really Modular?** *Contemporary debates in cognitive science*, Stainton, Robert J. (Ed). Malden: Blackwell Publishing, pp. 22-36, 2006.

PRINZ, Jesse J. **The Return of Concept Empiricism.** In *Categorization and Cognitive Science*, ed. By H. Cohen & C. Lefebvre. Elsevier, 2005.

PUTNAM, Hilary. **Brains and Behavior**. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. By David Chalmers. New York: Oxford University Press, 2002, 45-54.

PUTNAM, Hilary. **The Nature of Mental States**. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. By David Chalmers. New York: Oxford University Press, 2002, 73-79.

PYLYSHYN, Zenon. Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. **Behavioral and brain sciences**, v. 22, n. 03, p. 341-365, 1999.

QUINE, Willard van Orman. **Reference and Modality**. In *From a logical point of view*. Cambridge MA: Harvard University Press, 1953.

ROBBINS, Philip. **Modularity of Mind**. *Stanford Encyclopedia of Philosophy*, 1 abr. 2009. Disponível em: <http://plato.stanford.edu/entries/modularity-mind/>>. Acesso em: 7 set. 2014.

RUFFMAN, T., GARNHAM, W., IMPORT, A. & CONOLLY, D. **Does Eye Gaze Indicate Knowledge of False Belief? Charting Transitions in Knowledge**. *Journal of Experimental Child Psychology*, 80, 201-224, 2001.

RUFFMAN, T., PERNER, J. **Do infants really understand false belief?** *Trends in Cognitive Sciences*, Vol. 9, No. 10, 462-463, 2005.

RUSSELL, Bertrand. **Analogy**. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. By David Chalmers. New York: Oxford University Press, 2002, 667-669.

RUSSELL, Bertrand. **On Denoting**. *Mind, New Series*, Vol. 14, No. 56. pp. 479-493, 1905.

RUSSELL, Bertrand. **Os Problemas da Filosofia**. Tradução: Antônio Sérgio. Coimbra: Arménio Amado, 1996.

RYLE, Gilbert. **The Concept of Mind**. Londres: Penguin Books, 2000.

SALMON, Wesley C. **Four decades of scientific explanation**. University of Pittsburgh press, 2006.

SAMUELS, Richard. **Is the Human Mind Massively Modular?** *Contemporary debates in cognitive science*, Stainton, Robert J. (Ed). Malden: Blackwell Publishing, pp. 37-56, 2006.

SAMUELS, Richard. **The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds**. In *In two minds: Dual processes and beyond*, ed. by Jonathan Evans and Keith Frankish. New York: Oxford University Press, 2009.

SAXE, R., & KANWISHER, N. **People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind"**. *NeuroImage*, 19, 1835-1842, 2003.

SAXE, R., & POWELL, L. **It's the thought that counts: Specific brain regions for one component of Theory of Mind**. *Psychological Science*, 17, 692-699, 2006.

SAXE, R., CAREY, S., & KANWISHER, N. **Understanding other minds: Linking developmental psychology and functional neuroimaging**. *Annual Review of Psychology*, 55, 87-124, 2004.

SAXE, Rebecca. **The new puzzle of Theory of Mind development**. In *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*. Ed: M Banaji & S Gelman, 2013.

SAXE, Rebecca. **The right temporo-parietal junction: a specific brain region for thinking about thoughts**. *Handbook of Theory of Mind*. 'Ed'. Alan Leslie & Tamsin German, 2010.

SAXE, Rebecca. **Theory of Mind (Neural Basis)**. In *Encyclopedia of Consciousness*, ed. William P. Banks, 2009.

SCHOLL, B., LESLIE, A. M. **Minds, modules and meta-analysis**. *Child Development*, 72, 697-701, 2001.

SCHOLL, B., LESLIE, A. M. **Modularity, development and 'theory of mind'**. *Mind and Language*, 14, 131-153, 1999.

SCHOLL, Brian J., LESLIE, Alan M. **Modularity, Development and 'Theory of Mind'**. *Mind & Language* Vol. 14 No.1, pp. 131-153, 1999.

SEARLE, John. **Intentionality**. New York: Oxford University Press, 1983.

SEARLE, John. **Mente, cérebro e ciência**. Tradução: Artur Morão. Lisboa: Edições 70, 1987.

SEARLE, John. **Mind, brains and programs**. *Behavioral and Brain Sciences*: 417-57, 1980.

SEARLE, John. **The Rediscovery of the Mind**. Cambridge: MIT Press, 1992.

SEGAL, Gabriel. **The modularity of theory of mind**. In *Theories of Theories of Mind*. Peter Carruthers & Peter K. Smith (eds.), Cambridge University Press. 141-157, 1996.

SELLARS, Wilfrid. **Empirismo e filosofia da mente** (Stein, S.I.A., Trad.). São Paulo: Vozes, 2008.

SETH, Anil K. The cybernetic Bayesian brain. In: **Open Mind**. Open MIND. Frankfurt am Main: MIND Group, 2014.

SOUTHGATE, V. & SENJU A. **Action anticipation through attribution of false belief by 2 year olds.** *Psychological Science* 18 (7), 587-592, 2007.

SOUTHGATE, V. & VERNETTI A. **Belief-based action prediction in preverbal infants.** *Cognition* 130, 1-10, 2014

STEIN, Sofia Inês Albornoz. **Social Minds.** *Trends in Psychiatry and Psychotherapy.* 34(4) -167-170, 2012.

STICH, S, NICOLS, S. **Folk Psychology.** In *The Blackwell Guide to Philosophy of Mind*, ed. By Stephen Stich & Ted A. Warfield. Oxford: Basil Blackwell, 2003, 235-55.

STICH, Stephen P., & RAVENSCROFT, Ian. **What is Folk Psychology?** *Cognition* 50: 447-68, 1994.

TEIXEIRA, João de Fernandes. **A mente segundo Dennett.** São Paulo: Perspectiva, 2008.

THORNHILL, Dianne E.; VAN PETTEN, Cyma. Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, v. 83, n. 3, p. 382-392, 2012.

WELLMAN, H. M., J., CROSS, D., WATSON, J. **Meta-Analysis of Theory-of-mind Development: The Truth about False Belief.** *Child Development*, Volume 72, Number 3, 655-684, 2001.

WITTGENSTEIN, Ludwig. **Investigações Filosóficas.** Tradução: José Carlos Bruni. São Paulo: Editora Nova Cultural, 1999.

WITTGENSTEIN, Ludwig. **Tractatus Lógico-philosophicus.** Tradução: José Arthur Giannotti. São Paulo: Companhia Editora Nacional, 1968.

VON HELMHOLTZ, Hermann; SOUTHALL, James Powell Cocke. **Treatise on physiological optics.** Courier Corporation, 2005.