

UNIVERSIDADE DE LISBOA
FACULDADE DE PSICOLOGIA



**RACIOCÍNIO BAYESIANO EM PROBLEMAS DE
BASE-RATE: PENSAMENTO HEURÍSTICO,
INTEGRAÇÃO COGNITIVA E APRENDIZAGEM**

Hugo Duarte Silva da Assunção

MESTRADO INTEGRADO EM PSICOLOGIA

(Secção de Cognição Social Aplicada)

2018

UNIVERSIDADE DE LISBOA
FACULDADE DE PSICOLOGIA



**RACIOCÍNIO BAYESIANO EM PROBLEMAS DE
BASE-RATE: PENSAMENTO HEURÍSTICO,
INTEGRAÇÃO COGNITIVA E APRENDIZAGEM**

Hugo Duarte Silva da Assunção

Dissertação orientada pelo Prof. Doutor Mário Ferreira

MESTRADO INTEGRADO EM PSICOLOGIA

(Secção de Cognição Social Aplicada)

2018

"A teoria da probabilidade, no fundo, não é mais do que o senso comum reduzido ao cálculo."

Pierre Simon de Laplace (1749 - 1827)

Resumo

Nos últimos 50 anos, a investigação tem mostrado a existência de erros sistemáticos em tarefas de julgamento e decisão em condições de incerteza. Os Problemas de Base-Rate (PBR) criados por Kahneman e Tversky (1973) pedem uma previsão probabilística e fornecem duas fontes de informação onde o participante se pode apoiar para dar a sua resposta. A primeira informação apresentada é o base-rate, uma informação quantitativa sobre o número de elementos que compõem dois grupos de pessoas. A segunda informação apresentada é uma descrição individualizante, tipicamente diagnóstica de um dos grupos. Kahneman e Tversky, comparam as respostas dos participantes em problemas onde o base-rate reforçava a diagnosticidade da descrição (problemas inclusivos) com as respostas dos participantes em problemas onde o base-rate se opunha à diagnosticidade da descrição (problemas exclusivos). Não obtendo diferenças ao nível da resposta mediana, estes autores concluíram que a manipulação dos base-rates tinha pouca influência nas probabilidades subjetivas e que os participantes baseavam o seu julgamento maioritariamente na descrição individualizante.

Embora muitos estudos tenham confirmado estes resultados, poucos se focaram em estudar a capacidade de integração das duas informações. Nesta tese apresentamos um paradigma de Problemas de Base-Rate onde a diagnosticidade da descrição é quantificada. Esta quantificação permite-nos comparar a resposta de cada participante a uma resposta bayesiana. Respostas próximas da resposta bayesiana revelam um maior grau de integração cognitiva. Respostas afastadas da resposta bayesiana revelam um maior grau de enviesamento. Utilizando este paradigma criámos cinco experiências com a participação de mais de mil participantes. Os resultados obtidos nestas experiências levaram-nos às seguintes conclusões:

- 1- É possível encontrar diferentes estratégias de resposta aos PBR. Existem participantes que dão mais peso à descrição, participantes que dão mais peso aos base-rates e participantes que fazem a integração cognitiva ponderada das duas informações.
- 2- O desempenho em PBR correlaciona-se com o desempenho noutras medidas de racionalidade e com os valores obtidos em escalas de estilo cognitivo.
- 3- Participantes que recebem feedback corretivo obtêm um desempenho melhor em PBR. No entanto, um treino em PBR inclusivos não se generaliza para PBR exclusivos.

No final desta tese discutem-se as vantagens e as desvantagens do paradigma proposto, as suas limitações e potencialidades para investigações futuras.

Palavras-chave: Problemas de Base-Rate (PBR), Enviesamento Heurístico, Integração Cognitiva, Ambientes de Aprendizagem, Probabilidades Condicionadas, Fórmula de Bayes.

Agradecimentos

Gostaria de agradecer aos Grandes Professores da minha vida que tanto me ensinaram.

À minha Mãe que me ensinou a escrever.

Ao meu Pai que me ensinou a contar.

A todos os meus Avós que cuidaram de mim.

Aos Professores da primária que me ensinaram a ouvir.

Aos Professores do secundário que me ensinaram a pensar.

Aos Professores da faculdade que me ensinaram a teoria.

Aos Colegas da faculdade que me ensinaram a prática.

Gostaria de agradecer agora a todas as pessoas que contribuíram diretamente para esta tese.

Muito obrigado Prof. Doutor Mário Ferreira por ter sido um excelente mestre e orientador desta tese de mestrado. Obrigado Thomas Bayes por ter criado a fórmula da probabilidade condicionada. Obrigado André Vaz por me ter ajudado a deduzir a fórmula inversa de Bayes. Obrigado Joana Reis por fornecer os pré-testes do primeiro estudo e fazer a revisão científica da tese. Obrigado à minha Mãe por fazer a revisão ortográfica e sintática da tese, um trabalho exigente. Obrigado ao meu Pai por fazer a auditoria da tese, um trabalho de perícia. Obrigado ao meu Avô Hugo por fazer a revisão da tese, a sua perspicácia não diminui com a idade. Obrigado à Rita Trindade por ajudar a rever a tese e identificar as partes mais confusas. Agradeço a todos os meus amigos que sempre me apoiaram e me deram força para ir mais longe e nunca desistir. Gostava de agradecer também aos mais de mil participantes que, com a sua boa vontade, responderam aos cinco questionários referidos nesta tese de mestrado. Os questionários eram difíceis e o vosso esforço fez toda a diferença, uma diferença significativa.

Esta tese é o resultado do culminar de uma fase importante da minha vida, representa a combinação de conhecimentos teóricos, metodológicos e práticos na área de julgamento e decisão em incerteza, esperando que possa vir a inspirar investigações futuras nesta área, por isto, gostaria ainda de agradecer ao leitor desta tese: muito obrigado por ler esta tese! Espero que os conhecimentos que encontrar aqui, o inspirem a fazer bons julgamentos e a tomar boas decisões, tendo como certeza que no meio é que está a virtude.

Índice

Introdução.....	1
Paradigma clássico de Problemas de Base-Rate	2
Modelos Dualistas de Julgamento e Decisão na Incerteza.....	4
Paradigma de Problemas de Base-Rate proposto nesta tese.....	8
Resumo das experiências realizadas.....	11
Experiência 1: Paradigma Contínuo	15
Experiência 2: Escolha Livre.....	19
Experiência 3: Base-Rates Adaptados	22
Experiência 4: Base-Rates Adaptados - Replicação.....	29
Experiência 5: Paradigma de aprendizagem.....	36
Resumo dos resultados	44
Discussão geral.....	45
Os participantes são capazes de integrar as informações apresentadas em PBR? ..	45
Esta integração cognitiva correlaciona-se com outras medidas de racionalidade? ..	45
Os participantes melhoram o seu desempenho ao recebem feedback corretivo?....	45
Limitações do paradigma proposto.....	46
A escala contínua de resposta apresenta um elevado grau de ruído estatístico.....	46
Normatividade da resposta bayesiana e controlo experimental	46
A subjetividade objetiva e objetividade subjetiva da diagnosticidade	47
Investigações futuras	48
Transformar variáveis estranhas em variáveis independentes controladas	48
Adaptação do paradigma de PBR a outros métodos e materiais	49
Referências Bibliográficas.....	50
Anexos.....	51

Introdução

Quando tomamos decisões com base em informações complexas, temos tendência a utilizar heurísticas que simplificam o nosso pensamento. Ao longo do século XX vários investigadores mostraram que os participantes não aderiam às prescrições de modelos normativos quando respondiam a diversos tipos de problemas. Muitos destes problemas sugerem uma resposta intuitiva que resulta de processos heurísticos. No entanto, para responder corretamente é necessário que o participante assuma um ponto de vista crítico e que baseie a sua resposta, não na sua intuição, mas no seguimento de uma análise lógica e racional.

Os Problemas de Base-Rate (PBR) criados por Kahneman e Tversky (1973) são muito utilizados na área da investigação de julgamento probabilístico. Estes problemas pedem uma previsão probabilística e fornecem duas fontes de informação onde o participante se pode apoiar para dar a sua resposta. A primeira informação apresentada é o base-rate, uma informação quantitativa sobre o número ou percentagem elementos que compõem dois grupos de pessoas. A segunda informação apresentada é uma descrição individualizante que pode ser diagnóstica de um dos grupos. O PBR termina com uma pergunta de probabilidade condicionada à informação individualizante. Os investigadores Kahneman e Tversky encontraram evidência de que, quando as duas informações levavam a conclusões opostas, os participantes tendiam a basear as suas respostas na descrição individualizante, desvalorizando os base-rates.

Embora muitos estudos tenham confirmado a preferência geral de um tipo de informação em detrimento do outro, poucos estudos se focaram em estudar a capacidade individual de integração das duas informações. Nesta tese apresentamos um paradigma adaptado dos problemas originais de Kahneman e Tversky que compara a resposta dos participantes a uma norma bayesiana. Podendo comparar a resposta de cada participante a uma norma bayesiana levantamos as seguintes perguntas de investigação:

- 1- Os participantes são capazes de integrar as informações apresentadas em PBR?
- 2- Esta integração cognitiva correlaciona-se com outras medidas de racionalidade?
- 3- Os participantes melhoram o seu desempenho quando recebem feedback corretivo?

O objetivo desta tese é dar resposta a estas perguntas. De modo a enquadrar este trabalho na investigação feita anteriormente iremos apresentar o paradigma original e as implicações teóricas dos seus resultados, seguido de um exemplo prático. Após essa introdução fazemos uma breve revisão dos modelos dualistas na área de julgamento e decisão na incerteza. Por fim, apresentamos as limitações do paradigma original e explicamos em detalhe o paradigma proposto nesta tese.

Paradigma clássico de Problemas de Base-Rate

Vejam os seguintes exemplos de um PBR inclusivo no paradigma clássico:

Numa festa, estão 70 engenheiros e 30 advogados.

Destas 100 pessoas, foi escolhida uma pessoa ao acaso que gosta de ficção científica.

Qual é a probabilidade desta pessoa ser um engenheiro?

Este é um exemplo do paradigma clássico de PBR. Este problema apresenta dados quantitativos e qualitativos e a sua resolução envolve alguma subjetividade. Se considerarmos que a descrição não é diagnóstica dos dois grupos, a resposta deverá ser 70%. Se considerarmos que a descrição é mais típica de engenheiros do que de advogados, a resposta estará entre 70% e 100%. Se considerarmos que a descrição é mais típica de advogados do que de engenheiros a resposta deverá ser inferior a 70%. Quando ambas as informações apresentadas levam à mesma conclusão o problema diz-se ser inclusivo. Quando as duas informações levam a conclusões opostas, os problemas chamam-se exclusivos. Este PBR é considerado inclusivo porque para a maioria das pessoas, gostar de ficção científica é uma característica mais típica de engenheiros do que de advogados, no entanto, não isso não será necessariamente verdade para todas as pessoas.

Vejam os seguintes exemplos de um PBR exclusivo no paradigma clássico:

Numa festa, estão 30 engenheiros e 70 advogados.

Destas 100 pessoas, foi escolhida uma pessoa ao acaso que gosta de ficção científica.

Qual é a probabilidade de esta pessoa ser um engenheiro?

Esta versão do problema é mais difícil de responder, pois apresenta um dilema para a maioria dos participantes. Enquanto a informação quantitativa nos leva a crer que a resposta é inferior a 50%, a informação qualitativa leva-nos a crer que a resposta é superior a 50%. Ou seja, embora haja mais advogados na festa, gostar de ficção científica é uma característica mais típica de engenheiros do que de advogados. A resposta deverá estar algures entre 30% e 100%, dependendo da diagnosticidade subjetiva da descrição, que pode variar de pessoa para pessoa. Para responder ao problema os participantes devem considerar a percentagem de engenheiros e advogados que gostam de ficção científica e também as probabilidades de partida.

Uma das descobertas que os psicólogos Daniel Kahneman e Amos Tversky fizeram nos anos 70, foi que, quando apresentadas a dois grupos de pessoas este tipo de problemas com base-rates diferentes (30%/70% ou 70%/30%), as pessoas pareciam ignorar os base-rates, dando respostas com base apenas na semelhança da descrição ao grupo alvo. Embora estes autores tenham encontrado diferenças significativas nas respostas dadas em função dos base-rates apresentados ($p < 0,01$), a resposta mediana não se alterava significativamente. Poucos anos mais tarde, Lyon & Slovic (1976) replicaram estes resultados com base-rates extremos (1% e 99%).

Para explicar esta insensibilidade aparente aos base-rates, Kahneman e Tversky argumentam que os participantes ao resolverem PBR não estão a pensar em probabilidades, mas sim a fazer uso das heurísticas de disponibilidade e semelhança. Estas heurísticas facilitam o julgamento ao substituírem a integração probabilística das duas informações por julgamentos mais fáceis, como por exemplo: : “*Consigo lembra-me com facilidade de engenheiros ou advogados que gostam de ficção científica?*” e “*Uma pessoa que gosta de ficção científica é mais parecida com um engenheiro ou com um advogado?*”.

Este tipo de raciocínio pode ser útil quando as probabilidades de partida são 50%, mas quando elas não são 50%, o recurso exclusivo a estas heurísticas leva-nos a raciocínios enviesados.

Como referido, os PBR são muito utilizados na investigação na área do julgamento e decisão em incerteza (De Neys, 2012), mas será que estes problemas têm aplicações práticas fora do contexto académico?

Tomemos como exemplo a seguinte situação:

Existe uma doença com uma prevalência de 2% na população portuguesa.

O Sistema Nacional de Saúde disponibiliza milhares de kits para testar a doença.

Estes kits de última geração têm uma sensibilidade de 100% e uma especificidade de 93,9%.

Uma sensibilidade de 100% significa que todas as pessoas infetadas terão resultado positivo.

Uma especificidade de 93,9% significa que 6,1% das pessoas saudáveis terão um falso positivo.

Dezenas de milhares de pessoas fazem o teste para saberem se têm a doença.

Esta doença não tem sintomas específicos e a pessoa pode estar infetada sem saber.

Qual é a probabilidade de uma pessoa com um resultado positivo ter de facto a doença?

Poderá parecer que a probabilidade é alta, pois tanto a sensibilidade como a especificidade do teste são muito elevadas. Parece então que o teste é extremamente fiável. No entanto, como apenas 2% da população tem a doença, isso significa que 98% da população não tem a doença. Se escolhermos aleatoriamente 100 portugueses para aplicar este teste, 2 seriam corretamente diagnosticados, mas 6 das 98 pessoas saudáveis, seriam incorretamente diagnosticadas. Assim, a probabilidade de uma pessoa ter a doença dado que o teste deu positivo é apenas 25%.

Conclusão, de modo a evitar fazer diagnósticos errados, o teste não deve ser aplicado de modo indiscriminado, mas dirigido às populações com maior risco e maior taxa de incidência da doença. Este é um exemplo com implicações reais de como uma evidência diagnóstica deve ser relativizada quando existe uma baixa probabilidade de partida.

Embora possa parecer que esta integração é muito difícil, alguns estudos mostram que os base-rates são utilizados de maneira próxima à prescrita pelos modelos normativos e que a sua utilização depende da estrutura do problema e do tipo de representação (Koehler, 1996). Estes resultados podem deixar-nos otimistas em relação à possibilidade de estudar a capacidade das pessoas de integrar estas duas informações.

Curiosamente, apesar das evidências de que os participantes estão a atender às descrições e aos base-rates, a investigação em PBR não se tem focado em estudar a capacidade de integração das duas informações. Em estudos recentes, os PBR têm sido utilizados de forma simplificada, onde são apresentados base-rates extremos (0,5% e 99,5%) e a resposta é pedida num formato dicotómico (De Neys, 2012). Uma resposta dada de acordo com a descrição é interpretada como uma resposta resultante de um processamento heurístico e intuitivo, enquanto uma resposta de acordo com os base-rates é interpretada como resultante de um processamento deliberado e analítico. Esta distinção destes dois processos de pensamento é muito importante para a teorização na área de julgamento e decisão em incerteza. Vamos explorar um pouco mais a distinção entre os dois processos cognitivos antes de avançarmos.

Modelos Dualistas de Julgamento e Decisão na Incerteza

De acordo com a abordagem dualista, os julgamentos e as decisões que realizamos no quotidiano decorrem de dois tipos qualitativamente diferentes de processamento de informação: o processamento tipo 1 e o processamento tipo 2.

O processamento Tipo 1 caracteriza-se por ser intuitivo e autónomo. É despoletado automaticamente perante estímulos específicos que podem ter origem biológica, derivar de aprendizagens implícitas inconscientes ou de aprendizagens explícitas que sejam treinadas e automatizadas. Estes processos permitem-nos dar respostas rápidas e frequentemente satisfatórias, mas, em certas circunstâncias, estes processos podem levar ao enviesamento se não forem inibidos e substituídos por respostas mais deliberadas.

O processamento Tipo 2 é deliberado, normalmente mais lento e consciente. No entanto, este processamento lógico e abstrato possui recursos cognitivos limitados (e.g., Evans & Stanovich, 2013). Na perspetiva dualista, uma das características essenciais para ter um bom desempenho em tarefas de julgamento complexo é a disposição para não responder imediatamente (e.g., Stanovich, 2009). Muitas vezes, é essencial analisar cuidadosamente toda a informação disponível, considerar vários pontos de vista e refletir antes de dar uma resposta. Esta tendência para inibir a resposta intuitiva parece ser bem captada pelo Teste de Reflexão Cognitiva (CRT; Frederick, 2005), uma das melhores medidas de racionalidade conhecidas (Toplack, Stanovich & West, 2012).

O teste original é composto por três problemas:

1. Num lago de nenúfares, todos os dias o número de nenúfares duplica.

São necessários 48 dias para que todo o lago esteja coberto de nenúfares.

Quantos dias são necessários para que metade do lago esteja coberto de nenúfares?

2. Cinco máquinas demoram cinco minutos a fazer cinco peças.

Quanto tempo demoram cem máquinas a fazer cem peças?

3. Um bloco de notas e um lápis custam 1 euro e 10 cêntimos.

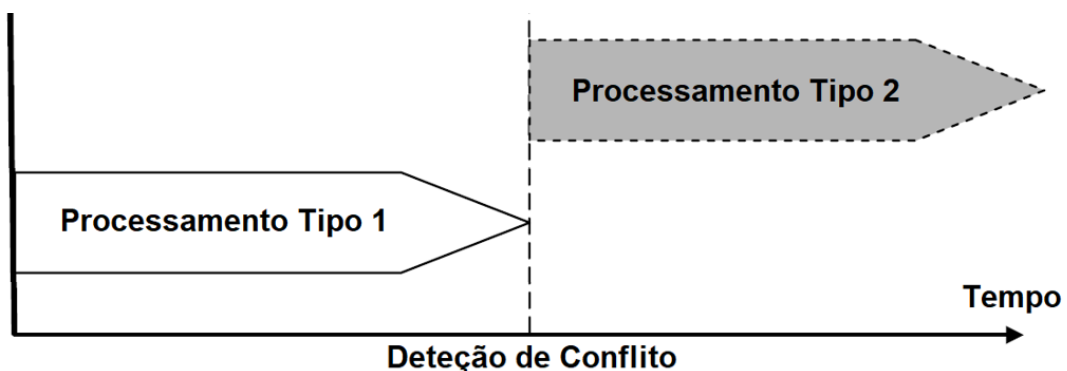
O bloco de notas custa um euro a mais do que o lápis.

Quanto custa o lápis?

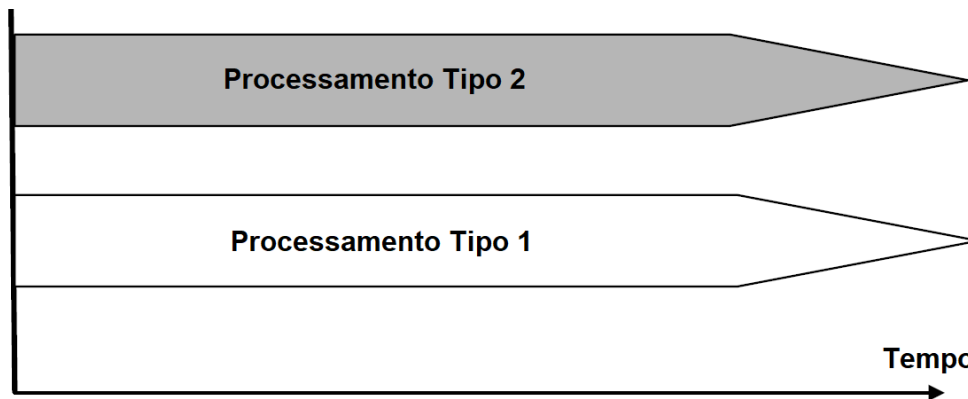
Estes problemas foram criados de modo a sugerir uma resposta intuitiva muito apelativa, mas errada. Contudo, se a pessoa inibir temporariamente a resposta intuitiva e raciocinar um pouco chegará à resposta certa. Este é um aspeto comum tanto em PBR como no CRT. Se ambos os tipos de problemas estiverem a medir disposições e processos cognitivos semelhantes então estas medidas deverão estar correlacionadas. Mas como é feita essa regulação? Como se relaciona o processamento Tipo 1 e o processamento Tipo 2?

Existem quatro modelos que explicam os dois processamentos interagirem: o modelo serial, o modelo paralelo, o modelo das intuições lógicas e o modelo trifásico

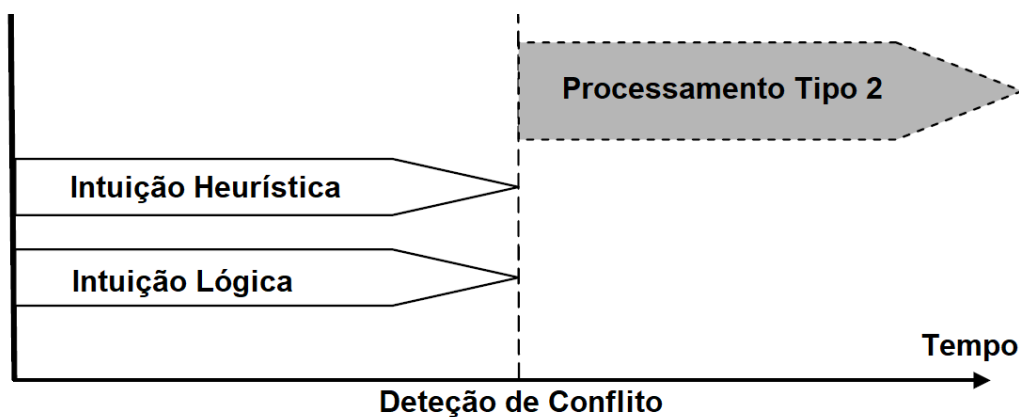
No **modelo serial** a primeira resposta ao problema é desencadeada por processos tipo 1 e os processos tipo 2 só são desencadeados se os resultados do processo tipo 1 forem considerados insuficientes (Kahneman & Frederick, 2005; Stanovich & West, 2000).



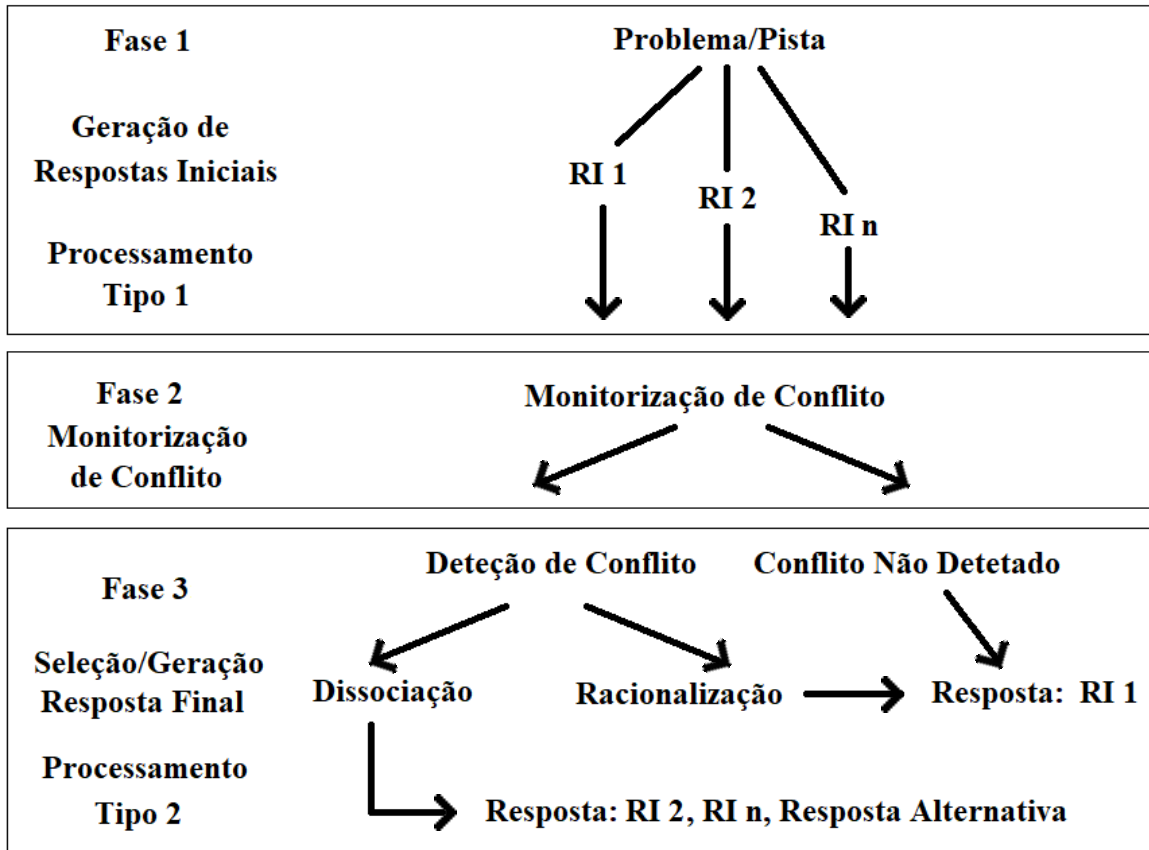
No **modelo paralelo**, ambos os sistemas são ativados desde o início (Epstein, 1994; Sloman, 1996). Este modelo tem mais facilidade em explicar a percepção do conflito entre os dois tipos de processamento, mas assume que os processos tipo 2 ocorrem logo desde o início a resolução do problema o que pode ser um gasto desnecessário de recursos cognitivos.



No **modelo das intuições lógicas**, tanto os base-rates como a descrição ativam processos tipo 1 (De Neys, 2012). Este modelo assume a existência de intuições heurísticas baseadas em associações semânticas e estereotípicas e também a existência de intuições lógicas baseadas em princípios estatísticos. O conflito entre as duas intuições irá ativar o processamento tipo 2 que pode ter sucesso em resolver o conflito ou falhar, mas o processamento tipo 2 só ocorre quando é detetado um conflito entre a intuição heurística e a intuição lógica.



Por fim, temos o modelo trifásico (Pennycook, Fugelsang & Koehler, 2015). Este modelo é semelhante ao modelo da intuição lógica no pois admite que, para além das respostas intuitivas heurísticas (ex: RI 1), podem haver respostas intuitivas lógicas (ex: RI 2), e que estas podem ser geradas antes ou depois das respostas intuitivas heurísticas. Para além disso, este modelo sugere que existem duas possíveis falhas do processamento tipo 2. O participante pode não conseguir detetar o conflito existente entre as várias respostas intuitivas, ou mesmo detetando o conflito, pode não o conseguir resolver, racionalizando a primeira resposta intuitiva como sendo a resposta correta, ignorando o conflito.



Em todos estes modelos, o resultado desejável é sempre a inibição de uma resposta intuitiva e a sobreposição de uma resposta analítica. Ou seja, os participantes que dão uma resposta normativamente corretam tiveram a capacidade de inibir a resposta intuitiva e utilizar um processamento analítico e deliberado para chegar à sua resposta. No entanto, como referido anteriormente, em estudos recentes com PBR a resposta é pedida num formato dicotómico. Neste paradigma, quando o participante dá uma resposta congruente com os base-rates esta resposta é vista como uma resposta analítica e deliberada.

Dependendo do modelo, podemos considerar que, para alguns participantes, os base-rates podem ser respostas intuitivas. Nesse caso, respostas congruentes com base-rates não refletem necessariamente um tipo de processamento do tipo 2, mas sim a utilização dos base-rates como uma intuitiva resultante de um processamento tipo 1. Esta é uma limitação metodológica dos estudos que utilizam um paradigma dicotómico em PBR. Estes estudos não distinguem entre os participantes com intuições lógicas e os participantes verdadeiramente racionais. Estudos recentes indicam que alguns participantes têm uma tendência inicial para responder de acordo com os base-rates (Newman, Gibb & Thompson, 2017) e que esta tendência de resposta se pode inverter quando os participantes têm tempo para refletir (Loureiro & Garcia-Marques, 2018). Para além de haver a possibilidade de o participante utilizar o base-rate como pista intuitiva existem outras explicações para ele dar uma resposta congruente com o base-rate. O participante pode estar a dar a resposta mais socialmente aceitável, ou seja, aquela que rejeita o estereótipo.

Concluindo, o paradigma de resposta dicotômica, por ser um paradigma de resposta forçada, pode levar a interpretações erradas. Num caso extremo, se uma pessoa responder ao acaso irá dar uma resposta “analítica” 50% das vezes. Um bom teste é aquele que apresenta muitas maneiras de errar e apenas uma maneira de estar correto e o paradigma dicotômico em PBR é problemático por esta mesma razão.

Os paradigmas de resposta dicotômica não distinguem os participantes que integram as descrições e os base-rates de participantes que estão a utilizar os base-rates de forma heurística. Utilizando uma escala de resposta contínua existe uma menor possibilidade de um participante acertar na resposta ao acaso. Poderíamos pensar que a solução seria regressar ao paradigma clássico de Kahneman e Tversky (1973), pois eles utilizam uma escala contínua de resposta. No entanto, a metodologia que estes autores utilizaram não permite obter um índice individualizado de integração cognitiva dos base-rates e da descrição. O paradigma que estes autores utilizaram compara a resposta mediana de dois grupos de participantes que recebem duas versões do mesmo PBR. Numa versão o base-rate é 30%/70%, na outra versão o PBR é 70%/30%. Estes autores verificaram que a resposta mediana não era diferente nas duas versões do problema, concluindo assim que a manipulação dos base-rates tinha um efeito mínimo sobre a probabilidade subjetiva condicionada às descrições. Embora a investigação posterior tenha vindo a mostrar que os base-rates podem ser utilizados de maneira próxima à prescrita pelos modelos normativos (Koehler, 1996), não houve um foco em estudar a capacidade de integração em PBR.

Paradigma de Problemas de Base-Rate proposto nesta tese:

Nesta tese propomos um paradigma em PBR que mede a integração cognitiva das duas fontes de informação, comparando a resposta dada pelo participante a uma resposta bayesiana.

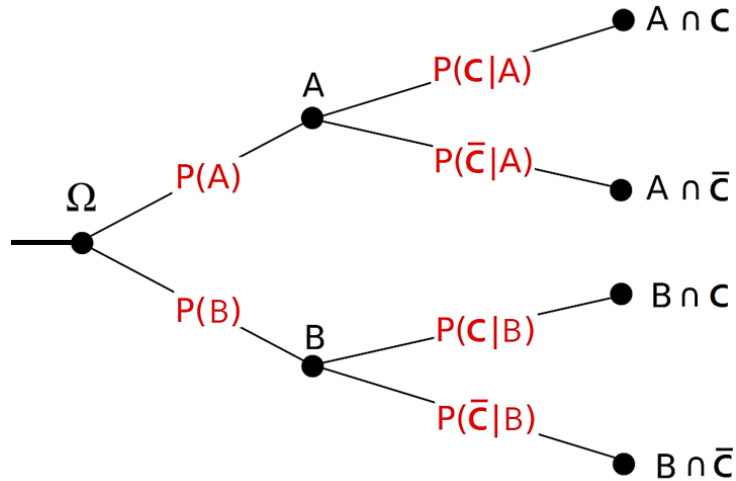
$$\text{Viés} = \text{Resposta ao PBR} - \text{Resposta Bayesiana}$$

O desvio absoluto á resposta bayesiana é visto como um índice do desempenho em PBR.

$$\text{Viés Absoluto} = | \text{Resposta ao PBR} - \text{Resposta Bayesiana} |$$

Utilizando este paradigma, ao apresentar PBR a uma amostra de participantes podemos obter dois índices a partir das suas respostas: o desempenho e a tendência de resposta. O desempenho representa a proximidade média à resposta bayesiana. Este valor é obtido a partir do viés absoluto médio. A tendência de resposta representa o tipo de enviesamento mais comum. Este valor é obtido a partir do viés médio. Como o viés ao PBR é uma medida com sinal, a tendência de resposta será um índice também com sinal. O sinal deste índice permite-nos identificar qual dos dois tipos de enviesamentos foram mais prevalentes na amostra.

O paradigma proposto assenta na ideia que todos os elementos apresentados no PBR (Ω) pertencem exclusivamente a uma de quatro categorias possíveis: pessoas do grupo A (com e sem a característica C) e pessoas do grupo B (com e sem a característica C). $P(A)$ é a probabilidade de uma pessoa ser do grupo A e $P(B)$ a probabilidade de uma pessoa ser do grupo B. $P(C|A)$ é a probabilidade de uma pessoa do grupo A ter a característica C e $P(C|B)$ a probabilidade de uma pessoa do grupo B ter a característica C.



De acordo com esta notação, a resposta bayesiana é calculada da seguinte forma:

$$P(A|C) = \frac{P(A \cap C)}{P(A \cap C) + P(B \cap C)} = \frac{P(A) * P(C|A)}{P(A) * P(C|A) + P(B) * P(C|B)}$$

Se quisermos calcular o base-rate necessário para se obter uma determinada resposta bayesiana com base nos valores da diagnosticidade, podemos usar a seguinte fórmula:

$$P(A) = \frac{P(C|B) * P(A|C)}{P(C|A) - P(C|A) * P(A|C) + P(C|B) * P(A|C)}$$

Observemos o seguinte exemplo de um PBR no paradigma proposto:

Numa festa estão 30 engenheiros e 70 advogados.

Destas 100 pessoas, foi escolhida uma pessoa ao acaso que gosta de ficção científica.

Na sua opinião, que percentagem de engenheiros gosta de ficção científica?

Na sua opinião, que percentagem de advogados gosta de ficção científica?

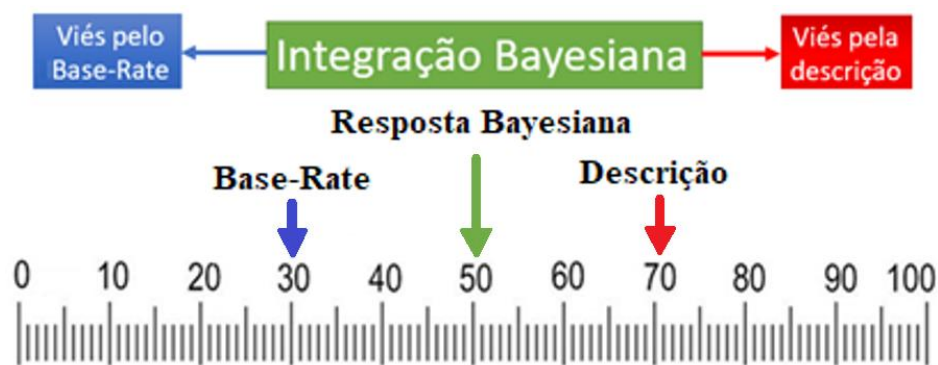
Qual é a probabilidade da pessoa escolhida, entre estas 100, ser um engenheiro?

Imaginemos que um participante que considera que 70% dos engenheiros e 30% dos advogados gostam de ler ficção científica. Neste caso, gostar de ficção científica tem uma diagnosticidade de 70% para diferenciar entre engenheiros e advogados. A diagnosticidade da descrição é calculada através da seguinte fórmula:

$$\text{Diagnosticidade} = 100 * \frac{P(C|A)}{P(C|A) + P(C|B)}$$

Os valores da diagnosticidade podem ser pedidos ou fornecidos ao participante. Neste exemplo os valores foram pedidos ao participante. Deste modo, cada participante terá uma resposta bayesiana individual de varia em função da diagnosticidade subjetiva da descrição. Este método tem a vantagem de não assumir o valor diagnóstico da descrição à priori. No entanto, os valores da diagnosticidade podem ser apresentados no próprio PBR, deste modo garante-se que todos os participantes têm acesso aos mesmos valores de diagnosticidade da descrição. Cada um dos dois métodos apresenta vantagens e desvantagens.

Voltando ao exemplo apresentado. Imaginemos que um participante considera que 70% dos engenheiros e 30% dos advogados gostam de ler ficção científica. Neste caso a resposta bayesiana a este PBR é 50%. Como 70% de 30 e 30% de 70 é igual a 21, existem na festa 21 engenheiros e 21 advogados que gostam de ler ficção científica. Se dermos a mesma importância aos base-rates e a descrição individualizante a resposta deverá ser 50%.



Utilizando este paradigma podemos medir dois tipos de enviesamento em função do tipo de desvio à resposta bayesiana. Este desvio pode ocorrer tanto no sentido da descrição como no sentido dos base-rates. Um viés pela descrição deve-se à utilização de heurísticas de disponibilidade e semelhança (Tversky & Kahneman, 1981). Um viés pelos base-rates deve-se à utilização de heurísticas de ancoragem e ajustamento insuficiente e é considerado um efeito de conservadorismo (Kahneman, 1966).

Neste paradigma a capacidade de integração é vista como um contínuo, uma aproximação a um ideal normativo e não apenas um binário, onde a pessoa é considerada racional ou enviesada.

Resumo das experiências realizadas

Ao longo de 5 experiências procurámos dar resposta às seguintes perguntas:

- 1- Serão os participantes capazes de integrar as informações apresentadas em PBR?
- 2- Esta integração cognitiva correlaciona-se com outras medidas de racionalidade?
- 3- Os participantes melhoram o seu desempenho quando recebem feedback corretivo?

De seguida apresentamos um resumo das experiências, os seus fundamentos teóricos, metodológicos e os seus resultados e conclusões.

Experiência 1: Paradigma Contínuo.

Nesta experiência procurámos dar resposta às primeiras duas perguntas de investigação.

Neste estudo os participantes responderam a 5 PBR e de seguida quantificaram a diagnosticidade da descrição. Com base nestes valores calculámos uma resposta bayesiana individual para cada participante. A proximidade da resposta dada à resposta bayesiana individual foi considerada como o índice do desempenho. O distância da resposta dada à resposta bayesiana individual foi considerada como um índice de enviesamento.

Embora os participantes tenham apresentado uma tendência geral para sobrevalorizar a descrição em detrimento dos base-rates, vários participantes deram respostas próximas às prescritas pelo modelo bayesiano.

Para além dos PBR os participantes também responderam a três problemas de CRT. O CRT é uma medida de racionalidade que reflete a capacidade de monitorização e inibição das respostas intuitivas, e análise deliberada de respostas alternativas.

O desempenho dos participantes no CRT correlacionou-se significativamente com o desempenho em PBR ($r = 0,193$; $p = 0,018$).

A relação entre estas duas medidas de desempenho apoia a ideia de que a capacidade de inibição das respostas intuitivas é uma característica comum em PBR e no CRT.

Em primeiro lugar encontramos evidência de que alguns participantes parecem estar a utilizar uma estratégia de integração cognitiva em PBR. Em segundo lugar a proximidade das suas respostas à resposta bayesiana correlaciona-se com o CRT, uma outra medida de racionalidade. Estes dados apoiam os resultados obtidos por Hoppe e Kusterer (2011). Estes autores verificaram que participantes com baixo desempenho em CRT apresentam uma tendência a serem mais enviesados em problemas onde existe uma resposta normativa, tanto no sentido da descrição como no sentido dos base-rates.

Experiência 2: Escolha Livre.

Os resultados obtidos na primeira experiência indicam a existência de participantes que respondem próximo da resposta bayesiana. Nesta experiência procurámos replicar este resultado e mostrar que esta capacidade de fazer uma integração bayesiana não depende do material utilizado. Para este fim pedimos que os participantes criassem os seus próprios problemas de base-rate, escolhendo dois grupos e uma característica que fosse típica de um e atípica do outro. De seguida, utilizando os grupos e as características escolhidas pelos participantes pedíamos aos participantes para quantificar a frequência da característica em cada um dos grupos. Por fim, utilizando os grupos e as características criadas pelos participantes apresentamos um PBR exclusivo com um base-rate de 7%.

Nesta experiência confirmamos a existência de um número significativo de participantes a responder próximo da resposta bayesiana. Concluímos que existem muitos participantes capazes de integrar as informações apresentadas em PBR e que esta capacidade não depende do tipo de material utilizado.

Experiência 3: Base-Rates Adaptados.

Na primeira experiência o desempenho em PBR correlacionou-se com o desempenho no CRT, uma medida concorrente de racionalidade. Nesta experiência procuramos replicar estes resultados. Para além disso tentamos correlacionar o desempenho em PBR com uma medida de inteligência fluida, o desempenho numa bateria das Matrizes Progressivas de Raven (MPR; Bilker et al., 2012). Esta bateria de testes é uma medida geral de inteligência (fator g) e relaciona-se com a capacidade de fazer a integração cognitiva de padrões abstratos de imagens. Para além disso, correlacionamos o desempenho em PBR com os valores obtidos nas escalas Faith in Intuition (FI; Epstein et al., 1996) e Need for Cognition (NFC; Cacioppo & Petty, 1982). Estas escalas são medidas do estilo de pensamento e encontram-se tipicamente correlacionadas com o desempenho em tarefas de raciocínio. Por fim, como variável de controlo correlacionamos o desempenho em PBR com a média do secundário de cada participante.

Ao contrário do esperado, não encontramos correlações significativas entre o desempenho em PBR e as outras medidas de raciocínio, inteligência e estilo cognitivo. Curiosamente, a única medida que se correlacionou significativamente com o desempenho em PBR nesta experiência foi a média do secundário ($r = 0,157$; $p = 0,014$).

Esta falha em detetar correlações significativas deveu-se possivelmente a uma diferença significativa no design experimental. Nesta experiência em vez de utilizarmos base-rates fixos apresentamos base-rates adaptados a cada participante de modo a que a resposta bayesiana fosse fixa. Esta manipulação baixou o nível de desempenho nos PBR e piorou a qualidade desta medida.

Experiência 4: Base-Rates Adaptados - Replicação

Por não termos obtido correlações significativas entre o desempenho em PBR e o desempenho no CRT, decidimos replicar a experiência 3 com algumas melhorias metodológicas:

1. Execução dos questionários em condições laboratoriais.
2. Apresentação de instruções mais claras sobre os objetivos da experiência.
3. Apresentação de um PBR de treino onde o base-rate apresentado era 50%.
4. Medição da diagnosticidade da descrição imediatamente antes do PBR.
5. Apresentação da diagnosticidade da descrição no enunciado do PBR.

Nesta experiência obtivemos uma correlação significativa entre o desempenho em PBR e o desempenho no CRT ($r = 0,169$; $p = 0,023$). A correlação entre o desempenho em PBR e o desempenho nas MPR foi marginalmente significativo ($r = 0,133$; $p = 0,074$). As correlações com as restantes medidas não foram significativas ($p > 0,2$).

Nesta experiência confirmámos os resultados obtidos na experiência 1, utilizando um design experimental diferente. O desempenho em PBR correlacionou-se significativamente com o desempenho no CRT, marginalmente com o desempenho nas MPR e não significativamente com as outras medidas. Isto significa que a capacidade de inibir as respostas intuitivas é mais relevante para a resolução bayesiana de PBR do que a capacidade de integração cognitiva de padrões abstratos de imagens, que o estilo cognitivo e que a média do secundário.

Experiência 5: Paradigma de Aprendizagem.

Nesta última experiência o nosso objetivo foi dar resposta às três perguntas de investigação desta tese e também esclarecer as hipóteses levantadas na quarta experiência.

Nesta experiência comparámos o efeito de ambientes de aprendizagem com feedback e sem feedback em PBR inclusivos e exclusivos (4 condições experimentais). Os participantes respondiam a 7 PBR na fase de aprendizagem e a 3 PBR na fase de teste. Na fase de teste os PBR apresentados eram todos exclusivos e os participantes não recebiam feedback. Os ambientes que não promovem a aprendizagem chamam-se ambientes wicked, os ambientes que apresentam feedback corretivo adequado chamam-se ambientes kind (Hogarth, 2003).

Na fase de aprendizagem encontrámos diferenças significativas entre o desempenho dos participantes em PBR inclusivos com e sem feedback ($t = 2,3$; $p = 0,024$). No entanto esta diferença não foi significativa entre PBR exclusivos com e sem feedback ($t = 1,3$; $p = 0,193$).

Na fase de teste encontrámos o padrão inverso. A diferença de desempenho dos participantes treinados em PBR inclusivos com e sem feedback não foi significativa ($t = 0,586$; $p = 0,56$) e a diferença de desempenho dos participantes treinados em PBR exclusivos com e sem feedback foi marginalmente significativa ($t = 1,767$; $p = 0,081$).

A partir deste padrão de resultados concluímos que é possível a aprendizagem em PBR através do feedback, pelo menos em PBR inclusivos. Estes dados apoiam os resultados obtidos por Lindeman, Brink e Hoogstraten (1988). Estes autores verificaram que os participantes aproximavam a sua estimativa ao standard bayesiano depois de receberem feedback em PBR.

O treino em PBR inclusivos pode ser considerado um ambiente wicked quando os participantes são testados em PBR exclusivos. Embora os participantes aprendam a fazer uma integração cognitiva em PBR inclusivos, quando são testados em PBR exclusivos, esta aprendizagem não lhes é útil.

Por outro lado, o treino em PBR exclusivos pode ser considerado um ambiente kind, pois embora não apresente resultados imediatos na fase de aprendizagem, promove uma melhoria do desempenho na fase de teste, mesmo que apenas marginalmente significativa.

Os participantes responderam também a um CRT, a duas escalas de estilo cognitivo (NFI e FI) e forneceram a sua a média do secundário. O desempenho nos PBR correlacionou-se significativamente com todas estas medidas, no entanto, apresentou uma correlação mais forte com o CRT ($r = 0,515$; $p < 0,001$). Atribuímos a magnitude desta correlação às melhorias metodológicas introduzidas nesta última experiência.

Experiência 1: Paradigma Contínuo

Introdução: Nesta primeira experiência o nosso objetivo foi dar resposta às seguintes perguntas:

- 1- Serão os participantes capazes de integrar as informações apresentadas em PBR?
- 2- Esta integração cognitiva correlaciona-se com outras medidas de racionalidade?

Neste estudo os participantes responderam a cinco PBR e de seguida quantificaram a diagnosticidade da descrição. Com base nestes valores calculámos uma resposta bayesiana individual para cada participante. A proximidade da resposta dada à resposta bayesiana individual foi considerada como o índice de integração cognitiva. A distância da resposta dada à resposta bayesiana individual foi considerada como um índice de enviesamento.

Hipóteses:

- 1- Alguns participantes vão ser capazes de integrar as informações apresentadas nos PBR.
- 2- Esperamos encontrar uma tendência de resposta no sentido da descrição, devido à utilização de heurísticas de semelhança e disponibilidade (Kahneman & Tversky, 1973).
- 3- O desempenho em PBR e no CRT vai estar significativamente correlacionados à semelhança dos resultados obtidos por Hoppe e Kusterer (2011).

Método:

Participantes: 150 pessoas responderam a um questionário distribuído online através da plataforma Qualtrics.

Materiais: O questionário foi composto por cinco PBR e três problemas de CRT. Os grupos e descrições dos PBR foram adaptados dos problemas utilizados por Wim De Neys e Tamara Glumicic (2008) e os problemas de CRT foram adaptados dos problemas utilizados por Shane Frederick (2005). (Mais detalhes sobre os materiais em anexo).

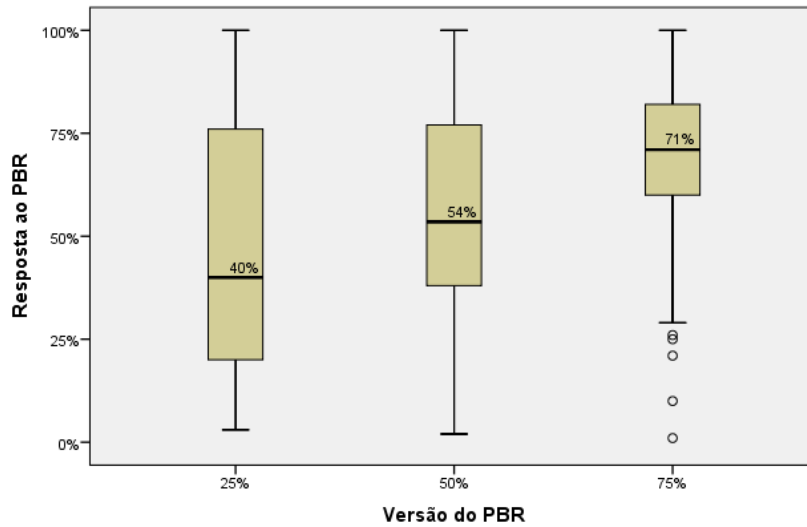
Procedimento: A diagnosticidade da descrição de cada PBR foi medida pré-teste. A partir dos valores médios da diagnosticidade das descrições foram criadas três versões diferentes. Para cada versão, foram calculados base-rates de modo a que a resposta bayesiana média fosse 25%, 50% e 75%. Esta manipulação teve dois objetivos: garantir que as respostas não se concentravam numa parte da escala de resposta e permitir a comparação das respostas em função do base-rate apresentado. De seguida, os participantes quantificaram a diagnosticidade das descrições. No final do questionário, os participantes responderam aos problemas do CRT.

Design: Nesta experiência tivemos uma variável independente, a versão do PBR (25%, 50% e 75%) e três variáveis dependentes (respostas ao PBR, diagnosticidade das descrições e desempenho no CRT). Cada participante recebeu uma das três versões do PBR aleatoriamente. Tanto os PBR como os problemas de CRT foram apresentados por ordem aleatória.

Resultados:

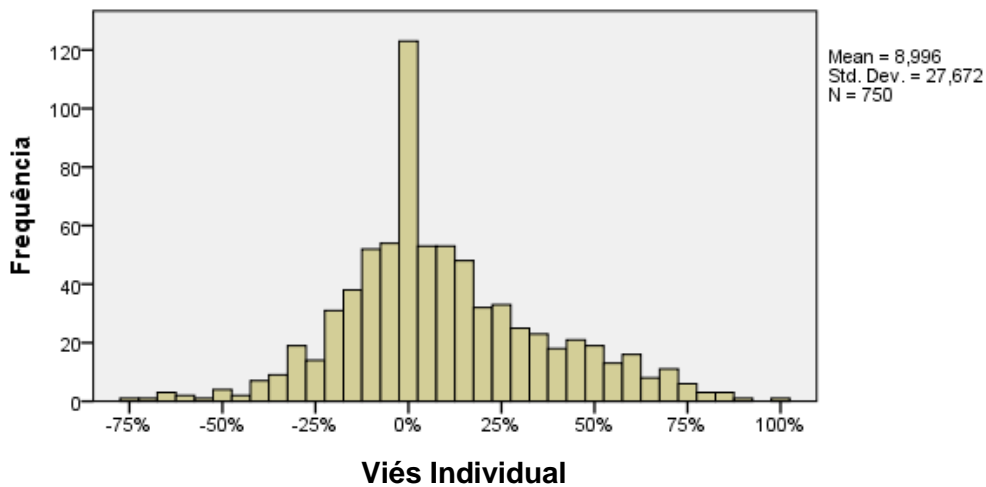
Observámos que as respostas ao PBR variaram em função da versão do problema apresentado ao participante. No gráfico 1, podemos observar a distribuição das respostas por quartis em função da versão do PBR apresentada ao participante.

Gráfico 1: Efeito da versão do PBR nas respostas dadas pelos participantes.



Enquanto a mediana na condição 25% está acima do esperado, na condição 50% e 75%, as medianas das respostas estão próximas do valor esperado. Usando uma ANOVA, encontramos diferenças significativas na média das respostas dadas aos PBR nas três versões do PBR, $F(2, 747) = 62, p < 0,001$. O output desta análise e os gráficos da distribuição das respostas em cada versão do PBR encontram-se em anexo. De seguida analisamos a proximidade das respostas dos participantes à resposta bayesiana individual. No gráfico 2, podemos observar a frequência dos vieses à resposta bayesiana individual em todas as versões.

Gráfico 2: Frequência de vieses à resposta bayesiana individual.



Um viés negativo significa que o participante respondeu abaixo da resposta bayesiana, ou seja, respondeu no sentido do base-rate apresentado. Um viés positivo significa que o participante

respondeu acima da resposta bayesiana, ou seja, respondeu no sentido da descrição. Como a média dos vieses é 9%, isso significa que em média as pessoas responderam acima da resposta bayesiana, ou seja, enviesadas para o lado da descrição. No geral os participantes obtiveram um viés absoluto médio de 21%. Mais de 10% dos participantes obtiveram um viés absoluto médio inferior a 10%, o que significa que utilizaram sistematicamente uma estratégia de resposta de integração bayesiana ao longo dos 5 PBR.

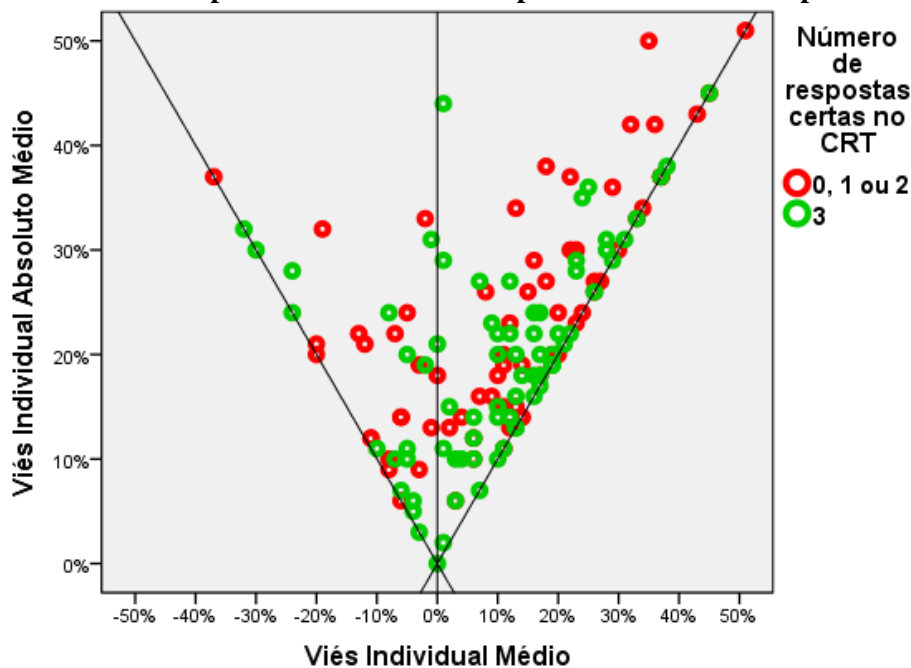
Nesta experiência é possível obter um índice alternativo do enviesamento que ignora as estimativas individuais da diagnosticidade das descrições. Este índice compara a resposta de cada participante a uma resposta bayesiana calculada em função da diagnosticidade média. A comparação entre estes dois índices encontra-se em anexo.

De seguida, analisámos o desempenho dos participantes nos problemas de CRT. Encontrámos um efeito de teto inesperado na distribuição das respostas aos problemas CRT (gráfico em anexo). Por esta razão decidimos agrupar os participantes que não responderam corretamente a todos os problemas de CRT de modo a obter dois grupos de contraste com o mesmo número de participantes. Verificámos que, quando agrupados deste modo, o desempenho nos PBR e no CRT estavam significativamente correlacionados ($r = 0,193$; $p = 0,018$).

Por fim, fizemos uma análise gráfica individualizada do desempenho e tendência de resposta, cruzando essa informação com o desempenho de cada participante no CRT.

No gráfico 3 podemos observar um gráfico de dispersão onde cada ponto representa um participante. A posição de cada ponto representa o desempenho e tendência de resposta e a cor representa o desempenho no CRT. Deste modo, num único gráfico podemos observar o desempenho, a tendência de resposta em PBR e o desempenho no CRT de cada participante.

Gráfico 3: Desempenho e tendência de resposta em PBR e desempenho no CRT.



No eixo vertical podemos observar o viés individual absoluto médio de cada participante. Este é um índice geral da capacidade de integração das informações apresentadas nos PBR. Na parte inferior do gráfico encontram-se os participantes com melhor desempenho, na parte superior do gráfico encontram-se os participantes com pior desempenho. Concluímos que os participantes com melhor desempenho em PBR (viés absoluto médio $< 10\%$, parte inferior do gráfico) tiveram maioritariamente um desempenho máximo no CRT.

No eixo horizontal podemos observar o viés individual médio de cada participante. Este é um índice da tendência de resposta ao longo dos 5 PBR. Participantes que se encontrem na diagonal esquerda do gráfico responderam sistematicamente abaixo da resposta bayesiana. Participantes que se encontram na parte direita do gráfico responderam sistematicamente acima da resposta bayesiana. Podemos observar uma tendência geral para os participantes responderem sistematicamente acima da resposta bayesiana individual, ou seja, que a estratégia de resposta mais comum é a utilização da descrição.

Por fim, os participantes situados na linha central do gráfico tiveram um tipo de viés em alguns PBR e o outro tipo de viés nos restantes PBR, ou seja foram inconsistentes com a sua estratégia de resposta. Podemos observar que houve poucos participantes nesta situação, porque a maioria apresentou uma tendência de resposta ao longo dos cinco PBR.

Discussão:

Nesta experiência todas as hipóteses avançadas foram confirmadas e os resultados obtidos permitem-nos responder às duas perguntas que colocámos na introdução desta experiência.

Em primeiro lugar encontramos evidência de que alguns participantes parecem estar a utilizar uma estratégia de integração cognitiva em PBR. Mais de 10% dos participantes obtiveram um viés absoluto médio inferior a 10%, o que significa que utilizaram sistematicamente uma estratégia de resposta de integração bayesiana ao longo dos 5 PBR.

Em segundo lugar a proximidade das suas respostas à resposta bayesiana correlaciona-se com o CRT, uma outra medida de racionalidade. Estes dados apoiam os resultados obtidos por Hoppe e Kusterer (2011). Estes autores verificaram que participantes com baixo desempenho em CRT apresentam uma tendência a serem mais enviesados em problemas onde existe uma resposta normativa. Este enviesamento pode ocorrer tanto no sentido da descrição como no sentido dos base-rates, o que apoia a ideia de que a utilização de uma estratégia de integração cognitiva envolve uma capacidade de inibir as respostas intuitivas e de dar uma resposta ponderada.

Os resultados obtidos nesta experiência apoiam a ideia de que é possível estudar a capacidade de integração cognitiva em PBR com a utilização do paradigma proposto nesta tese.

Experiência 2: Escolha Livre

Introdução: Os resultados obtidos na primeira experiência indicam que, embora encontremos participantes com respostas sistematicamente enviesadas, também encontramos participantes que respondem próximo da resposta bayesiana. Nesta experiência procurámos replicar estes resultados e mostrar que esta capacidade de fazer uma integração bayesiana não depende do material utilizado, reduzindo possíveis constrangimentos experimentais resultantes de ser o experimentador a definir o conteúdo dos problemas. Para este fim pedimos que os participantes para criarem os seus próprios problemas de base-rate.

Hipóteses:

- 1- Esperamos encontrar evidência da utilização de diferentes estratégias de resposta aos PBR. Por um lado, esperamos que os participantes sobrevalorizam o valor diagnóstico da descrição tal como previsto pela utilização de heurísticas judicativas de semelhança e disponibilidade (Kahneman & Tversky, 1973). Por outro lado, esperamos encontrar alguns participantes que sobrevalorizam a importância dos base-rates, dando respostas apresentam um efeito de conservadorismo (Phillips, 1966), possivelmente devido a uma heurística de ancoragem e ajustamento insuficiente. Finalmente, esperamos encontrar alguns participantes que utilizam uma estratégia de integração cognitiva, respondendo ao PBR de maneira próxima à prescrita pelo modelo normativo bayesiano.

Método:

Participantes: 256 pessoas responderam a um questionário distribuído online através da plataforma Qualtrics.

Materiais: O questionário era composto por 1 PBR exclusivo com um base-rate de 7%. Este valor foi escolhido por ter sido o valor de base-rate mais baixo utilizado na experiência 1.

Procedimento: Foi pedido aos participantes que pensassem numa característica “C” (descrição) que fosse frequente no grupo “A” (grupo típico) e pouco frequente no grupo “B” (grupo atípico). Depois foi pedida em separado a probabilidade de uma pessoa do grupo A e uma pessoa do grupo B terem a característica C. Por fim, foi apresentado um PBR onde o base-rate do grupo A era 7%. Tanto os grupos A e B como a característica C foram apresentados no problema tal como os participantes introduziam no questionário. No final do questionário os participantes recebiam feedback sobre o seu desempenho onde era apresentada a sua resposta, a resposta bayesiana e a diferença entre as duas.

Design: Nesta experiência tivemos duas variáveis dependentes principais: a resposta ao PBR e diagnosticidade da característica.

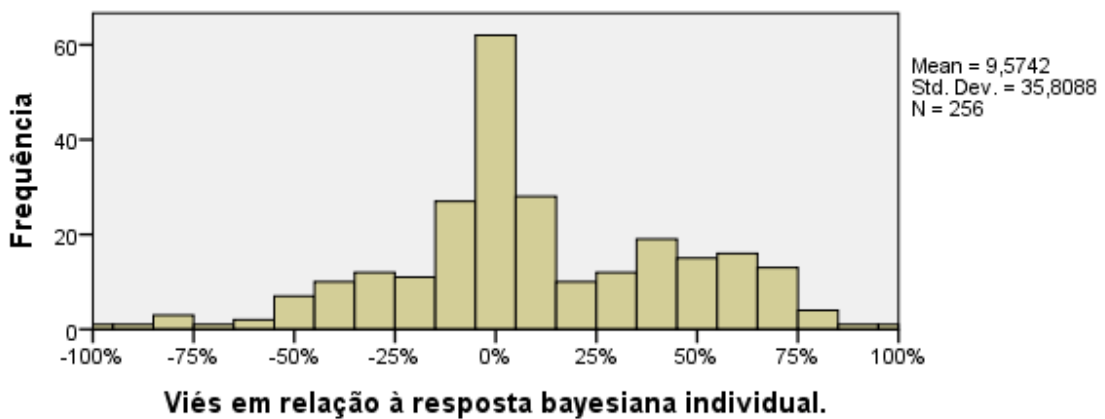
Resultados:

Os grupos escolhidos mais escolhidos pelos participantes foram profissões e a característica mais escolhida pelos participantes foram características psicológicas. Os gráficos que representam as categorias mais escolhidas pelos participantes encontram-se em anexo.

A maioria dos participantes respondeu adequadamente à quantificação da tipicidade da descrição C no grupo A e B, ou seja, para mais de 95% dos participantes a característica C era diagnóstica do grupo A. Uma análise detalhada dos valores da diagnosticidade existe em anexo.

No gráfico 4 podemos observar a frequência dos vieses em relação à resposta bayesiana.

Gráfico 4: Frequência do viés à resposta bayesiana individual.

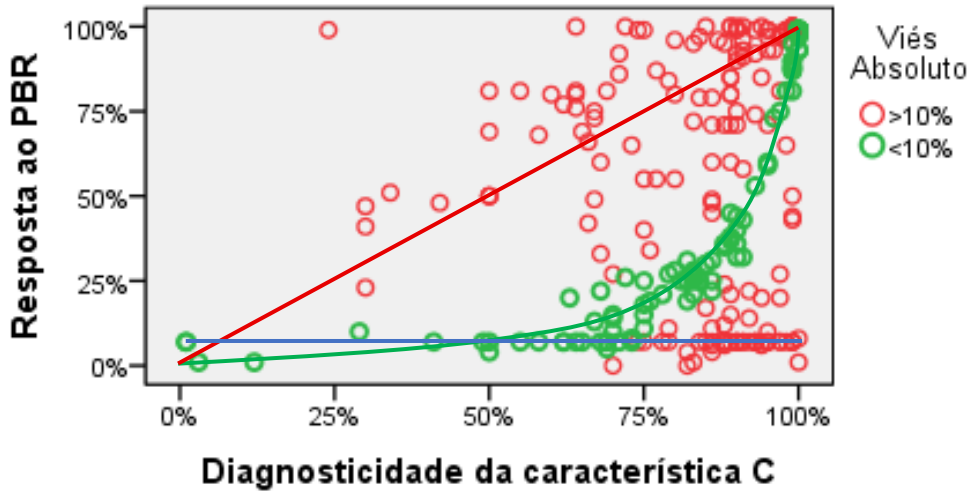


Aqui observamos um resultado semelhante ao gráfico da experiência 1. Existe uma tendência para os participantes se aproximarem da resposta bayesiana (barra central), mas simultaneamente existe uma tendência geral para os participantes responderem acima da resposta bayesiana e uma tendência menor para os participantes responderem abaixo da resposta bayesiana. Nesta experiência os participantes apresentando um viés médio de 9,6% e um viés absoluto médio de 27,4%. Dado que existe um grande número de participantes que respondeu próximo à resposta bayesiana, seria de esperar que à medida que a diagnosticidade da descrição aumentasse a resposta aumentasse de modo curvilíneo, tal como esperado pela fórmula de Bayes.

No gráfico 5 podemos observar um gráfico de dispersão de pontos onde cada ponto representa um participante. No eixo horizontal está representada a diagnosticidade da descrição do PBR. No eixo vertical está representada a resposta que o participante deu ao PBR. Cada ponto encontra-se colorido de verde se a resposta se encontra a 10% de distância da resposta bayesiana e colorido de vermelho no caso contrário. Para uma das três estratégias de resposta identificadas na experiência 1 encontram-se identificadas no gráfico. A estratégia da integração bayesiana encontra-se identificada por uma curva verde, a estratégia da utilização da diagnosticidade da descrição por uma linha vermelha e a estratégia da utilização do base-rate por uma linha azul.

A aglomeração de pontos na proximidade destes marcadores é vista como evidência da utilização destas estratégias de resposta.

Gráfico 5: Respostas ao PBR em função da diagnosticidade da característica C.



Observámos que existe uma grande aglomeração de respostas próximo à resposta bayesiana (curva verde) indicando que muitos participantes estão a utilizar esta estratégia de resposta. Encontrámos também muitos participantes que estão a responder utilizando apenas o base-rate (linha azul). Por fim não encontramos evidência de que os participantes estejam a utilizar a diagnosticidade da descrição como estratégia. Embora haja um aglomerado de pontos na parte superior do gráfico que correspondem a um viés positivo forte, estes pontos não se aproximam da reta vermelha, o que significa que este enviesamento não é facilmente explicado pelo valor da diagnosticidade.

Discussão:

O padrão de respostas nesta experiência confirma o padrão de respostas da experiência número 1, ou seja, existe evidência que existem três estratégias distintas de resposta aos PBR: utilização da descrição, utilização dos base-rates e integração bayesiana. Nesta experiência conseguimos ver uma agregação de dados próximos dos valores esperados para a estratégia de utilização dos base-rates e de integração bayesiana. No entanto, não encontramos agregação à volta dos valores esperados para a estratégia da utilização da diagnosticidade da descrição. Ou seja, os nossos dados não apoiam a ideia de que uma estratégia de utilização da descrição corresponda a uma resposta de acordo com a sua diagnosticidade. Pelo contrário, os nossos dados apontam que os processos envolvidos nesta estratégia não são contínuos, como por exemplo uma estratégia bayesiana. É possível que estas respostas que se situam acima da resposta bayesiana possam ser o resultado da utilização de heurísticas de semelhança e da utilização da escala de resposta de modo qualitativo, levando a respostas acima do valor da diagnosticidade da descrição.

Experiência 3: Base-Rates Adaptados

Introdução: Na primeira experiência o desempenho em PBR correlacionou-se com o desempenho no CRT, uma medida concorrente de racionalidade. Nesta experiência procuramos replicar estes resultados. Para além disso tentamos correlacionar o desempenho em PBR com uma medida de inteligência fluida, o desempenho numa bateria das Matrizes Progressivas de Raven (MPR; Bilker et al., 2012). Esta bateria de testes é uma medida geral de inteligência (fator g) e relaciona-se com a capacidade de fazer a integração cognitiva de padrões abstratos de imagens. Para além disso, correlacionamos o desempenho em PBR com os valores obtidos nas escalas Faith in Intuition (FI; Epstein et al., 1996) e Need for Cognition (NFC; Cacioppo & Petty, 1982). Estas escalas são medidas do estilo de pensamento e encontram-se tipicamente correlacionadas com o desempenho em tarefas de raciocínio. Por fim, como variável de controlo correlacionamos o desempenho em PBR com a média do secundário (MS) de cada participante.

Hipóteses:

- 1- Esperamos encontrar evidência de três estratégias de resposta: sobrevalorização da descrição, sobrevalorização dos base-rates, e integração cognitiva.
- 2- O desempenho nos PBR estará correlacionado com outras medidas de racionalidade, inteligência e disposição cognitiva.

Método:

Participantes: 245 pessoas responderam a um questionário distribuído online através da plataforma Qualtrics.

Materiais: O questionário era composto por 6 PBR, por 6 problemas CRT, por 9 problemas das Matrizes Progressivas de Raven (MPR; Bilker et al., 2012), 5 itens da escala Need for Cognition (NFC; Cacioppo & Petty, 1982) e Faith in Intuition (FI ; Epstein et al., 1996). Com a exceção das MPR estes materiais encontram-se em anexo.

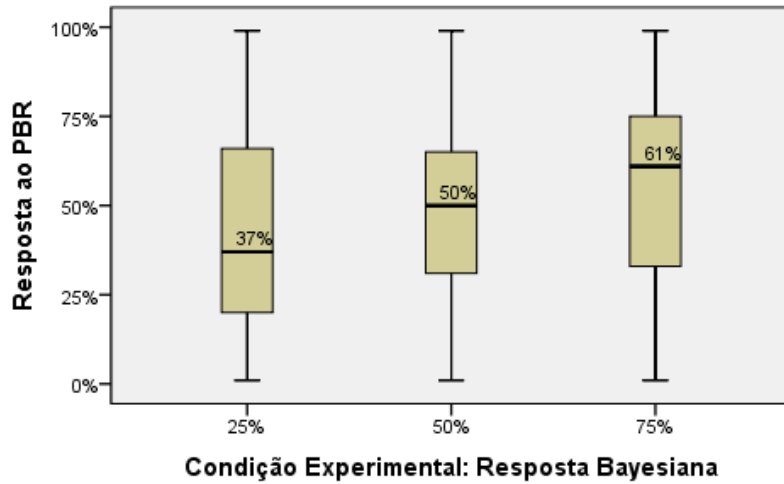
Procedimento: Os participantes receberam 2 PBR onde a resposta bayesiana era 25%, 2 PBR onde a resposta bayesiana era 50% e 2 PBR onde a resposta era 75%. Os base-rates eram calculados online em função diagnosticidade subjetiva das descrições. A diagnosticidade das descrições era pedida no início do questionário (condição “separado”) ou imediatamente antes do problema de modo intercalado (condição “junto”).

Design: Design misto 3 x 2: versão do PBR (25%, 50% e 75%) intraparticipante, proximidade da quantificação da diagnosticidade (separado ou junto) interparticipante. Variáveis dependentes: Desempenho nos PBR, no CRT, valores nas escalas NFC, FI e MS.

Resultados:

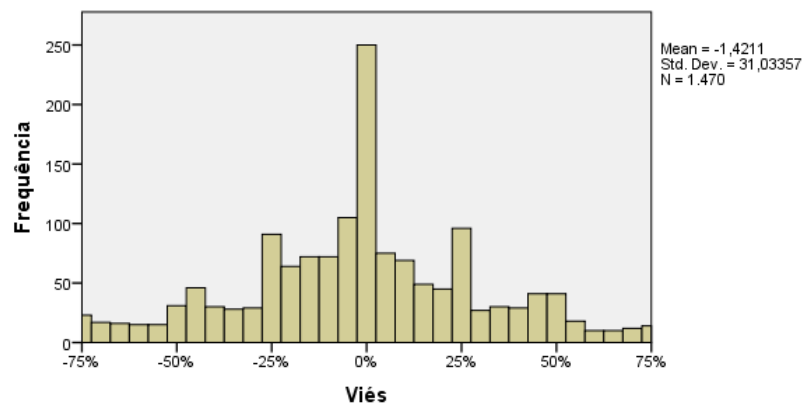
Observámos que as respostas ao PBR variaram em função da condição experimental. Para garantirmos que a resposta bayesiana era sempre 25%, 50% ou 75%, um pequeno número de respostas onde a diagnosticidade da descrição era extrema (0% ou 100%) foram retiradas da análise. No gráfico 6 podemos observar a distribuição das respostas por quartis em função da versão do PBR apresentada ao participante.

Gráfico 6: Efeito da condição experimental nas respostas dadas pelos



Observámos que a mediana das respostas ao PBR varia em função da condição experimental, no entanto varia menos do que na experiência 1. Na condição 25% a mediana está acima do esperado, na condição 50% a medianas das respostas está no valor esperado e na condição 75% a mediana está abaixo do valor esperado. Usando uma ANOVA, encontramos diferenças significativas na média das respostas dadas aos PBR nas três versões do PBR, $F(2, 1467) = 27, p < 0,001$. O output desta análise e os gráficos da distribuição das respostas em cada versão do PBR encontram-se em anexo. Analisamos a proximidade das respostas dos participantes à resposta bayesiana individual. No gráfico 7, podemos observar a frequência dos vieses à resposta bayesiana em todas as versões dos PBR. Observámos um pico de respostas em torno da resposta bayesiana (0%), mas também podemos observar um pico de respostas nos valores -25% e 25%. O viés médio foi de -1,4% e o viés absoluto médio foi de 23,4%.

Gráfico 7: Frequência do viés à resposta bayesiana.

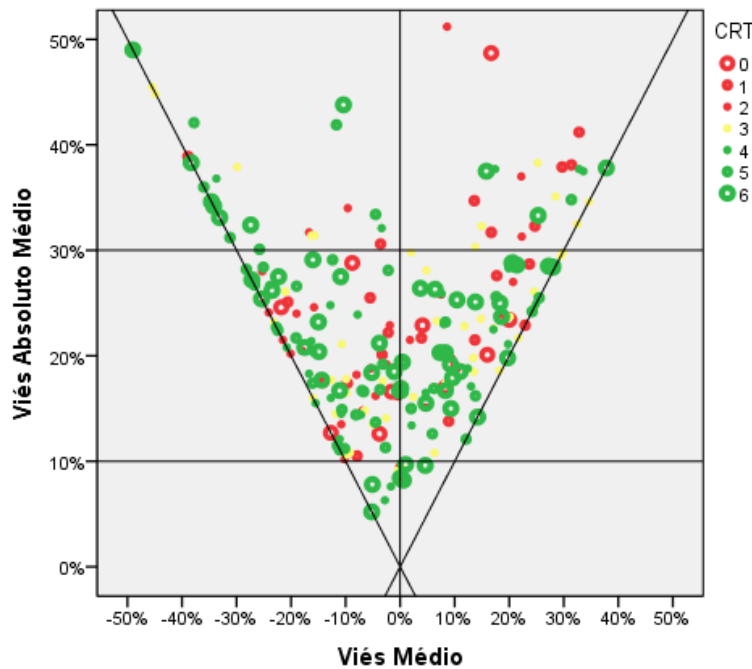


Analisámos o efeito da manipulação da proximidade da medição da diagnosticidade no desempenho nos PBR. Era esperado que quando a diagnosticidade fosse medida no início do questionário que isso piorasse o desempenho nos PBR, mas não encontramos esse efeito ($t = -0,3$; $p > 0,7$). A representação gráfica e o output do teste estatístico encontram-se em anexo.

De seguida analisámos se o desempenho nos PBR se correlacionava com o desempenho no CRT, replicando os resultados obtidos na experiência 1. Para além disso, correlacionámos o desempenho nos PBR com o desempenho nas Matrizes Progressivas de Raven (MPR), com a escala Need For Cognition (NFC), Faith in Intuition (FI) e a Média do Secundário (MS).

Observámos que, à exceção da média do secundário ($r = 0,157$; $p = 0,014$), não se encontraram correlações significativas entre o desempenho em PBR e outras medidas à exceção da média do secundário (tabela em anexo). Mesmo não tendo encontrado estas correlações decidimos recriar a mesma análise gráfica que a utilizada na experiência 1 (gráfico 8).

Gráfico 8: Desempenho em PBR VS Desempenho no CRT.



Os resultados desta experiência são bastante diferentes dos resultados da experiência 1. Nesta experiência o viés médio foi de -1,4% e o viés absoluto médio foi de 23,4%. Na experiência 1 o viés médio foi de 9% e o viés absoluto médio foi de 21%.

Podemos ver que, ao contrário da experiência 1, houve muitos participantes com um bom desempenho no CRT a responder sistematicamente abaixo da resposta bayesiana, provavelmente a responder sistematicamente utilizando os base-rates. Esta pode ser uma explicação para a falta de correlações entre o viés absoluto médio em PBR com outras medidas concorrentes de racionalidade, inteligência e estilo cognitivo.

De modo a confirmar a existência de três tipos de estratégia de resposta criámos um conjunto de gráficos de dispersão, onde cada ponto representa uma resposta a um PBR. A interpretação destes gráficos é semelhante à do gráfico 5 da experiência 2. As análises foram repetidas com margens progressivamente mais conservadoras (15%, 10%, 5%).

No gráfico 9, 10 e 11 podemos observar a existência de respostas que correspondem às 3 estratégias quando a resposta bayesiana era 25%.

Gráfico 9: Resposta ao PBR em função da diagnosticidade – Margens de 15%

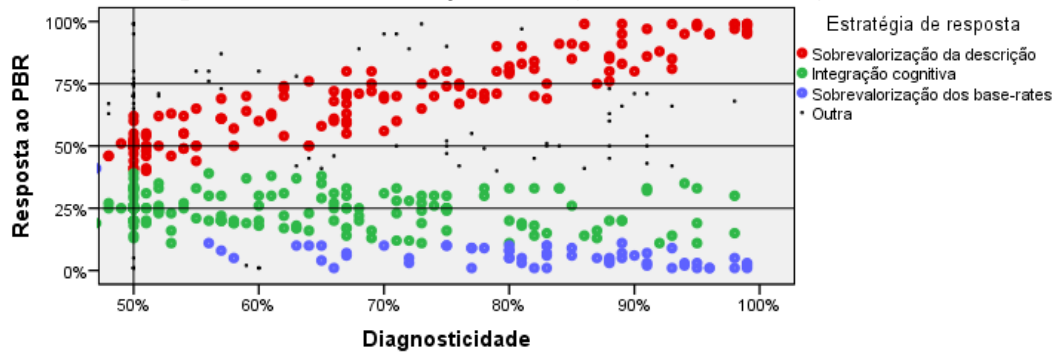


Gráfico 10: Resposta ao PBR em função da diagnosticidade – Margens de 10%

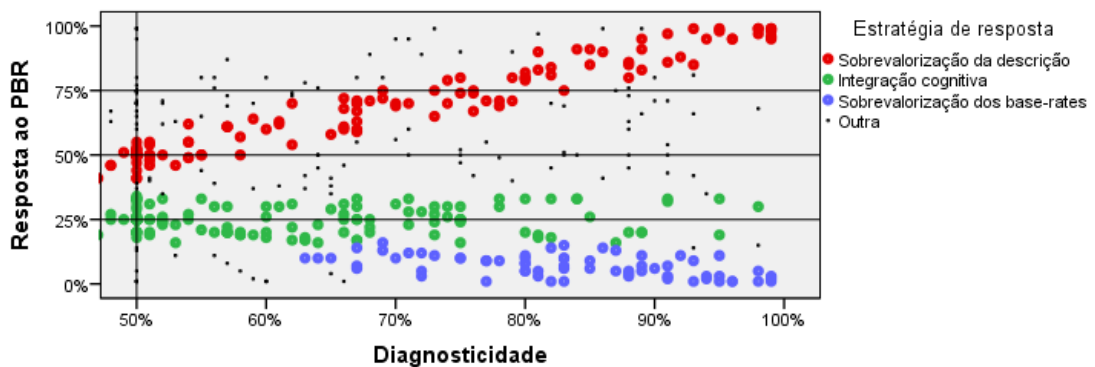
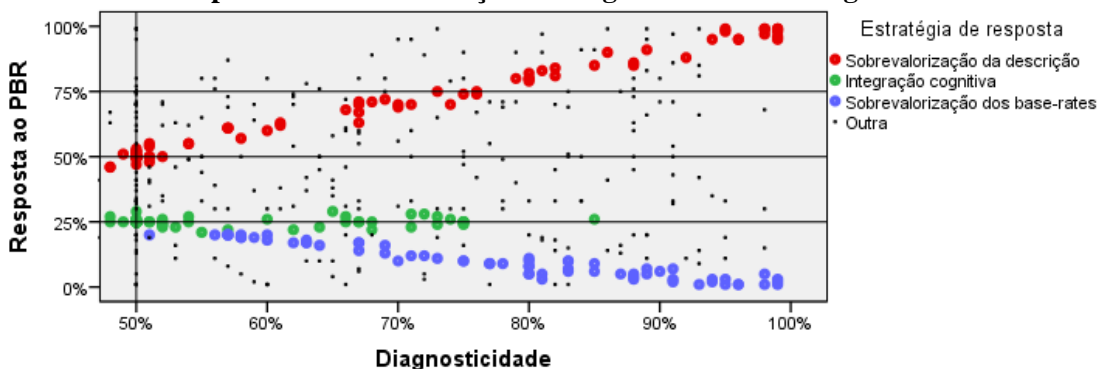


Gráfico 11: Resposta ao PBR em função da diagnosticidade – Margens de 5%



Nestes três gráficos podemos observar que, mesmo utilizando um critério conservador podemos identificar as três estratégias de resposta. Observamos também que à medida que a diagnosticidade aumenta, diminuem o número de respostas próximas à resposta bayesiana.

Nos gráficos 12, 13 e 14 podemos observar a existência de respostas que correspondem às 3 estratégias quando a resposta bayesiana era 50%.

Gráfico 12: Resposta ao PBR em função da diagnosticidade – Margens de 15%

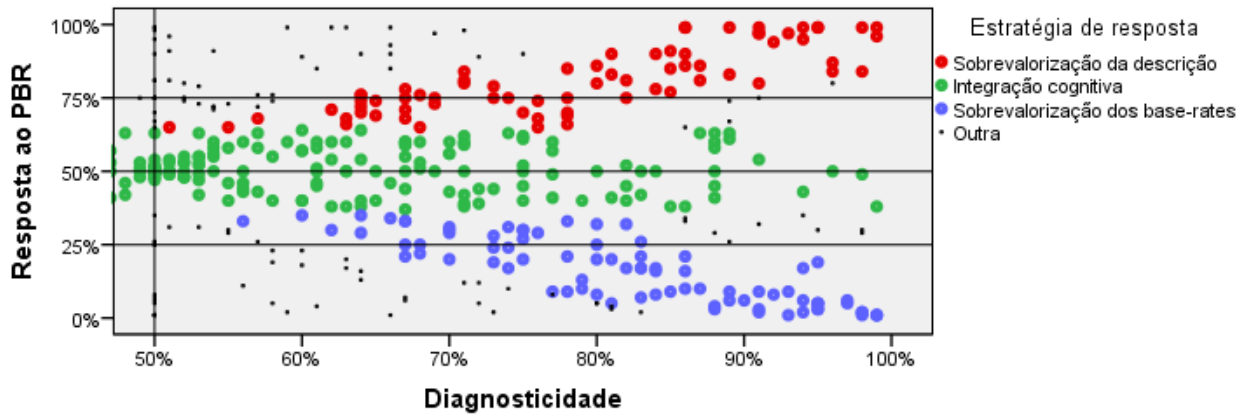


Gráfico 13: Resposta ao PBR em função da diagnosticidade – Margens de 10%

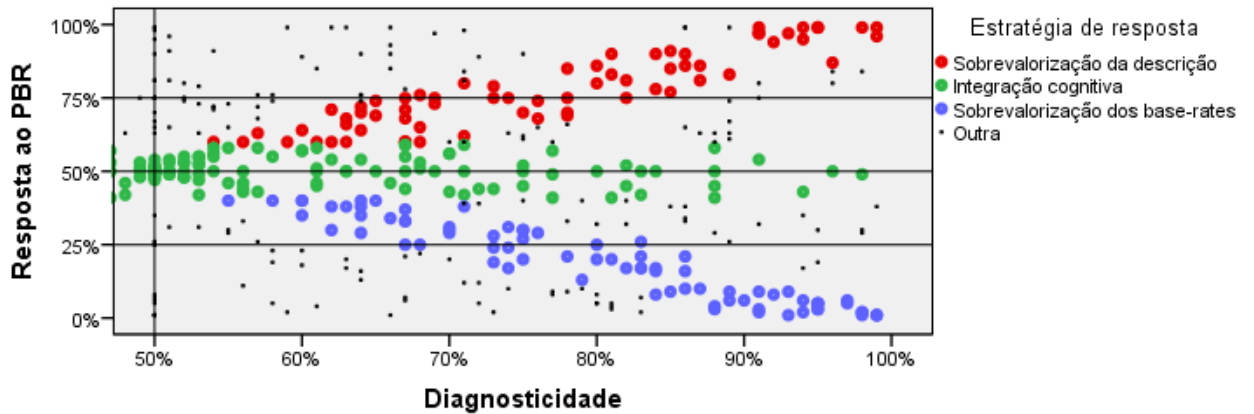
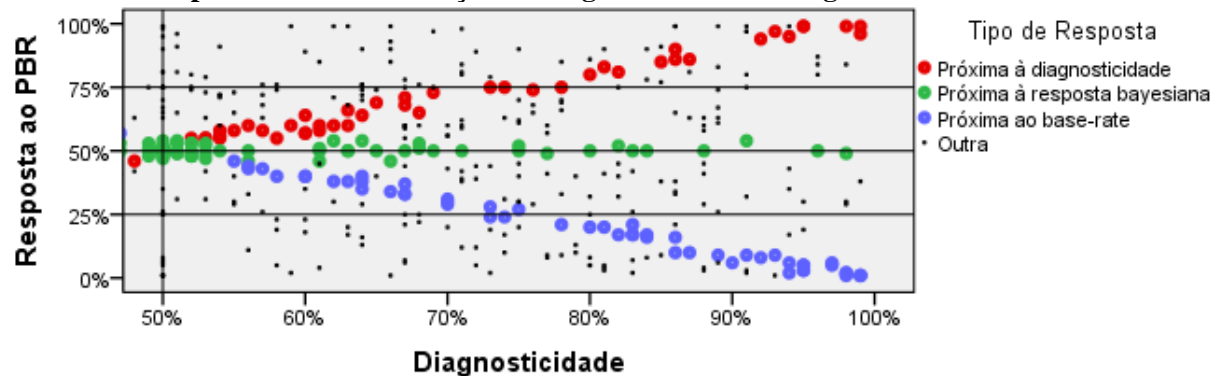


Gráfico 14: Resposta ao PBR em função da diagnosticidade – Margens de 5%



Nestes três gráficos podemos observar que, à semelhança dos gráficos 9, 10 e 11, mesmo utilizando um critério conservador podemos identificar as três estratégias de resposta. Observamos também que à medida que a diagnosticidade aumenta, diminuem o número de respostas próximas à resposta bayesiana.

Nos gráficos 15, 16 e 17 podemos observar a existência de respostas que correspondem às 3 estratégias quando a resposta bayesiana era 75%.

Gráfico 15: Resposta ao PBR em função da diagnosticidade da descrição – Margens 15%

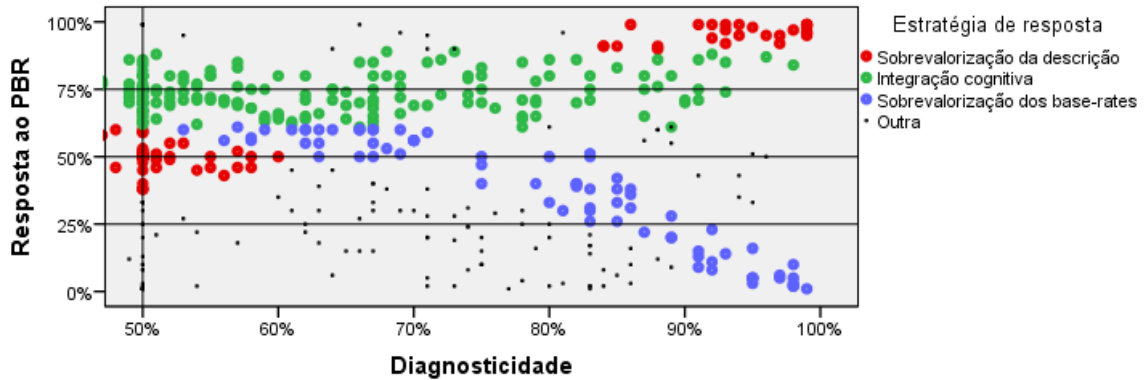


Gráfico 16: Resposta ao PBR em função da diagnosticidade da descrição – Margens 10%

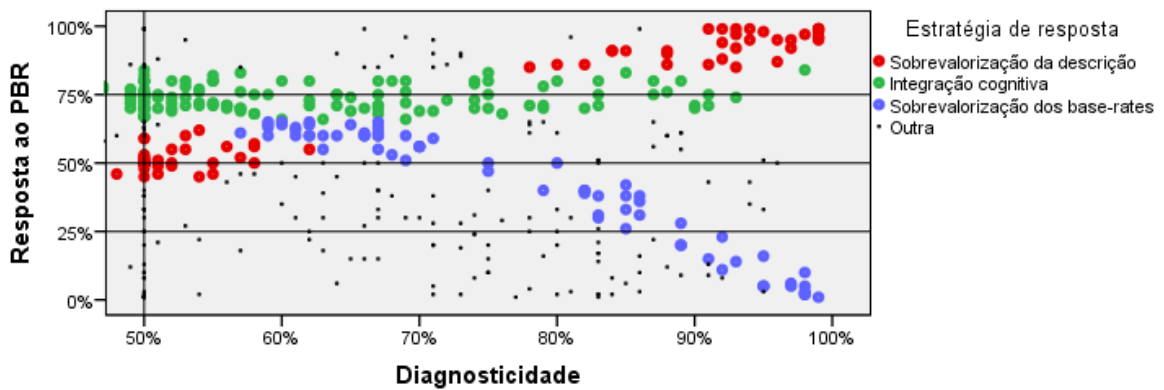
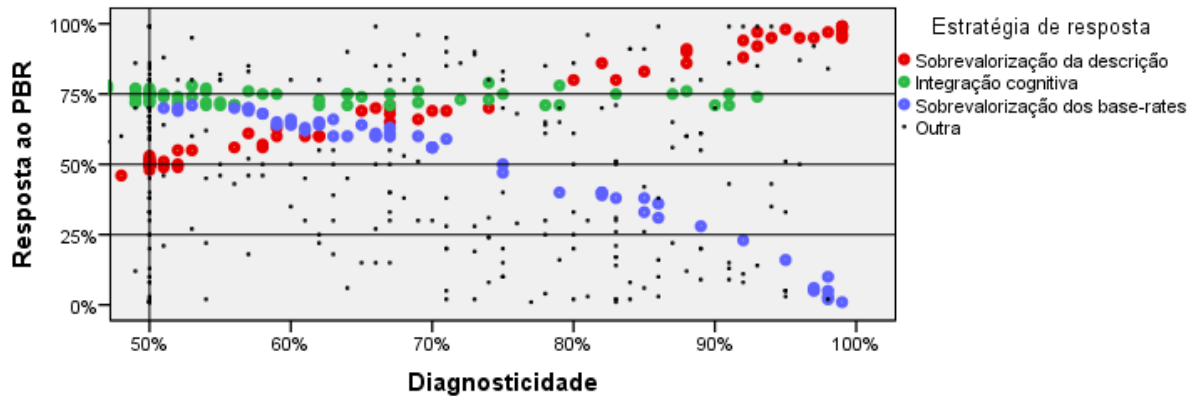


Gráfico 17: Resposta ao PBR em função da diagnosticidade da descrição – Margens 5%



Observamos que nesta condição existe uma sobreposição das estratégias de resposta, mas mesmo assim é possível identificar cada uma delas com um critério conservador de inclusão.

No geral, podemos observar que, quanto maior a diagnosticidade da descrição, mais difícil se torna a integração bayesiana. Na condição 25% é mais dificuldade integrar as duas informações do que a condição 50% e 75%.

Discussão:

Efeito da condição experimental (resposta bayesiana) na resposta ao PBR: Os dados apontam para um efeito da condição experimental na resposta dada pelos participantes aos PBR, mostrando evidência que esta informação está a ser levada em conta por alguns participantes. Tal como esperado, existe uma maior dispersão dos dados (e viés associado) na condição 25% do que na condição 50%. Por fim, ao contrário dos resultados da experiência 1, nesta experiência a condição 75% teve uma dispersão de respostas semelhante à condição 25%.

Estratégias de resposta aos PBR: Embora as correlações com medidas complementares não tenham sido frutíferas, foi possível, mesmo assim, observar três estratégias de resposta diferentes: respostas próximas à diagnosticidade da descrição, respostas próximas à resposta bayesiana e respostas próximas ao base-rate apresentado.

Limitações e investigações futuras:

Ao contrário do esperado, não encontramos correlações significativas entre o desempenho em PBR e as outras medidas de raciocínio, inteligência e estilo cognitivo. Curiosamente, a única medida que se correlacionou significativamente com o desempenho em PBR nesta experiência foi a média do secundário ($r = 0,157$; $p = 0,014$).

Esta falha em detetar correlações significativas deveu-se possivelmente a uma diferença significativa no design experimental. Nesta experiência em vez de utilizarmos base-rates fixos apresentamos base-rates adaptados a cada participante de modo a que a resposta bayesiana fosse fixa. Esta manipulação baixou o nível de desempenho nos PBR e piorou a qualidade desta medida.

Vários resultados diferentes que eram esperados não foram obtidos nesta experiência. Por exemplo, a manipulação da proximidade da quantificação da diagnosticidade da descrição ao PBR não teve qualquer efeito significativo nas respostas aos PBR. Tanto os participantes que quantificaram a descrição do PBR em conjunto antes ou imediatamente antes dos PBR tiveram o mesmo desempenho. Para além disto, o desempenho em PBR não se correlacionou com nenhuma outra medida complementar (CRT, MPR, NFC e FI) com a exceção da média do secundário que deveria funcionar como variável de controlo. Esta falta de correlação persistiu mesmo perante a exclusão de outliers e utilizando contrastes planeados. Estes resultados podem ser um artefacto do paradigma de base-rates adaptados ser mais difícil do que o paradigma de base-rates fixos. No paradigma desta experiência, quando a diagnosticidade é muito alta o base-rate apresentado ser muito baixo, o que pode levar a mais erros por parte dos participantes. Enquanto a média do viés absoluto médio na primeira experiência foi 21%, nesta experiência foi 23,3%, o que significa que houve mais desvio à resposta bayesiana.

Experiência 4: Base-Rates Adaptados - Replicação

Introdução: Insatisfeitos com os resultados da última experiência, replicamos o questionário 3 com melhorias metodológicas e uma amostragem de participantes mais rigorosa. Com esta replicação esperamos obter correlações significativas entre o desempenho nos PBR e outras medidas de racionalidade, inteligência e disposição cognitiva.

Hipóteses:

- 1- As alterações metodológicas introduzidas nesta experiência irão melhorar o desempenho em PBR em relação à experiência anterior.
- 2- Iremos encontrar três tipos de respostas: respostas que utilizam exclusivamente os base-rates, respostas que utilizam exclusivamente a diagnosticidade da descrição e respostas que integram as duas informações presentes nos PBR.
- 3- O desempenho nos PBR estará correlacionado com outras medidas de racionalidade, inteligência e disposição cognitiva.

Método:

Participantes: 181 participantes responderam a um questionário através da plataforma Qualtrics. Destes 181 participantes, 78 responderam em condições laboratoriais.

Materiais: O material utilizado nesta experiência com igual ao da experiência anterior com a introdução de um PBR de treino.

Procedimento: O procedimento desta experiência foi em tudo igual à anterior com as seguintes melhorias metodológicas:

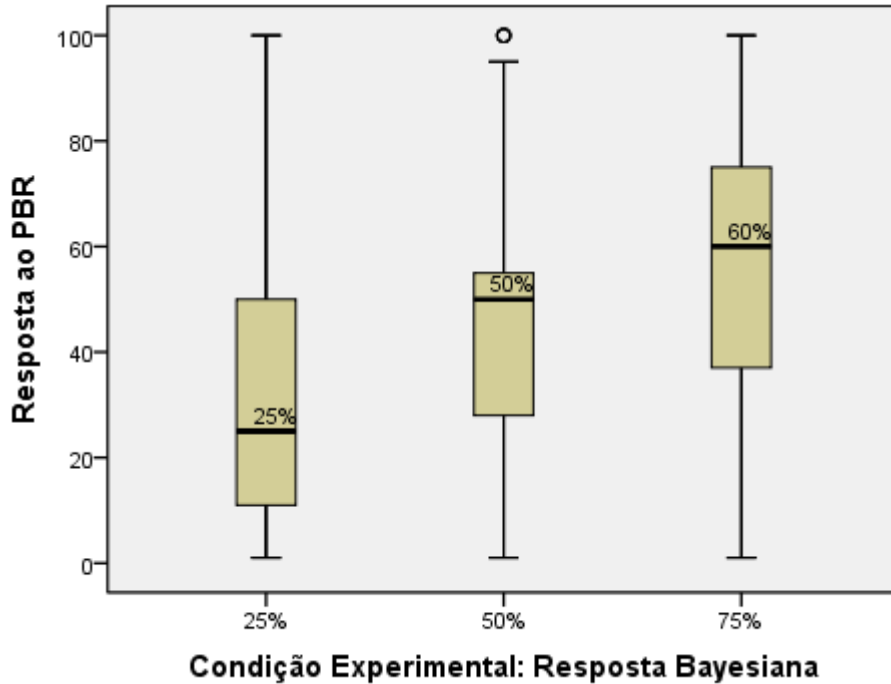
- 1- Apresentação de instruções mais claras sobre os objetivos da experiência de modo a dar ênfase à utilização das duas informações apresentadas.
- 2- Apresentação de um PBR de treino onde o base-rate apresentado era 50%.
- 3- Medição da diagnosticidade da descrição imediatamente antes do PBR.
- 4- Apresentação da diagnosticidade da descrição no enunciado do PBR.

Design: Variável independente: Versão do PBR (25%, 50% e 75%) intraparticipante.
Variáveis dependentes: Desempenho nos PBR, no CRT, valores nas escalas NFC, FI e MS.

Resultados:

Observámos que as respostas ao PBR variaram em função da condição experimental. Para garantirmos que a resposta bayesiana era sempre 25%, 50% ou 75%, um pequeno número de respostas onde a diagnosticidade da descrição era extrema (0% ou 100%) foram retiradas da análise. No gráfico 18 podemos observar a distribuição das respostas por quartis em função da versão do PBR apresentada ao participante.

Gráfico 18: Efeito da condição experimental nas respostas aos PBR.



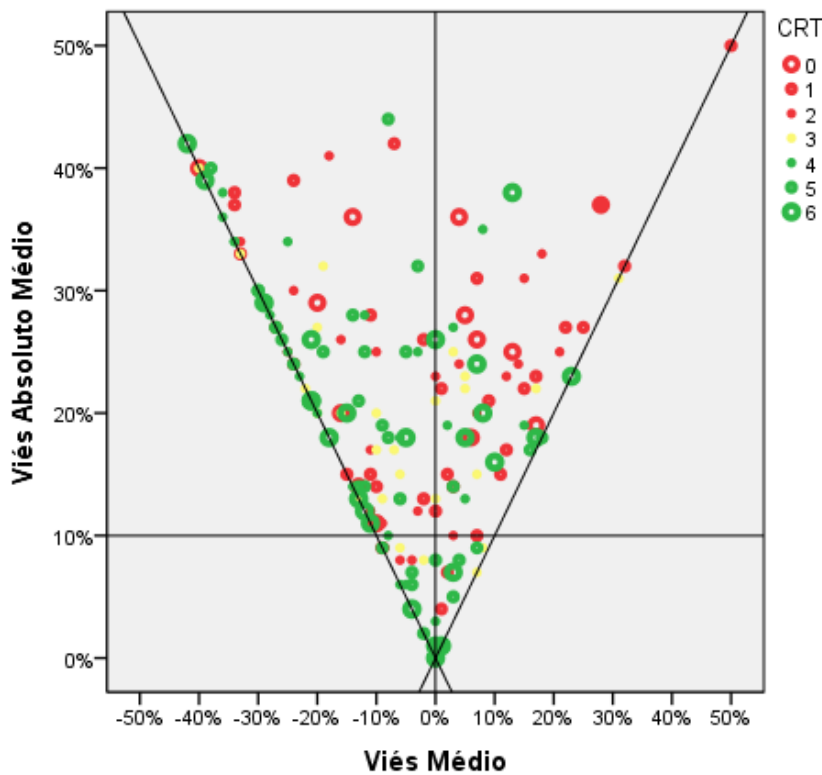
Enquanto na condição 25% e 50% estão as medianas estão no valor esperado, na condição 75% está abaixo do valor esperado. Usando uma ANOVA, encontramos diferenças significativas na média das respostas dadas aos PBR nas três versões do PBR, $F(2, 1054) = 71$, $p < 0,001$. O output desta análise e os gráficos da distribuição das respostas em cada versão do PBR encontram-se em anexo. O efeito da condição experimental foi mais forte na replicação do que na experiência 3, pois o valor do teste F passa de 27 para 71. Este é um resultado muito importante pois é um indicador importante de que as melhorias metodológicas introduzidas aumentaram o efeito da condição experimental. Em comparação com a experiência 3 o viés médio desceu de -1,4% para -6,6% e o viés absoluto médio desceu de 23,4% para 20,5%. Estes resultados indicam que embora os participantes nesta experiência em média se tenham aproximado mais da resposta bayesiana, paradoxalmente tiveram uma tendência de resposta mais forte no sentido dos base-rates.

O desempenho e tendência de resposta médio da amostra recolhida em laboratório e da amostra recolhida fora do laboratório encontra-se em anexo.

De seguida analisámos se o desempenho nos PBR se correlacionava com o desempenho no CRT, replicando os resultados obtidos na experiência 1. Para além disso, correlacionámos o desempenho nos PBR com o desempenho nas Matrizes Progressivas de Raven (MPR), com a escala Need For Cognition (NFC), Faith in Intuition (FI) e a Média do Secundário (MS). O desempenho nos PBR correlacionou-se significativamente com CRT ($r = 0,169$; $p = 0,023$), marginalmente com o desempenho nas MPR ($r = 0,133$; $p = 0,074$) e não significativamente com as restantes medidas ($p > 0,2$) (tabela de correlações em anexo).

No gráfico 19 podemos observar a mesma análise gráfica utilizada na experiência 1 onde podemos visualizar a relação entre o desempenho nos PBR e no CRT.

Gráfico 19: Desempenho em PBR VS Desempenho no CRT.



Ao contrário da experiência 3, nesta experiência já obtivemos bastantes participantes com um bom desempenho em PBR (viés absoluto médio $< 10\%$). Estes participantes responderam em média a 4 problemas de CRT corretamente enquanto os outros participantes responderam corretamente em média a 3. Para além desta correlação existe um efeito de lateralização, os participantes que responderam sistematicamente com um viés no sentido dos base-rates (diagonal esquerda) apresentam um melhor desempenho que os restantes no CRT (cor verde). A correlação entre o viés médio e o desempenho no CRT é significativa ($r = -0,16$; $p < 0,03$). Este é um resultado já existia como tendência na experiência 3, mas não era significativa e possivelmente também existia na experiência 1, mas não havia participantes suficientes a ter uma tendência de resposta sistematicamente no sentido dos base-rates.

De modo a confirmar a existência de três tipos de estratégia de resposta criámos um conjunto de gráficos de dispersão, onde cada ponto representa uma resposta a um PBR. A interpretação destes gráficos é semelhante à do gráfico 5 da experiência 2. As análises foram repetidas com margens progressivamente mais conservadoras (15%, 10%, 5%).

No gráfico 20, 21 e 22 podemos observar a existência de respostas que correspondem às 3 estratégias quando a resposta bayesiana era 25%.

Gráfico 20: Resposta ao PBR em função da diagnosticidade – Margens de 15%

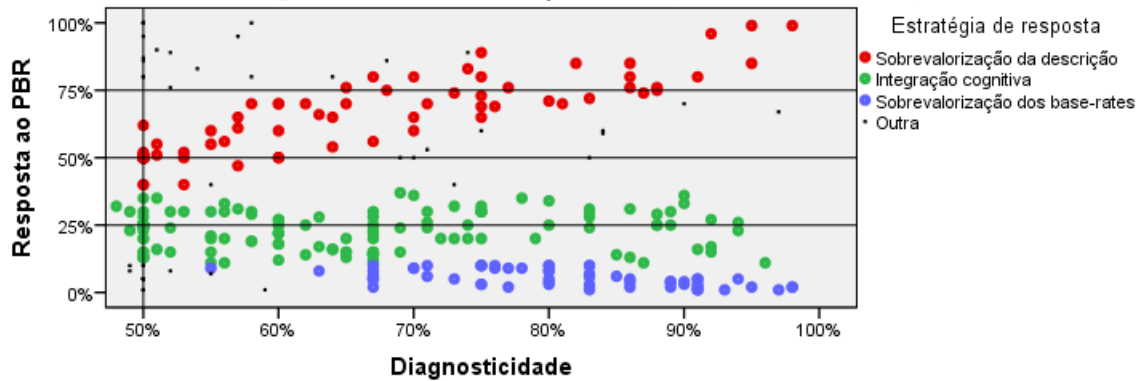


Gráfico 21: Resposta ao PBR em função da diagnosticidade – Margens de 10%

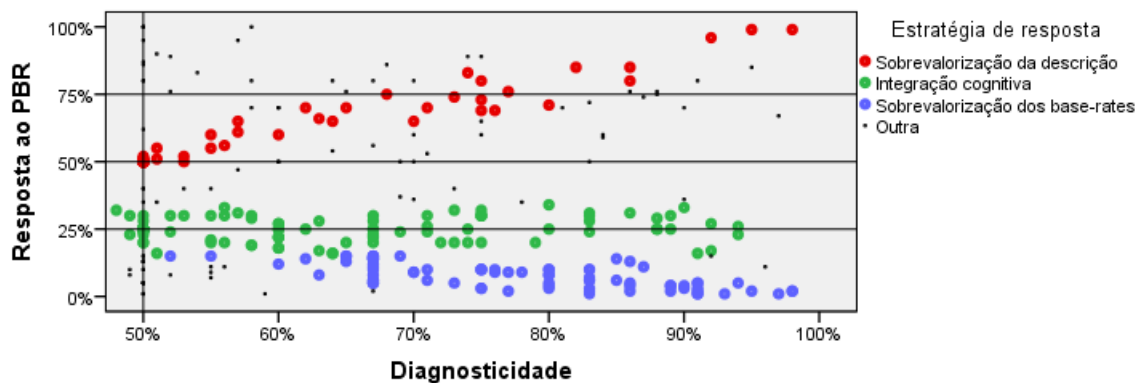
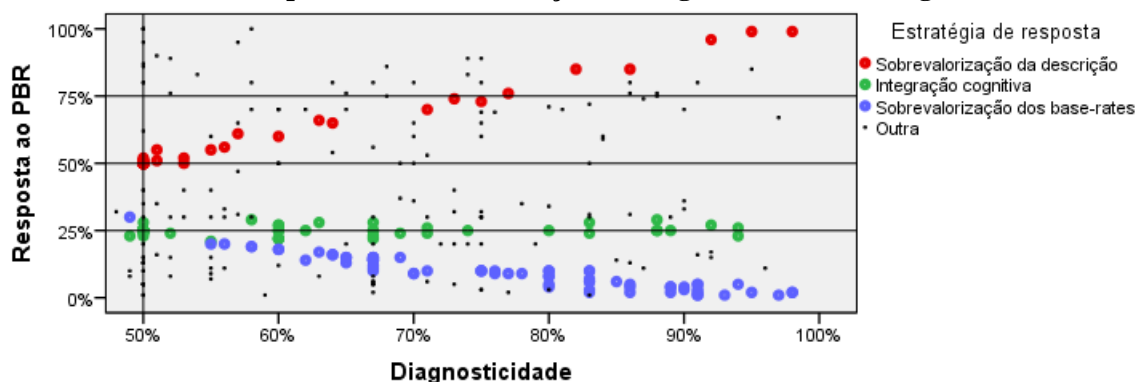


Gráfico 22: Resposta ao PBR em função da diagnosticidade – Margens de 5%



Nestes três gráficos podemos observar que, mesmo utilizando um critério conservador podemos identificar as três estratégias de resposta. Observamos também que à medida que a diagnosticidade aumenta, o número de respostas próximas à resposta bayesiana se mantém.

Nos gráficos 23, 24 e 25 podemos observar a existência de respostas que correspondem às 3 estratégias quando a resposta bayesiana era 50%.

Gráfico 23: Resposta ao PBR em função da diagnosticidade – Margens de 15%

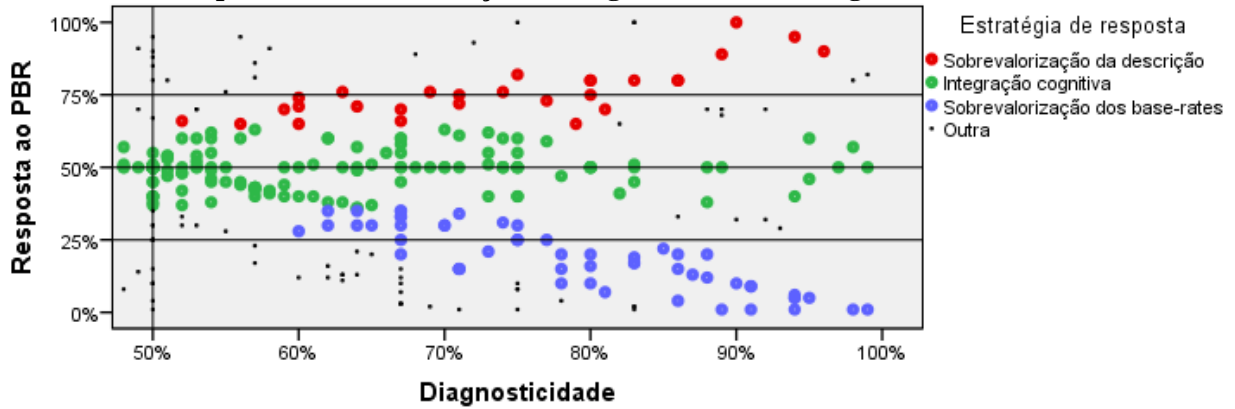


Gráfico 24: Resposta ao PBR em função da diagnosticidade – Margens de 10%

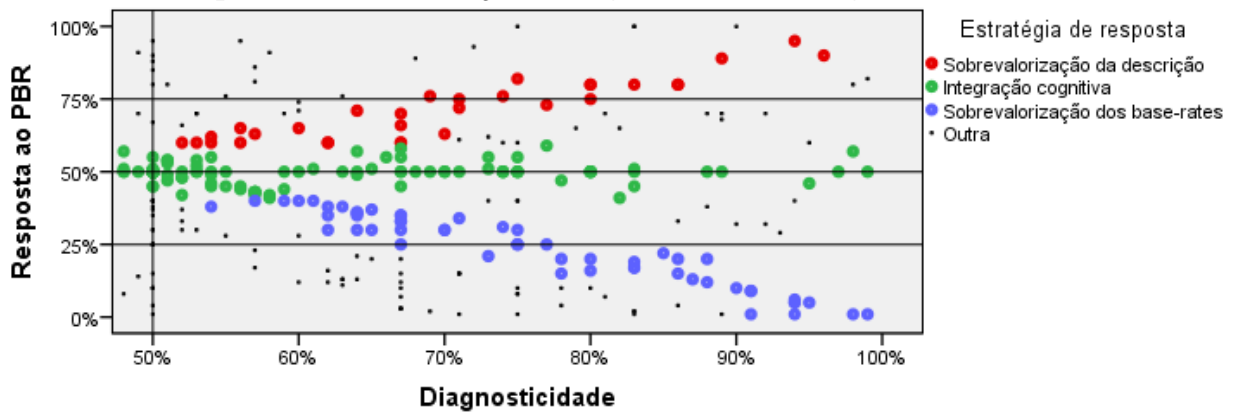
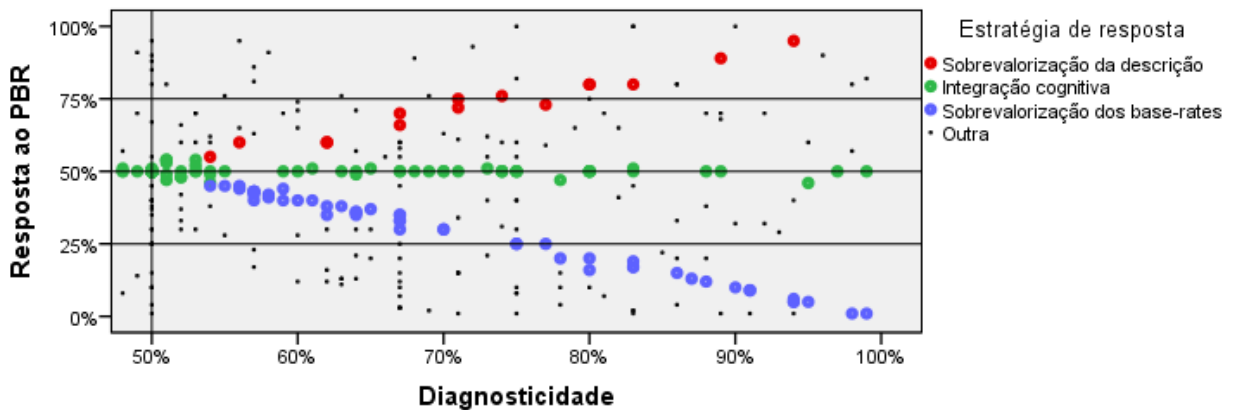


Gráfico 25: Resposta ao PBR em função da diagnosticidade – Margens de 5%



Nestes três gráficos podemos observar que, à semelhança dos gráficos 9, 10 e 11, mesmo utilizando um critério conservador podemos identificar as três estratégias de resposta. Observamos também que à medida que a diagnosticidade aumenta, diminuem o número de respostas próximas à resposta bayesiana.

Nos gráficos 26, 27 e 28 podemos observar a existência de respostas que correspondem às 3 estratégias quando a resposta bayesiana era 75%.

Gráfico 26: Resposta ao PBR em função da diagnosticidade da descrição – Margens 15%

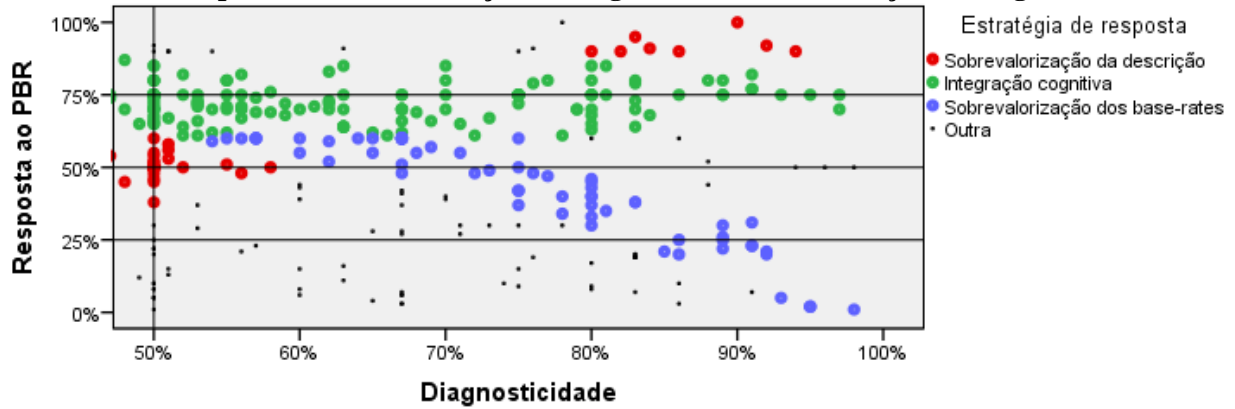


Gráfico 27: Resposta ao PBR em função da diagnosticidade da descrição – Margens 10%

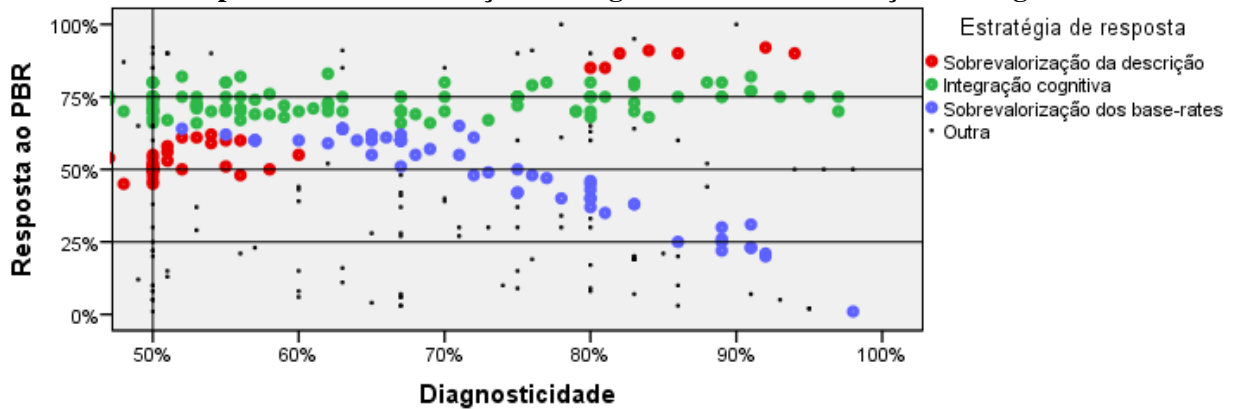
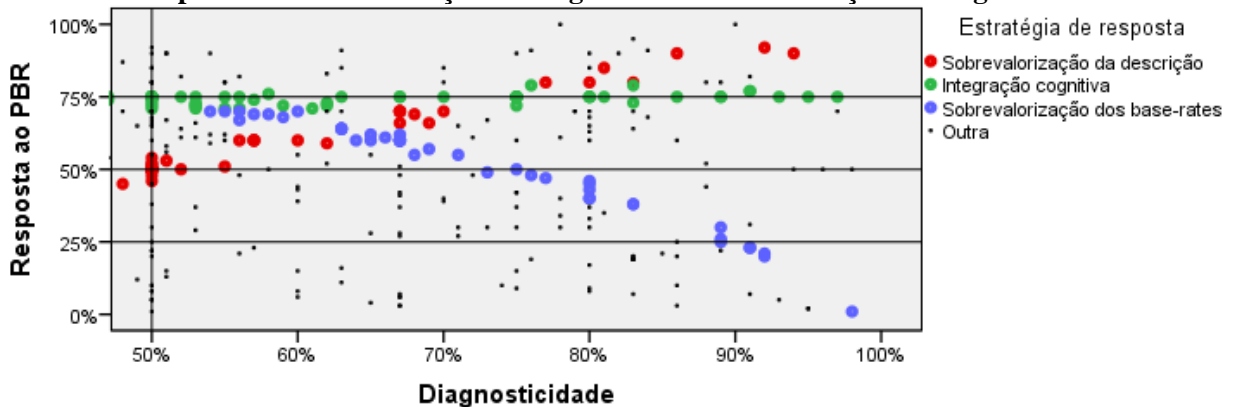


Gráfico 28: Resposta ao PBR em função da diagnosticidade da descrição – Margens 5%



Nestes três gráficos podemos observar que, à semelhança dos gráficos anteriores, mesmo utilizando um critério conservador podemos identificar as três estratégias de resposta. Observamos também que nesta condição existe uma sobreposição das estratégias de resposta.

No geral, podemos observar que, quanto maior a diagnosticidade da descrição, mais difícil se torna a integração bayesiana. A condição 25% apresenta mais dificuldade de integrar as duas informações do que a condição 50% e 75%.

Discussão:

Correlações com medidas complementares: Nesta experiência, ao contrário da experiência anterior, encontramos correlações significativas entre o desempenho em PBR e o desempenho no CRT e nas MPR. Estes dados corroboram os resultados obtidos na experiência 1. Em quase todas as análises o desempenho em PBR correlacionou-se mais significativamente com o CRT do que com as MPR. Isto pode significar que a capacidade para responder corretamente ao PBR depende mais da capacidade de inibir a primeira resposta intuitiva (capacidade medida pelo CRT) do que fazer o reconhecimento de padrões e combinações simbólicas com recurso da memória de trabalho (capacidade medida pela MPR). Por fim, ao contrário do esperado as escalas de NFC e FI não se correlacionaram significativamente com o desempenho em PBR.

Efeito das melhorias metodológicas: Nesta replicação obtivemos um conjunto de resultados significativamente melhores que os obtidos na experiência 3. Por exemplo, os dados apontam para um efeito da condição experimental na resposta dada pelos participantes aos PBR mais forte do que na experiência anterior, pois o valor do teste F sobe de 27 para 72. Atribuímos estes resultados se às melhorias metodológicas, nomeadamente a uma estratégia de recolha de participantes e execução do questionário de modo mais criterioso, a melhorias das instruções apresentadas e à apresentação da frequência subjetiva da característica no grupo típico e atípico no próprio PBR. Esta última alteração pode ter sido determinante na melhoria da capacidade dos participantes de darem respostas próximas da resposta bayesiana. Investigações futuras poderão beneficiar desta manipulação para melhorar a qualidade das respostas obtidas a PBR.

Limitações e investigações futuras:

Tendência de resposta no sentido dos base-rates: Em comparação com a experiência 3 o viés médio desceu de -1,4% para -6,6% e o viés absoluto médio desceu de 23,4% para 20,5%. Embora o desempenho tenha melhorado significativamente, a tendência de resposta desceu ainda mais no sentido da base-rates. Este é um resultado inesperado, pois contradiz a ideia de que existem mais participantes a fazer julgamentos enviesados no sentido das descrições do que participantes enviesados no sentido dos base-rates. Suspeitamos que esta mudança da tendência média de resposta se deva à mudança do paradigma da experiência 1 para a 3 e 4, no entanto não temos uma explicação simples para a mudança da tendência de resposta. Investigações futuras poderão tentar replicar estes resultados e tentar compreender a estratégia de resposta de um participante quando ele responde abaixo do valor dos base-rates.

Experiência 5: Paradigma de aprendizagem.

Introdução: Nesta experiência queríamos ver o efeito de um ambiente de aprendizagem favorável na melhoria do desempenho em PBR exclusivos. Para isso os participantes responderam a um bloco de treino de 7 PBR exclusivos ou inclusivos, com ou sem feedback e depois responderam a 3 PBR de teste exclusivos sem feedback. Ambientes de treino onde os problemas eram apenas inclusivos não treinavam os participantes a fazer uma integração de informações contrárias, e ambientes sem feedback não mostravam as respostas bayesianas aos participantes de modo a que eles fossem melhorando o seu desempenho. Os ambientes que não promovem a aprendizagem chamam-se ambientes wicked, e os ambientes que apresentam dificuldades desejáveis e feedback corretivo adequado chamam-se ambientes kind (Hogarth, 2003). Inspirados nas melhorias de resultados das experiências anteriores utilizamos um conjunto de melhorias metodológicas. Após resolverem estes 10 PBR os participantes responderam a um CRT de 6 itens, três escalas de estilo cognitivo, uma escala de dificuldade subjetiva e um problema de interpretação abstrata de PBR.

Hipóteses:

1. Os participantes que tiverem feedback na fase de aprendizagem vão ter um melhor desempenho nos na fase de teste, replicando os resultados obtidos por Lindeman, Brink e Hoogstraten (1988). Estes autores verificaram que os participantes aproximavam as suas respostas às normas bayesianas depois de receberem feedback em PBR.
2. A generalização da aprendizagem para a fase de teste irá apenas ocorrer para os participantes que tiveram treino com feedback em PBR exclusivo.
3. O desempenho nos PBR estará correlacionado com o desempenho no CRT e outras medidas de estilo cognitivo.

Método:

Participantes: 171 participantes responderam a um questionário através da plataforma Qualtrics. Os participantes realizaram o questionário em condição não controladas.

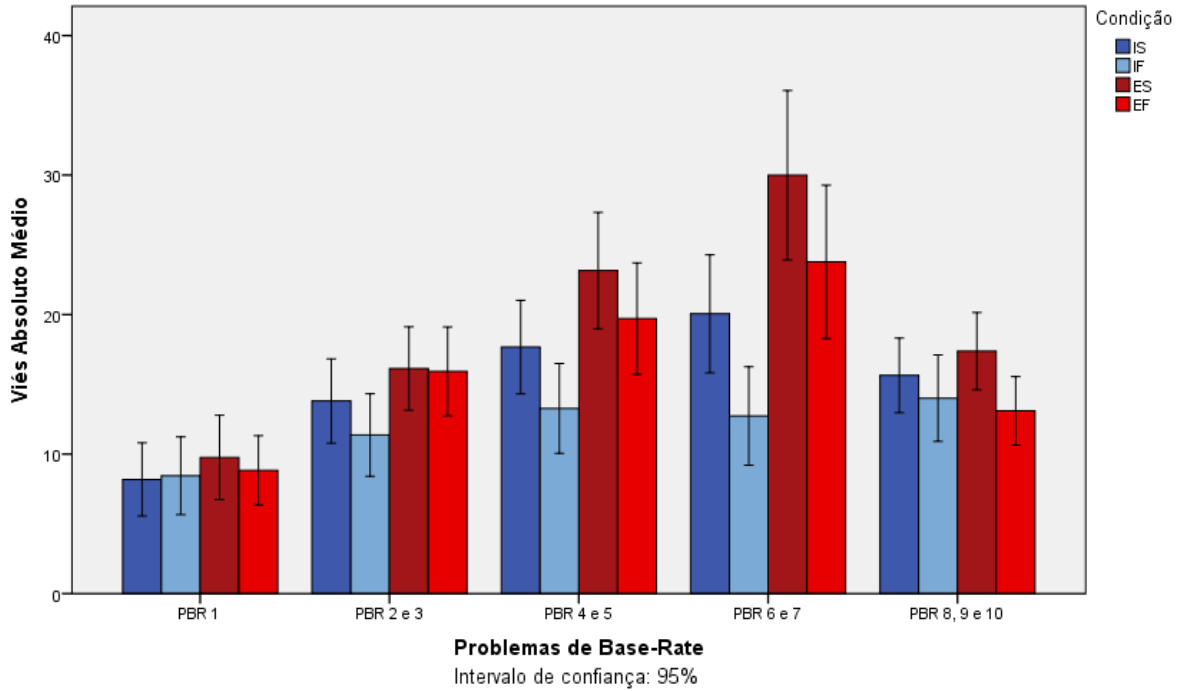
Materiais: Aos 6 PBR já utilizados na experiência 3 e 4 foram acrescentados mais 4 PBR de modo a perfazer um total de 10 PBR (ver anexos). O Cognitive Reflection Test (CRT) e as escalas Need For Cognition (NFC) e Faith in Intuition (FI) apresentadas aos participantes foram iguais às experiências 3 e 4.

Procedimento: À semelhança da experiência 4 os participantes responderam a um PBR de treino onde a descrição apresentada era neutra e a resposta bayesiana era igual ao base-rate. De seguida os participantes responderam a 7 PBR de treino numa de 4 condições. Uma condição com PBR inclusivos sem feedback (Condição “IS”), uma condição com PBR inclusivos com feedback (Condição “IF”), uma condição com PBR exclusivos sem feedback (Condição “ES”) e uma condição com PBR exclusivos com feedback (Condição “EF”). O feedback fornecido apresentava a resposta que o participante deveria ter dado, a resposta que ele deveria ter dado e a distância entre as duas respostas em valor absoluto. Após acabarem a fase de aprendizagem, os participantes responderam a uma fase de teste composta por 3 PBR exclusivos sem feedback. Em todos os PBR a frequência da descrição no grupo típico e atípico foi apresentada no próprio PBR. Estes valores foram obtidos através dos valores médios obtidos nas experiências anteriores para 6 PBR e através dos valores médios obtidos em pré-teste (N=31). O valor diagnóstico da descrição foi aumentando progressivamente ao longo da fase de aprendizagem (min=52%; max=73%). Na fase de teste o valor diagnóstico da descrição foi estandardizado (média=63%) e os PBR foram apresentados por ordem aleatória. Em todos os PBR o valor dos base-rates foram calculados de modo a que a resposta normativa bayesiana fosse estandardizada para todos os participantes, independentemente se o PBR era inclusivo ou exclusivo. Para garantir que a resposta normativa era igual para PBR inclusivos e exclusivos o grupo alvo por vezes era o grupo típico e por vezes era o grupo atípico. Esta manipulação foi contrabalançada entre condições na fase de aprendizagem e estandardizada na fase de teste. O valor da resposta normativa bayesiana foi escolhido para os 10 PBR em intervalos de 5% de modo aleatório. A escala de resposta também foi apresentada em intervalos de 5%. O objetivo desta simplificação é o facilitar a resposta dos participantes, pois eles já têm preferência por estes valores da escala e também evitar que os participantes utilizassem uma estratégia de resposta igual aos base-rates apresentados, pois estes não estavam numa escala de 5%. Os detalhes destes PBR (grupos, descrições, diagnosticidades, base-rates, respostas bayesianas e grupos alvo) encontram-se em anexo.

Design: Nesta experiência utilizamos um design experimental 2 x 2 entre participantes. As variáveis independentes foram: Tipo de treino (PBR inclusivos ou PBR exclusivos) e Feedback (aprendizagem com feedback ou aprendizagem sem feedback). Cada participante foi aleatoriamente atribuído a uma destas quatro condições experimentais: PBR Inclusivos Sem Feedback (IS), PBR Inclusivos com Feedback (IF), PBR Exclusivos Sem Feedback (ES) e PBR Exclusivos com Feedback (EF). As variáveis dependentes foram: Resposta ao PBR, Desempenho no CRT, resultado nas escalas NFC e FI e média do secundário.

Resultados: O gráfico 29 apresenta o viés absoluto médio em cada grupo de PBR. Agregámos os problemas (2 e 3), (4 e 5) e (8, 9 e 10) para facilitar a interpretação. A azul estão as condições inclusivas e a vermelho estão as condições exclusivas. As cores mais claras representam as condições com feedback.

Gráfico 29: Viés Absoluto Médio por bloco de PBR.

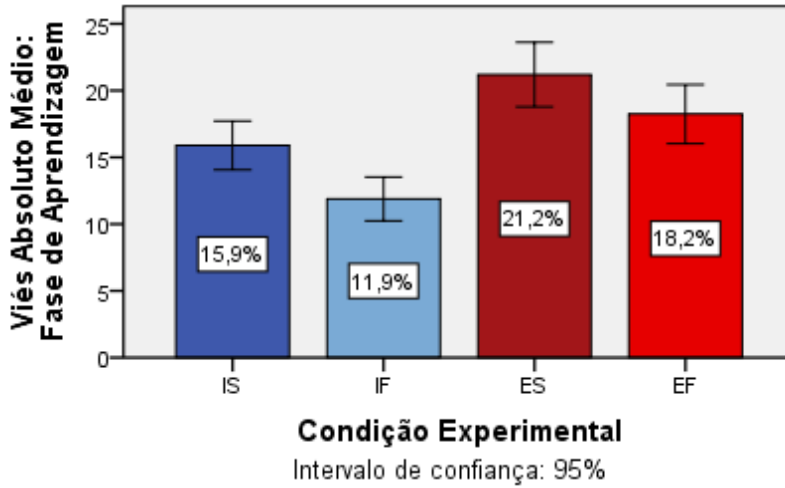


Neste gráfico podemos observar que o desempenho geral dos participantes piorou ao longo da fase de aprendizagem (primeiros 7 PBR). Este resultado deve-se provavelmente ao aumento do progressivo da diagnosticidade da descrição. Embora o desempenho em geral piore, podemos observar que a diferença entre as condições com feedback e sem feedback se vai acentuando ao longo da fase de aprendizagem. Nos PBR 2 e 3 podemos observar uma ligeira melhoria de desempenho devido ao feedback na condição de PBR inclusivos. Esta tendência vai aumentando, tornando-se máxima nos PBR 6 e 7. A melhoria do desempenho devido ao feedback na condição de PBR exclusivos apenas se observa a partir do PBR 4 e 5, tornando-se mais evidente nos PBR 6 e 7. Na fase de teste encontramos também alguma evidência de melhoria do desempenho nas condições que receberam feedback na fase de aprendizagem. Esta melhoria é mais evidente para as condições de PBR exclusivos do que PBR inclusivos.

Para confirmar a significância destes resultados, agregamos os dados dos primeiros 7 PBR e dos últimos 3 PBR. Deste modo podemos comparar o efeito das condições experimentais no viés absoluto médio obtido durante fase de aprendizagem e a sua generalização para a fase de teste utilizando o teste t de Student.

Verificámos a existência de diferenças significativas no desempenho dos participantes obtiveram na fase de aprendizagem em função da condição experimental. O gráfico 30 mostra o viés absoluto médio na fase de aprendizagem por condição experimental.

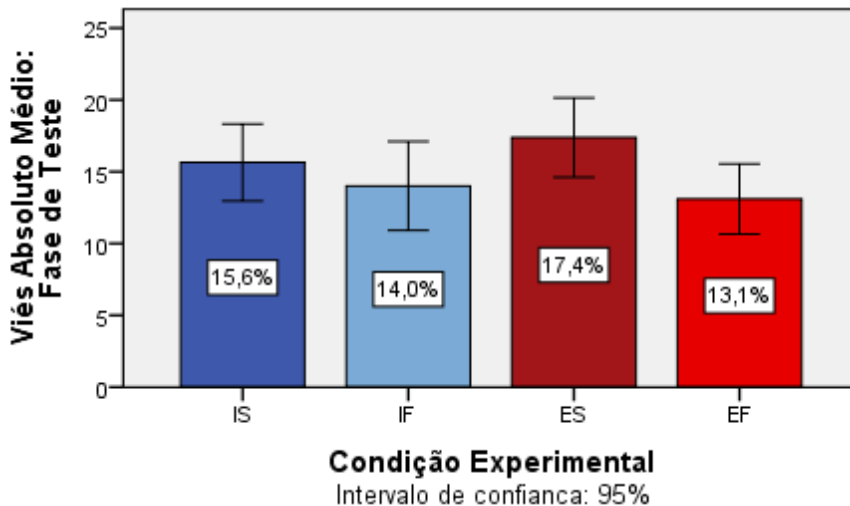
Gráfico 30: Viés Absoluto Médio na Fase de Aprendizagem.



Na condição IS os participantes obtiveram um viés absoluto médio de 15,9% e na condição IF este foi 11,9%. Esta melhoria de desempenho foi significativa ($t = 2,3$; $p = 0,024$) (output em anexo). Na condição ES os participantes obtiveram um viés absoluto médio de 20% e na condição EF de 16,7%. Esta melhoria não foi significativa ($p > 0,1$) (output em anexo).

Estes resultados apoiam a ideia da existência de aprendizagem através da utilização de feedback corretivo. Esta aprendizagem foi mais significativa quando os PBR eram inclusivos do que quando eram exclusivos. Mas será que esta aprendizagem se generalizou para a fase de teste? O gráfico 31 apresenta a mesma análise que o gráfico 30, mas para a fase de teste.

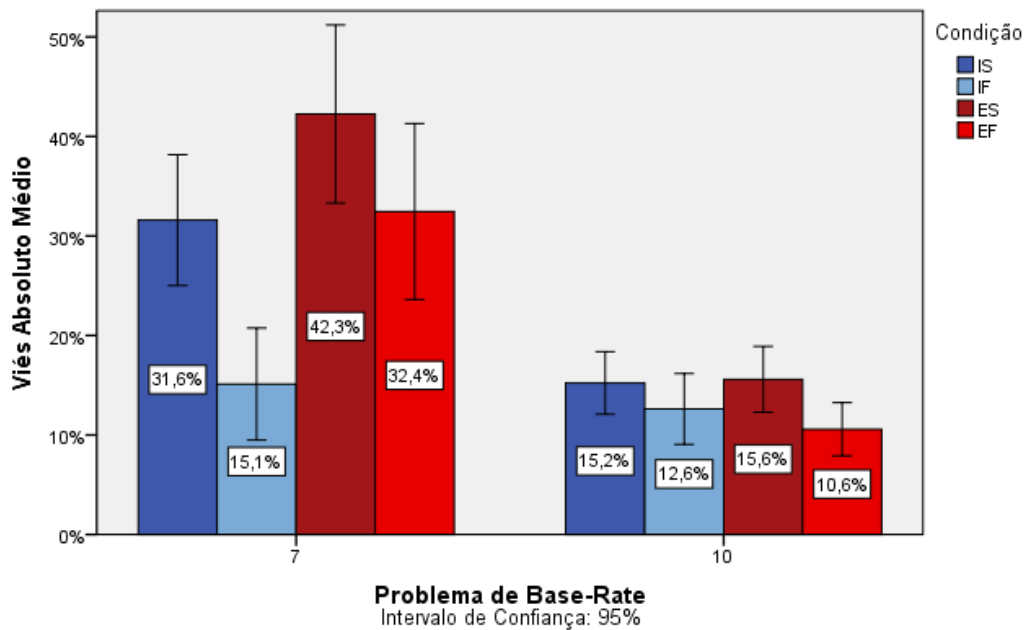
Gráfico 31: Viés Absoluto Médio na Fase de Teste.



Nas respostas dadas aos PBR na fase de teste encontrámos o mesmo padrão de respostas que na fase de aprendizagem. Na condição IS o viés absoluto médio foi de 15,6% e na condição IF foi de 14%. Esta diferença de desempenho não foi significativa ($p > 0,5$). Na condição IS o viés absoluto médio foi de 15,6% e na condição IF foi de 14%. Esta diferença de desempenho foi marginalmente significativa ($t = 1,767$; $p = 0,081$) (output em anexo).

Embora tivéssemos obtido um efeito significativo do feedback em PBR inclusivos na fase de aprendizagem e um efeito marginalmente significativo do feedback em PBR exclusivos na fase de teste, esperávamos encontrar também um efeito significativo do feedback em PBR exclusivos na fase de aprendizagem e um efeito significativo do feedback em PBR inclusivos na fase de teste. É possível que ao agruparmos as respostas estejamos a encobrir a significância dos resultados. De modo a evitar essa situação, escolhemos os PBR onde se encontraram diferenças mais significativas entre as condições IS-IF e ES-EF. O gráfico 32 mostra o viés absoluto médio no PBR 7 e PBR 10 por condição experimental.

Gráfico 32: Viés Absoluto Médio no PBR 7 e PBR 10.

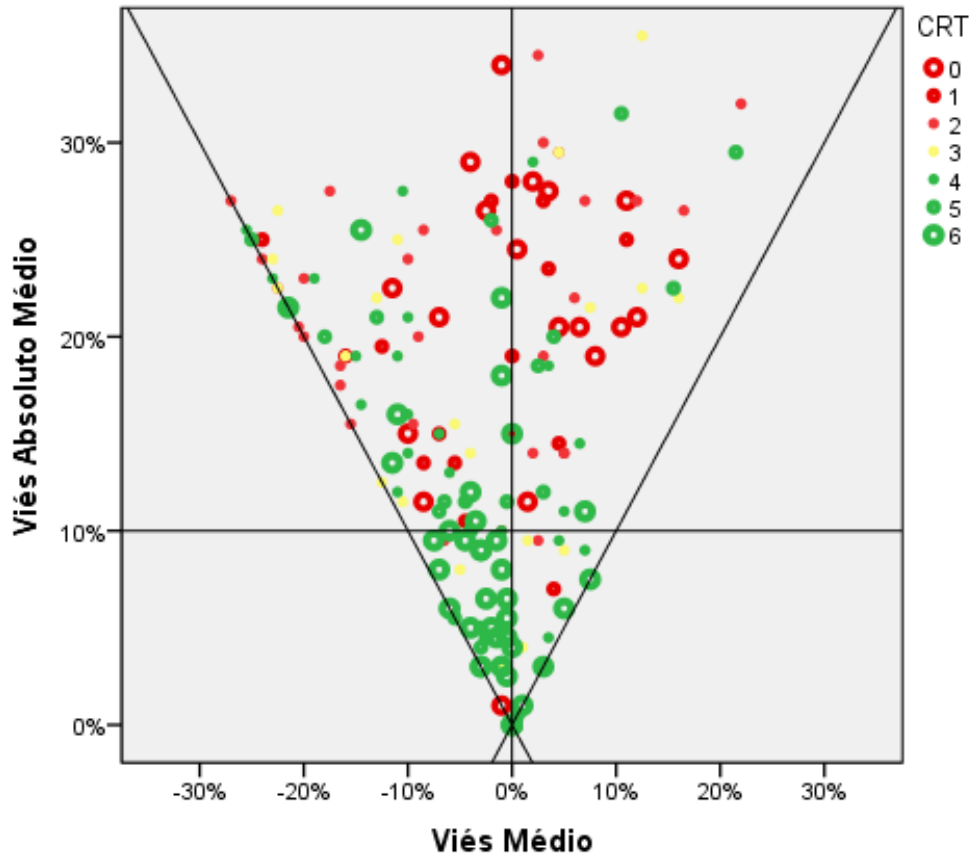


O padrão de resultados nestes dois PBR é semelhante ao padrão obtido quando os dados são agregados por fase (aprendizagem VS teste). A diferença entre a condição ES-EF no PBR 7 continua a não ser significativa ($p > 0,1$) e a diferença entre a condição IS-IF no PBR 10 continua a não ser significativa ($p > 0,2$). No entanto, no PBR 10 a diferença entre a condição ES-EF já é significativa ($t = 2,389$; $p = 0,019$) (output nos anexos).

Este resultado significa que embora a aprendizagem com feedback de PBR exclusivos não tenha atingido diferenças significativas no último PBR da fase de aprendizagem, essa diferença já foi significativa no último PBR da fase de teste. Por outro lado, embora a aprendizagem com feedback de PBR inclusivos tenha atingido diferenças extremamente significativas no último PBR da fase de aprendizagem, essa diferença não foi significativa no último PBR da fase de teste.

De seguida analisámos a relação entre o desempenho e a tendência de resposta com o desempenho no CRT. No gráfico 33 podemos observar a mesma análise gráfica utilizada na experiência 1 onde podemos visualizar a relação entre o desempenho nos PBR e no CRT.

Gráfico 33: Desempenho em PBR VS Desempenho no CRT.



Neste gráfico podemos observar que os participantes com melhor desempenho em PBR (viés absoluto médio < 10%) tiveram quase todos um desempenho máximo no CRT (marcadores verdes grandes), enquanto muitos dos participantes com prior desempenho em PBR (viés absoluto médio > 10%) falharam todos os problemas do CRT (marcadores vermelhos grandes). Na tabela 1 apresentamos as correlações entre o desempenho em PBR e as medidas complementares.

Tabela 1: Correlação entre o Viés Absoluto Médio e as medidas complementares.

		CRT	NFC	FI	MS
Viés Absoluto Médio	Correlação de Pearson	-,515**	-,289**	,237**	-,267**
	Sig.	,000	,000	,002	,001
	N	171	171	171	157

** . A correlação é significativa com $p < 0,01$.

Nesta tabela podemos observar que o desempenho nos 10 PBR se correlacionou significativamente com o desempenho no CRT e medidas complementares com $p < 0,01$.

Discussão:

Ambientes de aprendizagem em PBR: Os resultados obtidos nesta experiência estão de acordo com as nossas hipóteses. Em primeiro lugar, os participantes que tiveram feedback na fase de aprendizagem tiveram um melhor desempenho do que aqueles que não tiveram feedback, mas este efeito foi significativo apenas em problemas inclusivos. Isto pode significar que a os erros cometidos quando os PBR são inclusivos são mais fáceis de corrigir do que quando os problemas são exclusivos. Por outro lado, relativamente à generalização da melhoria do desempenho para a fase de teste, apenas os participantes que receberam PBR exclusivos beneficiaram do feedback e esta melhoria foi particularmente evidente no PBR 10. Isto significa que os participantes, mesmo com feedback, quando são treinados a fazer uma integração inclusiva, quando se deparam com problemas exclusivos não beneficiam do feedback.

Correlação entre o desempenho em PBR e medidas complementares: A corroborar os resultados da experiência 1 e 4 encontramos uma correlação muito forte ($r > 0,5$) e significativa ($p < 0,001$) entre o desempenho nos PBR e no CRT. Para além disso, o desempenho correlacionou-se significativamente com o resultado na escala NFC e FI, um resultado que ainda não tínhamos obtido até aqui. O viés absoluto médio correlacionou-se negativamente com o resultado na escala NFC e positivamente com a escala FI, tal como esperado teoricamente. Para além destes resultados também se correlacionou significativamente com a MS, apoiando os resultados obtidos na experiência 3.

Acreditamos que a melhoria de todas estas correlações em relação às experiências anteriores se deve a vários fatores. Em primeiro lugar, esta foi a experiência que apresentou aos participantes um maior número de PBR, sendo que isso poderá aumentar o significado do viés absoluto médio de cada participante. Para além disto, nesta experiência apresentámos feedback corretivo a metade dos participantes. Embora a correlação com o CRT exista significativamente em todas as condições experimentais, nas condições onde foi apresentado feedback aos participantes a correlação foi mais forte ($r > 0,6$), o que é um indicador de que o feedback melhora a qualidade da medida de viés absoluto médio. Por fim, uma outra razão para se terem obtido correlações desta magnitude foi a simplificação da escala de resposta para intervalos de 5% e a apresentação da frequência da descrição no grupo típico e atípico no próprio PBR.. Investigações futuras deverão usar estas melhorias metodológicas de modo a obterem uma medida de desempenho em PBR tão fidedigna como aquela que obtivemos nesta experiência.

Limitações e investigações futuras:

A comparação direta da dificuldade dos PBR através do viés absoluto médio, não é possível pois existem pelo menos três variáveis estranhas a confundir os resultados: a diagnosticidade da descrição, o grupo alvo e a resposta bayesiana.

O valor da diagnosticidade da descrição não foi estandardizado na fase de aprendizagem, aumentando progressivamente (min=52%; max=73%). Embora o aumento progressivo da diagnosticidade possa facilitar a aprendizagem, pode também ser considerado uma variável estranha na comparação direta da dificuldade entre PBR.

De modo a estandardizar a resposta bayesiana para as quatro condições experimentais apresentámos grupos alvo diferentes. Apesar desta manipulação ter sido contrabalançada e não se ter encontrado diferenças significativas ao nível do viés absoluto médio ($p > 0,6$), esta decisão faz com que as condições inclusivas e exclusivas não pudessem diretamente comparáveis entre PBR, apenas de modo agregado.

Por fim, a própria resposta bayesiana que se escolhe para o PBR irá influenciar o viés absoluto médio. Quando a resposta bayesiana for 45% ou 55%, o viés absoluto máximo será de 55%, mas se a resposta bayesiana for 5% ou 95% o viés absoluto máximo é de 95%. Por esta razão não é possível comparar diretamente a dificuldade de dois PBR com respostas bayesianas diferentes.

No entanto, embora estas decisões impeçam a comparação direta da dificuldade entre PBR, elas permitiram a implementação deste paradigma de aprendizagem. Ao apresentarmos PBR onde a descrição tinha valores diagnósticos progressivamente maiores facilitamos a aprendizagem. Ao apresentarmos grupos alvo diferentes de modo contrabalançado, permitimos a comparação das condições inclusivo e exclusivo de modo agregado. E por fim, ao variarmos de modo aleatória a resposta bayesiana aos PBR, garantimos que o feedback obtido num PBR de modo descontextualizado, não seja preditivo da resposta aos PBR seguintes.

Todas estas decisões foram tomadas para permitir o funcionamento deste paradigma com objetivos de medir a aprendizagem em ambientes diferentes. Investigações futuras, se tiverem por objetivo comparar a dificuldade entre PBR terão de ter em conta estas variáveis estranhas antes de poderem tirar conclusões sobre esse tipo de comparação.

Resumo dos resultados

Na **experiência 1** verificámos que, embora haja uma tendência geral para sobrevalorizar a descrição em PBR, existem vários participantes que dão respostas próximas da resposta bayesiana. Estes participantes que se aproximaram da resposta bayesiana também tiveram um melhor desempenho no CRT, dando a entender que as duas medidas estão relacionadas.

Na **experiência 2** verificámos que, ao responderem a um PBR exclusivo onde o base-rate apresentado era 7%, os participantes utilizavam três estratégias de resposta distintos. Encontrámos participantes a responder com o base-rate, participantes a responder próximo da resposta bayesiana e participantes a sobrevalorizar o valor da diagnosticidade da descrição.

Na **experiência 3** procurámos adaptar os base-rates apresentados à diagnosticidade subjetiva de cada participante e deste modo standardizar a resposta bayesiana. No entanto, este método, em conjunto com uma amostragem pouco controlada, apresentou resultados fracos.

Na **experiência 4** replicámos a experiência 3 com um conjunto de melhorias metodológicas, nomeadamente a apresentação da frequência da descrição no grupo típico e atípico diretamente no próprio PBR. Em conjunto com uma amostragem mais rigorosa e a aplicação deste questionário em condições laboratoriais conseguimos replicar a correlação entre o desempenho nos PBR e no CRT, tal como encontrado na experiência 1. No entanto, ao contrário da experiência 1, encontrámos uma tendência geral para sobrevalorizar os base-rates. Esta tendência já existia na experiência 3, mas tornou-se mais forte na experiência 4.

Na **experiência 5** encontrámos evidência de que os participantes podem aprender a responder de forma mais próxima à resposta bayesiana quando recebem feedback. Os dados desta experiência apontam para uma maior facilidade de aprendizagem quando os PBR são inclusivos do que quando estes problemas são exclusivos. No entanto, a aprendizagem feita com problemas inclusivos não se generaliza para problemas exclusivos. Apenas uma aprendizagem em problemas exclusivos permite um melhor desempenho em problemas exclusivos. Para além destas análises, correlacionámos o desempenho nos PBR com o desempenho no CRT, com o valor obtido na escala NFC/FI e também com a média do secundário. Ao contrário da experiência 1 e semelhante à experiência 3 e 4, encontrámos uma tendência de resposta no sentido dos base-rates.

Discussão geral

Os participantes são capazes de integrar as informações apresentadas em PBR?

Sim, os dados recolhidos nestas experiências revelam que muitos participantes são capazes de fazer a integração das informações apresentadas em PBR. Em todas as experiências encontramos um número significativo de participantes a responder sistematicamente aos PBR integrando as duas informações de modo aproximado ao esperado pelo modelo bayesiano.

Para além disso, encontrámos participantes a sobrevalorizar sistematicamente o valor diagnóstico da descrição tal como previsto pela utilização de heurísticas judicativas de semelhança e disponibilidade (Kahneman & Tversky, 1973).

Por outro lado, encontramos alguns participantes que sobrevalorizam a importância dos base-rates. Este fenómeno pode ser explicado pela utilização de heurísticas de ancoragem e ajuste insuficiente, revelando um efeito de conservadorismo (Phillips, 1966).

Esta integração cognitiva correlaciona-se com outras medidas de racionalidade?

Sim, o desempenho em PBR correlaciona-se com o desempenho noutras medidas de racionalidade. A proximidade a esta norma bayesiana correlacionou-se com o desempenho no CRT (experiência 1, 4 e 5), e valores na escala NFC e FI (experiência 5).

Isto significa que os participantes capazes de analisar as suas primeiras intuições e resistir à tentação de as utilizar nas suas respostas, preferindo fazer uma análise mais profunda antes de responder, têm um melhor desempenho em PBR. A capacidade de resistir às intuições heurísticas e fazer uma análise crítica e ponderada é uma característica essencial no julgamento e tomada de decisão na incerteza e o bom desempenho em PBR parece correlacionar-se com essa capacidade.

Os participantes melhoram o seu desempenho ao recebem feedback corretivo?

Sim, os participantes podem aprender a responder de forma mais próxima à resposta bayesiana quando recebem feedback, em particular se os PBR forem inclusivos. No entanto, o treino em PBR inclusivos pode ser considerado um ambiente wicked, pois a melhoria do desempenho não é generalizara para PBR exclusivos. Embora os participantes aprendam a fazer uma integração cognitiva próxima das normas bayesianas em PBR inclusivos, quando são testados em PBR exclusivos, esta aprendizagem não lhes é útil. O treino em PBR exclusivos pode ser considerado um ambiente kind, pois embora não apresente resultados imediatos na fase de aprendizagem, promove uma melhoria do desempenho na fase de teste.

Este resultado pode apresentar algumas implicações importantes. Uma implicação possível é que, embora seja difícil aprender a integrar informações contrárias, devemos procurar fazê-lo. Pois, se não o fizermos, quando nos depararmos com informações contraditórias, corremos o risco de sobrevalorizar apenas as informações que nos parecem ser intuitivamente verdadeiras.

Limitações do paradigma proposto

A escala contínua de resposta apresenta um elevado grau de ruído estatístico:

Embora a escala de resposta possa ter uma resolução de 0,1% os participantes quase sempre preferem responder com valores inteiros, com valores de resolução de 5% ou 10%. Esta limitação pode ser superada reduzindo a resolução da escala de resposta para 5% (experiência 5).

Alguns participantes dão respostas que não são interpretáveis quantitativamente. Respostas que estão abaixo do base-rate e acima da diagnosticidade são indicadoras de que os participantes estão a utilizar a escala de resposta de modo qualitativo, o que pode apresentar uma limitação para a aplicação de modelos mais complexos de análise qualitativa.

A quantificação da diagnosticidade subjetiva individual não só acrescenta ruído estatístico, por acrescentar ao modelo duas novas variáveis, $P(C|A)$ e $P(C|B)$, como pode levar a valores de diagnosticidade extrema quando uma destas variáveis toma o valor de 0%. No pior dos casos, o participante pode atribuir a estas duas variáveis o valor de 0%, fazendo com que o valor da diagnosticidade seja indeterminado. Este problema pode ser superado apresentando aos participantes estes valores no próprio PBR (experiência 5).

Normatividade da resposta bayesiana e controlo experimental:

Embora tenhamos encontrado evidência de que existe uma tendência sistemática dos participantes utilizarem uma de três estratégias de resposta, isso não quer dizer que uma das estratégias de resposta seja melhor que a outra, porque os problemas apresentados nesta tese não são completamente abstratos. Ao incluírem neles estereótipos sobre grupos sociais levantam também questões éticas e morais que ultrapassam um paradigma meramente cognitivo. Por esta razão, a estratégia de resposta que ignora a diagnosticidade das descrições e utiliza apenas o base-rate pode ser explicada para alguns participantes como o negar do estereótipo, pois embora o estereótipo seja diagnóstico pode ser considerado errado estar a considerar essa informação. Esta possível crítica ao paradigma era conhecida logo desde o início e por isso procuramos utilizar sempre estereótipos positivos. No entanto, ao contrastar dois grupos, ao afirmar uma característica positiva e diagnóstica de um grupo é implícitamente afirmar a ausência dessa característica positiva num outro grupo. Por esta razão, ao contrastar grupos sociais, para além da componente cognitiva de integração de informação qualitativa e quantitativa, para além da componente cognitiva, existe também uma componente ética e moral que pode levar o participante a ignorar a descrição em vez de ignorar os base-rates.

Em nenhuma das experiências se utilizou um controlo experimental rigoroso para as descrições utilizadas. Embora tenhamos controlado o valor dos base-rates e encontramos diferenças significativas nas respostas dadas pelos participantes, não fizemos nenhum controlo experimental para as descrições. Isso por si só não é muito grave, pois utilizamos materiais com temáticas diferentes de modo a não ferir as sensibilidades de nenhum participante.

O conceito de normatividade estatística é um conceito relativamente recente na história da humanidade e embora os participantes nestes estudos sejam na sua grande maioria estudantes universitários, isso não quer dizer que tenham aprendido e interiorizado o cálculo das probabilidades condicionadas. Esta é uma dificuldade desejável, pois impede a pessoa de dar uma resposta analítica, obrigando-a a recorrer a processos heurísticos e aproximações. Por esta razão, os desvios à resposta normativa podem ser mais interessantes do que acertar na resposta correta.

A subjetividade objetiva e objetividade subjetiva da diagnosticidade:

Será possível quantificar com precisão a diagnosticidade subjetiva de uma característica?

Será que a representação mental da diagnosticidade corresponde exatamente ao índice de diagnosticidade quantificado?

Os dados recolhidos nesta teste não apoiam esta hipótese, pois muitos participantes responderam acima dos valores da diagnosticidade que eles próprios definiram. Este resultado leva-nos a interrogar até que ponto a escala contínua de 100 pontos está a ser utilizada de modo quantitativo. Para Kahneman e Tversky quando um participante tenta responder ao um PBR ele não está a fazer uma contagem das pessoas do grupo típico e do grupo atípico que encaixam na descrição, mas sim a fazer uma análise da semelhança da descrição ao grupo típico. Deste ponto de vista é natural que os participantes ignorem os valores quantitativos que acabaram de introduzir.

Neste paradigma, a diagnosticidade da descrição de um PBR foi medida em função da frequência de uma descrição num grupo, onde ela é típica e noutra group, onde é atípica. Esta medida pode ser pedida ao participante de forma subjetiva ou apresentada ao participante de modo objetivo. Estas duas abordagens têm as suas limitações. Na abordagem subjetiva estamos a adicionar ruído estatístico ao modelo, considerando que é possível obter medidas quantitativas rigorosas através da introspeção. Na abordagem objetiva, mesmo apresentando os valores médios obtidos em pré-teste corremos o risco de estes valores se afastarem muito da diagnosticidade subjetiva de alguns participantes, e deste modo adicionar uma variável estranha ao modelo.

Quando trabalhamos com quantificações de representações subjetivas temos sempre de ter cuidado pelo modo como operacionalizamos os conceitos. Embora o método utilizado nesta tese nos tenha permitido obter resultados bons, há sempre espaço para melhorar. Na próxima secção vamos discutir algumas destas melhorias e possibilidades de investigação futura.

Investigações futuras

Transformar variáveis estranhas em variáveis independentes controladas:

Nas experiências relatadas existem variáveis estranhas que limitam a interpretação dos resultados obtidos. Na maioria das experiências (1, 3, 4 e 5) tanto os base-rate como as descrições apresentadas foram variadas. No entanto, para sermos rigorosos a medir o efeito de um destes fatores devíamos estandardizar o outro. Investigações futuras podem procurar fixar algumas destas variáveis de modo a verificar como elas interagem de maneira mais rigorosa. Por exemplo, seria interessante fazer uma experiência semelhante à experiência 2, onde se fixasse o base-rate apresentado e se variasse de modo controlado a diagnosticidade da descrição em toda a escala. Seria também interessante fazer o contrário e estandardizar a descrição. Neste caso poderíamos apresentar um PBR que começasse com um base-rate de 50% e que vai diminuindo tornando o problema progressivamente mais exclusivo. O objetivo seria medir o ponto a partir do qual o participante muda a sua opinião e acredita que é mais provável a pessoa escolhida ao acaso ser do grupo atípico. Assumindo que a sua primeira resposta não é 100%, existirá um base-rate para o qual a sua resposta será inferior a 50%.

Adaptação do paradigma de PBR a outras áreas de investigação:

A integração bayesiana, vista nesta tese como um standard de racionalidade, quando aplicada fora de modelos meramente cognitivos, pode não ser a resposta mais correta. Por exemplo, um participante pode rejeitar a diagnosticidade da descrição por não querer afirmar um determinado estereótipo com o qual não concorda. Podem haver várias razões para um participante se desviar da resposta bayesiana, no entanto, esta pode ser uma vantagem deste paradigma quando o aplicamos a outras áreas de investigação.

O paradigma de PBR, por exigir a integração de informação quantitativa e qualitativa, permite-nos estudar os estereótipos. Por exemplo, se estandardizarmos tanto o base-rate como a diagnosticidade da descrição, mas manipularmos a valência do conteúdo estereotípico, podemos observar efeitos relacionados com a valência dos estereótipos. Se um participante apresenta uma tendência de resposta diferente para descrições negativas do que para descrições neutras com o mesmo nível de diagnosticidade, isso pode ser uma medida indireta e implícita de discriminação. Ou seja, o foco deixa de ser o desempenho absoluto do participante, mas sim se o participante é mais enviesado quando a característica é negativa do que quando é neutra e o sentido deste enviesamento. Existem algumas limitações em relação a este tipo de investigação, nomeadamente encontrar descrições neutras que sejam tão diagnósticas como descrições negativas. No entanto, se o paradigma for bem implementado e as suas limitações forem controladas, este poderá tornar-se um método viável de investigar processos complexos de avaliação dos estereótipos.

Na área da investigação em dilemas morais este paradigma também pode ser utilizado. Embora seja interessante estudar como as pessoas podem integrar várias informações contraditórias numa decisão ética nuanciada, o julgamento moral é normalmente mais clivado do que o julgamento probabilístico abstrato. É relativamente comum um participante responder que a probabilidade de uma pessoa ser do grupo típico é 50% quando essa é a resposta bayesiana. No entanto, se estes problemas forem adaptados a questões morais é possível que encontremos respostas mais extremadas, pelo que seria interessante perceber o que diferencia um participante que dá respostas extremadas de um que dá respostas nuanciadas. Embora esta linha de investigação pareça interessante do ponto de vista teórico, na prática não será fácil adaptar um paradigma probabilístico a uma temática de dilemas morais.

Adaptação do paradigma de PBR a outros métodos e materiais:

O maior problema do paradigma proposto nesta tese é que as escalas de quantificação da diagnosticidade e as escalas de resposta existem num contínuo de 0% a 100%. Embora metodologicamente isto apresente uma grande vantagem, pois cada resposta pode ser medida como um desvio contínuo a uma resposta bayesiana objetiva, os participantes não interagem com esta escala necessariamente de maneira contínua. Para colmatar este problema podemos simplificar a resolução da escala de resposta a uma resolução de 5% e apresentar os valores da diagnosticidade no próprio PBR (experiência 5). Esta melhoria metodológica melhora em grande medida a qualidade dos resultados. No entanto, mesmo assim é necessário que o participante tenha alguma sensibilidade a escalas de resposta quantitativas. As escalas mais utilizadas em investigação são escalas de resposta qualitativas ordinais e com boa razão, pois as escalas de resposta quantitativas não são intuitivas. Respostas dadas nesta escala não mapeiam bem os processos mentais que pretendem representar, pois a escala tem uma resolução superior aquela que o participante consegue diferenciar. Para superar esta limitação, investigações futuras deverão procurar melhorar a metodologia deste paradigma apresentando os PBR de forma figurativa e permitindo respostas de forma visual. Uma das críticas originais a este paradigma é que os base-rates eram apresentados na forma de percentagens e que os participantes interpretavam os problemas de modo mais rigoroso quando esta informação era apresentada na forma de frequências naturais. No entanto, esta informação, para ser representada de maneira correta e semelhante por todos os participantes deve ser apresentada na forma de imagens. Este método pode ser feito tanto pela apresentação de imagens sequenciais, onde a pessoa faz uma impressão subjetiva do tamanho dos dois grupos, como pode ser apresentada de maneira abstrata e codificada. Podemos, por exemplo, apresentar ao participante um conjunto de pontos distribuídos aleatoriamente onde a cor e a forma podem representar o grupo e a presença da característica em questão. Deste modo, o participante tem acesso numa única imagem a várias informações que teria de processar de maneira abstrata.

Referências Bibliográficas

- Bilker, W. B., Hansen, J. A., Brensing, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, 19(3), 354-369.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, 42(1), 116.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7(1).
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of personality and social psychology*, 71(2), 390.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3).
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4).
- Hogarth, R. M. (2003). Educating intuition: a challenge for the 21st century. CREI, Centre de Recerca en Economia Internacional.
- Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, 110(2).
- Lindeman, S. T., Van Den Brink, W. P., & Hoogstraten, J. (1988). Effect of feedback on base-rate utilization. *Perceptual and Motor Skills*, 67(2), 343-350.
- Loureiro, F., & Garcia-Marques, T. (2018). I know the rule, but I'll just go with my gut: is there a rational use of intuition?. *Thinking & Reasoning*.
- Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, 40(4).
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, 80, 34-72.
- Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of experimental psychology*, 72(3).
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, 80(4), 237.
- Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of experimental psychology: learning, memory, and cognition*, 43(7), 1154.

Anexos

Experiência 1 – Paradigma Contínuo:

Na tabela 2 podemos ver os grupos e as descrições utilizadas nesta experiência.

Tabela 2: Problemas de Base-Rate: Grupos e descrição

Grupo A	Grupo B	Descrição
Engenheiro	Advogado	Tem 36 anos, é solteiro, gosta de ler ficção científica e programar.
20 Anos	50 Anos	Gosta de ouvir música techno e eletrônica, usa jeans e óculos escuros, tem um pequeno piercing no nariz e gostar de dançar.
Budista	Muçulmana	Gosta de filosofia, tem aversão a bens materiais, usa habitualmente roupa em segunda mão e ambiciona poder viajar até à Índia.
Italiana	Americana	É bilingue, gosta de ler nos tempos livres, estar na moda e cozinhar.
Rapaz	Rapariga	Gosta de sair à noite com amigos, ouvir música e beber cervejas.

Na tabela 3 podemos ver os base-rates utilizados nos 5 PBR em cada uma das versões.

Tabela 3: Problemas de Base-Rate: Base-Rates

Grupo A	Versão do PBR			Grupo B	Versão do PBR		
	25%	50%	75%		25%	50%	75%
Engenheiro	117	285	456	Advogado	883	715	544
20 Anos	7	18	40	50 Anos	93	82	60
Budista	145	338	821*	Muçulmana	855	662	179*
Italiana	17	38	65	Americana	83	62	35
Rapaz	201	431	694	Rapariga	799	569	306

*Este base-rate foi mal calculado

Na tabela 4 podemos ver os problebas do CRT e as respetivas respostas.

Tabela 4: Problemas de CRT

Pergunta	Resposta
1- Um bloco de notas e um lápis custam 1 euro e 10 cêntimos. O bloco de notas é um euro mais caro que o lápis. Quanto custa o bloco de notas?	O bloco de notas custa 1,05 euros.
2- Se cinco máquinas demoram cinco minutos para fazer cinco peças, quanto tempo demoram sessenta máquinas para fazer sessenta peças?	5 minutos
3- Num lago de nenúfares, todos os dias o número de nenúfares duplica. Se são necessários 48 dias para que todo o lago esteja coberto de nenúfares, quantos dias são necessários para que metade do lago esteja coberto de nenúfares?	47 dias

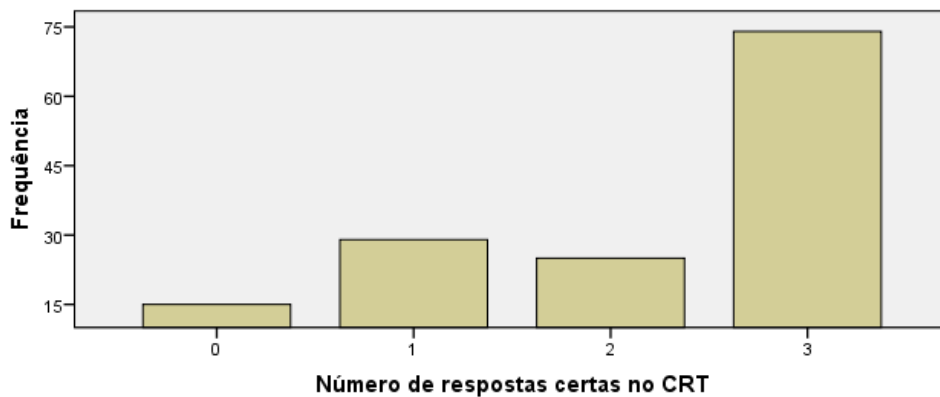
Na tabela 5 podemos ver o output da ANOVA entre a resposta ao PBR e a versão do PBR.

Tabela 5: ANOVA da Resposta ao PBR VS Versão do PBR

	Soma dos Quadrados	gl	Média Dos Quadrados	F	Sig.
Entre Grupos	69988	2	34994	62	,000
No Mesmo Grupo	422477	747	566		
Total	492465	749			

No gráfico 34 vemos que poucos participantes falharam em todos os problemas de CRT.

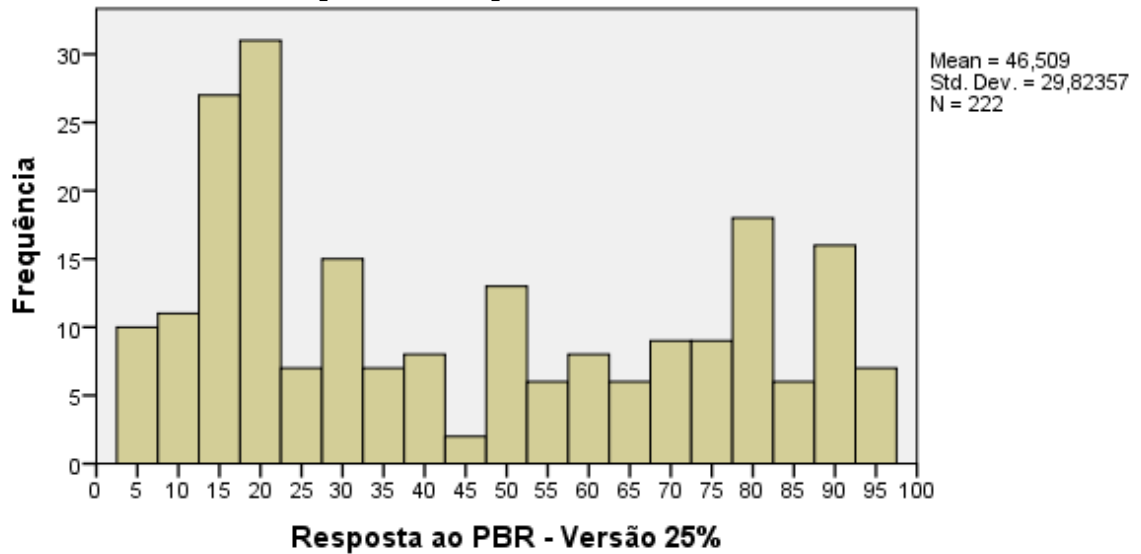
Gráfico 34: Frequência de participantes por número de respostas certas no CRT.



Por esta razão, nesta experiência, os participantes foram agregados em dois grupos de tamanhos comparáveis, participantes que tiveram um desempenho ótimo no CRT e os restantes.

No gráfico 35 podemos ver a frequência das respostas aos PBR na versão 25%.

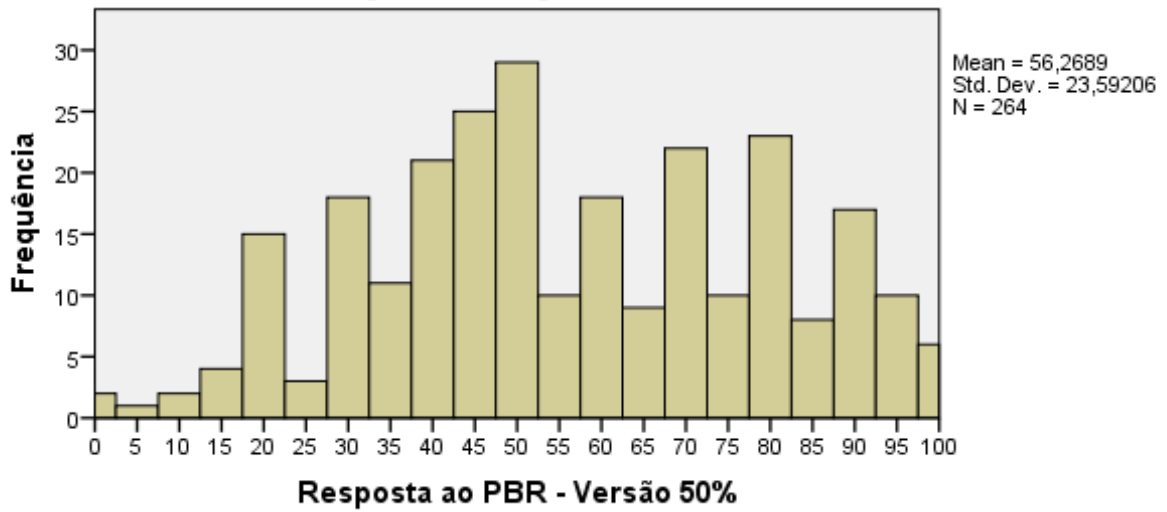
Gráfico 35: Frequência de respostas ao PBR na versão 25%.



Podemos observar que os dados não se distribuem simetricamente em torno do valor 25%. Identificamos um conjunto de participantes que responde em torno do valor de 20% e um conjunto de participantes que responde em torno do valor 80%.

No gráfico 36 podemos ver a frequência das respostas aos PBR na versão 50%.

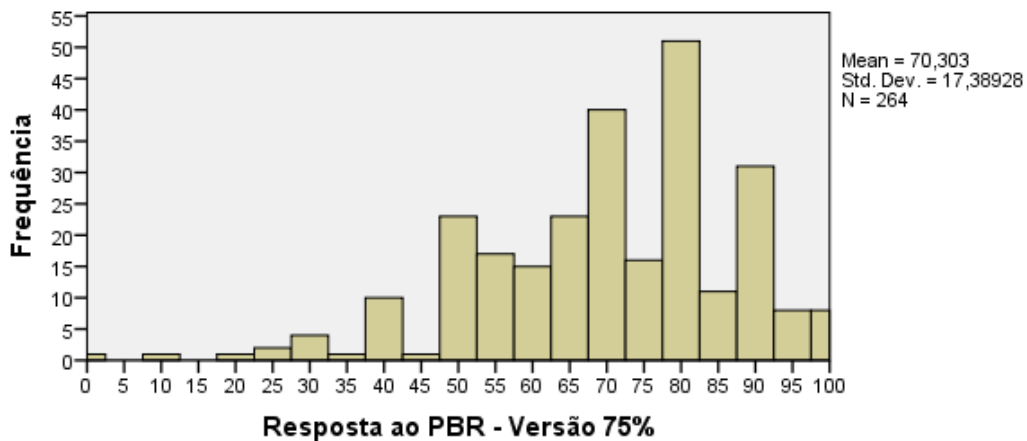
Gráfico 36: Frequência de respostas ao PBR na versão 50%.



Quando a versão é 75% a maioria das respostas concentra-se no valor 80%, 70% e 90%, sendo que a qualidade bimodal da distribuição é muito menos evidente.

No gráfico 37 podemos ver a frequência das respostas aos PBR na versão 75%.

Gráfico 37: Frequência de respostas ao PBR na versão 75%.



Este padrão de respostas é compatível com a utilização de duas estratégias de resposta, uma envolve a utilização do valor dos base-rates ou do valor da diagnosticidade da descrição. Estas duas estratégias levam a respostas diferentes na versão 25%, mas a respostas semelhantes na versão 75%. Por esta razão é que o desvio padrão é máximo na versão 25% (DP=30%) e mínimo na versão 75% (DP=17%).

De seguida, com base na quantificação da diagnosticidade das descrições, calculámos duas medidas: uma resposta bayesiana individual e uma resposta bayesiana grupal. Esta segunda medida serve apenas como medida de controlo de modo a verificar se os participantes estão a refletir na sua resposta os valores subjetivos de diagnosticidade ou se os valores grupais de subjetividade seriam igualmente válidos neste paradigma.

A tabela 6 apresenta o viés individual e grupal absoluto médio em cada versão do PBR.

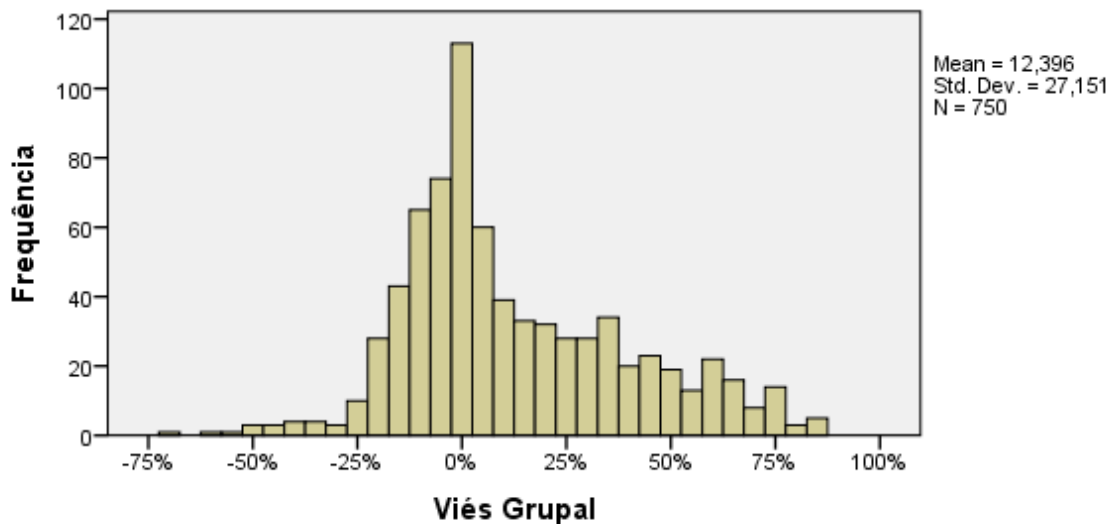
Tabela 6: Valores médios do viés individual absoluto e viés grupal absoluto.

Versão do PBR		VIA	VGA
25%	Média	30,4	30,3
	N	222	222
	Desvio Padrão	25,7	27,7
50%	Média	20,	22,10
	N	264	264
	Desvio Padrão	16,3	17,8
75%	Média	13,8	13,5
	N	264	264
	Desvio Padrão	14,3	12,310
Total	Média	21,0	21,5
	N	750	750
	Desvio Padrão	20,1	20,7

Como esperado a versão 25% leva a um maior viés absoluto médio do que a versão 50% e 75%. No entanto, ao contrário do que se esperava, a medida de viés grupal absoluto médio apresenta valores muito próximos da medida de viés individual absoluto médio. Adiferença de médias do viés individual absoluto e do viés grupal absoluto é apenas de 0,5% ($p < 0,4$).

No gráfico 38 podemos observar a frequência dos vieses à resposta bayesiana grupal.

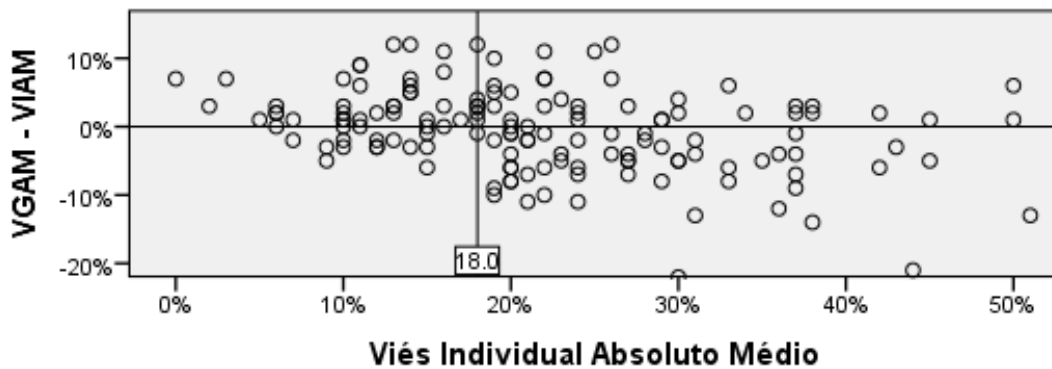
Gráfico 38: Frequência de vieses à resposta bayesiana grupal.



Estes vieses apresentam uma média mais positiva e uma assimetria e curtose superiores à distribuição de vieses à resposta bayesiana individual. Isto é um indicador de que a medida de viés grupal é uma medida mais pobre do que a medida de viés individual.

De seguida verificamos se a diferença entre estas duas medidas aumentava para os participantes com melhor desempenho em PBR. Se os participantes basearem as suas respostas nos valores individuais de diagnosticidade, então uma medida de viés individual terá uma precisão maior do que uma medida de viés grupal. Verificamos que, para os participantes com um viés individual absoluto médio inferior a 18%, o viés individual é uma medida mais preditiva da resposta dos participantes do que o viés grupal. No gráfico 39 podemos observar um gráfico de dispersão onde cada ponto é um participante, o eixo horizontal representa o viés individual absoluto médio e o eixo vertical representa a diferença entre estas duas medidas.

Gráfico 39: Benefício da medida de viés individual em função do viés absoluto médio.



Observámos que os participantes com bom desempenho ($VIAM > 18\%$) beneficiam de uma medida individual de viés, enquanto o contrário é verdade para os restantes participantes.

Como última análise de comparação da qualidade destas duas medidas de viés correlacionamo-las com o desempenho no CRT. Na tabela 7 podemos observar esta correlação.

Tabela 7: Correlação entre o VIA e o VGA com o desempenho no CRT.

		VIA	VGA	CRT
Viés Individual Absoluto	Correlação de Pearson	1	,729**	-,193*
	Sig.		,000	,018
	N	150	150	150
Viés Grupal Absoluto	Correlação de Pearson	,729**	1	-,154
	Sig.	,000		,060
	N	150	150	150

** . A correlação é significativa com $p < 0,01$.

* . A correlação é significativa com $p < 0,05$.

Verificamos que embora as duas medidas sejam semelhantes em média, elas apresentam correlações de magnitudes e níveis de significância distintos com o desempenho no CRT.

No geral podemos considerar que o viés individual absoluto médio é uma medida de qualidade superior à medida de viés grupal absoluto médio. No entanto, estas medidas estão suficientemente próximas para justificar a apresentação da diagnosticidade média nos PBR sem comprometer gravemente a qualidade da medida do viés absoluto médio.

Experiência 2 – Escolha Livre:

Escolha das características e dos grupos: Os participantes escolheram como grupos profissões, grupos etários, sexos, nacionalidades, grupos sociais, etnias, partidos políticos, religiões, sexualidades, partidos políticos, etc. As características escolhidas foram características físicas, psicológicas, gostos, tendências, actos, capacidades, saúde, etc. Os gráficos 40, 41 e 42 representam a proporção das categorias e descrições escolhidas pelos participantes.

Gráfico 40: Categorias principais dos grupos típicos.

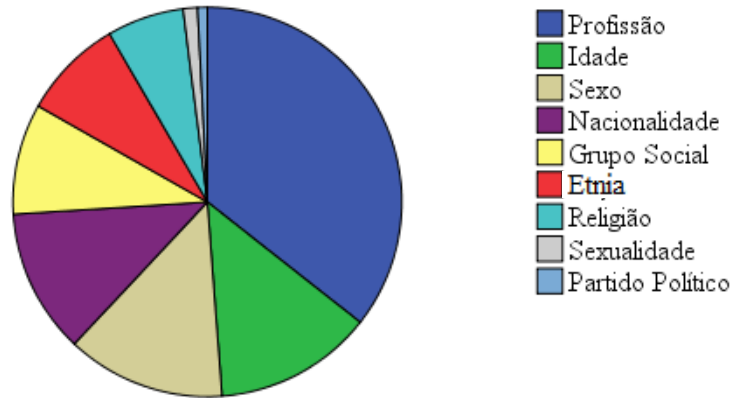


Gráfico 41: Categorias principais dos grupos atípicos.

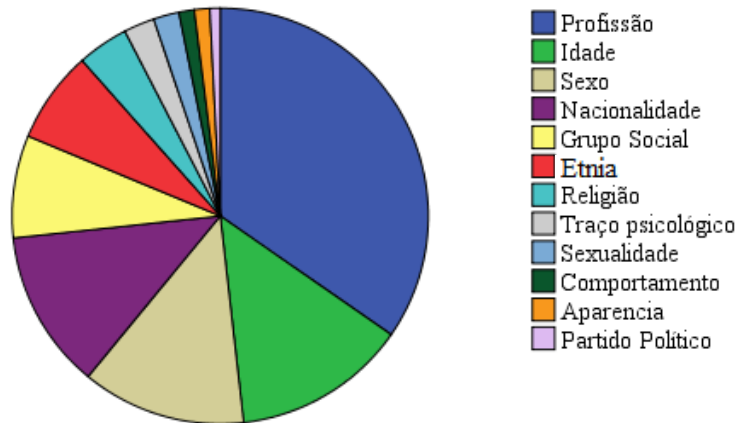
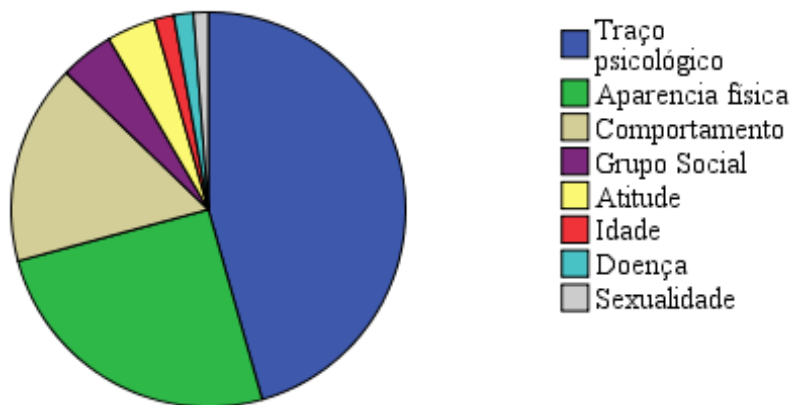


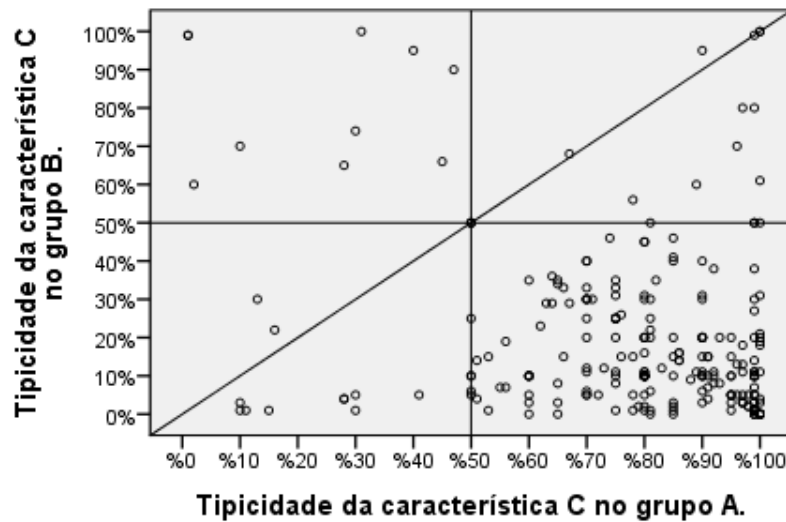
Gráfico 42: Categorias principais das descrições.



O resultado mais surpreendente foi que os participantes decidiram escolher profissões acima de qualquer outra temática para responder ao questionário. A característica mais típica foi a escolha de um traço de personalidade, aspeto físico ou comportamental que fosse diagnóstico de duas profissões diferentes. Este resultado foi inesperado pois a categoria das profissões foi a última categoria a ser sugerida nas instruções. Porque é que esta temática foi mais utilizada do que as outras? Será que é mais fácil diferenciar entre profissões do que entre idades, sexos, nacionalidades, etnias e religiões? Será que uma profissão é composta por elementos mais homogêneos e contrastantes com outros grupos e por isso é que foi o tema mais vezes escolhido? Esta é uma pergunta pertinente para a qual não tínhamos nenhuma hipótese prévia. Por fim, verificámos que um pequeno número de participantes utilizou mais do que um grupo social para definir os grupos. Esta combinação de vários grupos num só poderá ser uma ideia interessante a ser explorada em investigações futuras pois foi uma categorização múltipla espontânea que não foi pedida nas instruções e que pode ajudar a maximizar a diagnosticidade da característica.

De seguida analisamos cuidadosamente os valores escolhidos para a frequência da característica no grupo típico e atípico. O gráfico 43 podemos observar esta relação.

Gráfico 43: Relação entre a frequência da característica para o grupo A e B.

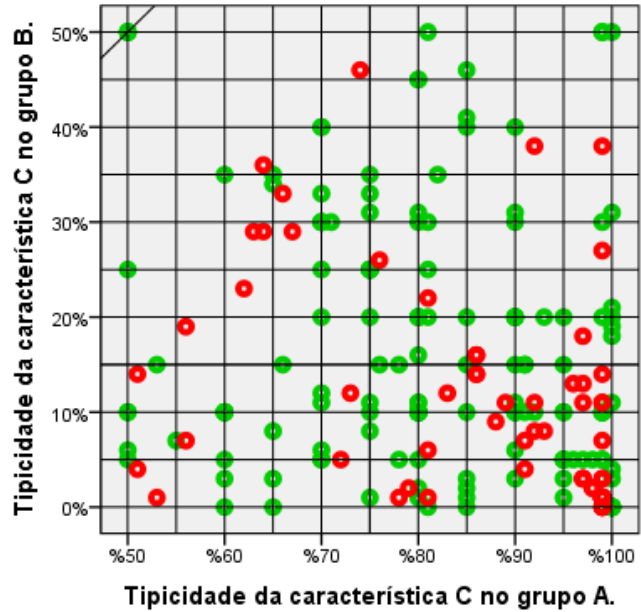


Os pontos abaixo da diagonal representam os participantes para os quais a característica era diagnóstica no sentido esperado. No quadrante inferior direito situam-se os participantes para os quais a característica tinha uma probabilidade maior que 50% no grupo típico e menor que 50% para o grupo atípico. Verificamos também que existem poucos participantes para os quais a descrição era pouco diagnóstica ou neutra, pois existem poucos pontos na proximidade da diagonal do gráfico.

Embora a escala de resposta fosse contínua, os valores utilizados pelos participantes para quantificar a percentagem de pessoas do grupo A e grupo B que tinham a característica C eram ancorados em percentagens de 5%. No gráfico 44 podemos observar este fenómeno.

Gráfico 44: Relação entre a frequência da característica para o grupo A e B.

Neste gráfico podemos observar o detalhe do quadrante inferior direito do gráfico acima com a sobreposição de uma grelha em intervalos de 5%. Os pontos verdes são os participantes que utilizaram valores que coincidiam com esta grelha, os pontos vermelhos são os participantes que não coincidiam com esta grelha. Podemos observar uma preferência por valores em intervalos de 5%.



A grande maioria dos participantes respondeu ao questionário com valores da tipicidade da descrição no grupo A e B de modo a que esta fosse diagnóstica, tal como instruídos pelo questionário. No entanto, os valores destas tipicidades não foram escolhidos de forma contínua, pois, mesmo tendo a escala uma precisão de 0,1, ao dar a sua resposta os participantes escolheram maioritariamente valores múltiplos de 5% em pelo menos uma das escalas. Isto confirma a nossa hipótese de que os participantes não estão a utilizar a escala de probabilidades de modo contínuo, mas dão a sua resposta em valores discretos. Isto leva-nos a questionar se a força diagnóstica da característica escolhida é representada mentalmente como um contínuo ou não. Seja qual for a explicação para este fenómeno o pressuposto de que as pessoas estão a utilizar estas escalas de modo contínuo não é suportado pelos dados. Este resultado apoia a ideia de que escalas de resposta e de quantificação da diagnosticidade com uma resolução de 5% talvez possam ser mais práticas e mais intuitivas para os participantes do que escalas com uma resolução inferior.

Experiência 3 – Base-Rates Adaptados:

Na tabela 8 podemos ver os grupos e descrições dos PBR desta experiência.

Tabela 8: Problemas de Base-Rate: Grupos e Descrição

Grupo A	Grupo B	Descrição	Diagnosticidade
Rapaz	Rapariga	Gosta de sair à noite com amigos, ouvir música e beber cerveja.	51%
Italiana	Americana	É bilingue, gosta de ler nos tempos livres e de cozinhar.	62%
Budista	Muçulmana	Gosta de filosofia, tem aversão a bens materiais, usa habitualmente roupa em segunda mão.	63%
Engenheiro	Advogados	Gosta de ler ficção científica e programar no computador.	69%
20 Anos	50 Anos	Gosta de ouvir música techno e eletrónica, usa jeans e óculos escuros, tem um pequeno piercing no nariz e gostar de dançar.	73%
Só comprar roupas em viagens ao estrangeiro	Só comprar roupas em supermercados	Vive no último andar de um dos melhores prédios da cidade e conduz um Porche.	73%

Na tabela 9 vemos as perguntas e respostas dos problemas de CRT desta experiência.

Tabela 9: Problemas de CRT

Pergunta	Resposta
1- Um bloco de notas e um lápis custam 1 euro e 10 cêntimos. O bloco de notas é um euro mais caro que o lápis. Quanto custa o lapis?	5 cêntimos
2- Se cinco máquinas demoram cinco minutos para fazer cinco peças, quanto tempo demoram sessenta máquinas para fazer sessenta peças?	5 minutos
3- Num lago de nenúfares, todos os dias o número de nenúfares duplica. Se são necessários 48 dias para que todo o lago esteja coberto de nenúfares, quantos dias são necessários para que metade do lago esteja coberto de nenúfares?	32 dias
4- Tem cinco moedas no bolso. O seu amigo também tem cinco moedas no bolso. Dá uma das suas cinco moedas ao seu amigo. Com quantas moedas a mais é que o seu amigo ficou em relação a si?	Duas moedas
5- Está a participar numa corrida e ultrapassa a pessoa que está em segundo lugar. Em que lugar ficou?	Segundo lugar
6- Um tijolo pesa um quilo mais meio tijolo. Quanto pesa um tijolo?	Dois quilos

Na tabela 10 podemos ver os itens da escala Need for Cognition.

Tabela 10: Escala Need For Cognition

Item
Não gosto de ter de pensar muito.
Tento evitar situações onde tenha que pensar profundamente sobre alguma coisa.
Prefiro problemas complexos a problemas simples.
Prefiro fazer algo que me desafie intelectualmente em vez de algo que exija pouco pensamento.
Pensar profundamente e durante muito tempo sobre uma coisa dá-me pouca satisfação.?

Na tabela 11 podemos ver os itens da escala Faith in Intuition

Tabela 11: Escala Faith in Intuition

Item
Confio nas minhas primeiras impressões sobre as pessoas. Creio poder confiar nos meus palpites. As minhas primeiras impressões acerca dos outros estão quase sempre certas. No que se refere a confiar nas pessoas, geralmente confio nos meus instintos. Geralmente consigo sentir se uma pessoa está certa ou errada, mesmo sem conseguir explicar como.

Na tabela 12 vemos a análise ANOVA da resposta ao PBR por condição experimental.

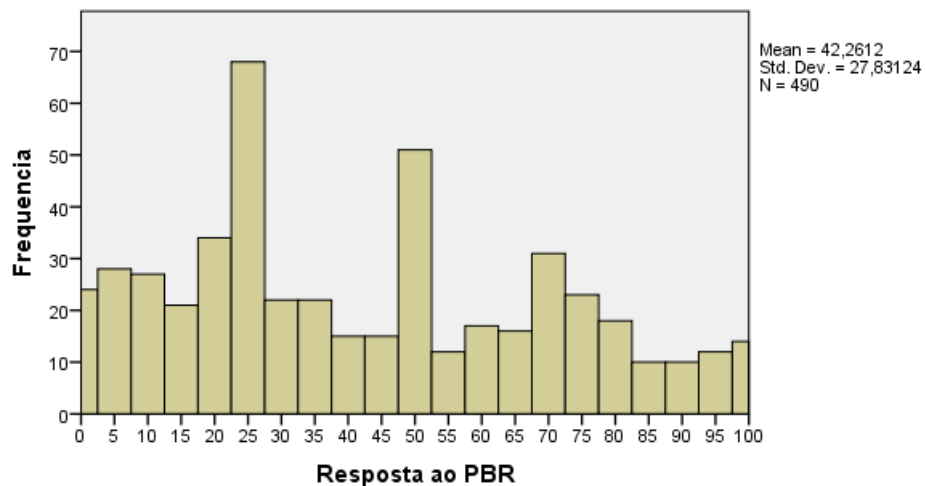
Tabela 12: ANOVA da Resposta ao PBR VS Condição experimental.

	Soma dos Quadrados	gl	Média dos Quadrados	F	Sig.
Entre Grupos	38750	2	19375	27	,000
No Mesmo Grupo	1068673	1467	728		
Total	1107423	1469			

De seguida analisámos a distribuição das respostas dadas aos PBR em frequência absoluta. Esta análise permitiu-nos determinar se os dados se distribuem de forma unimodal e simétrica, como se esperaria se os participantes estiverem todos a usar a mesma estratégia de resposta, ou se pelo contrário existe mais do que uma moda, o que pode significar a existência de mais de uma estratégia de resposta aos PBR.

O gráfico 45 apresenta a frequência de respostas ao PBR na condição 25%.

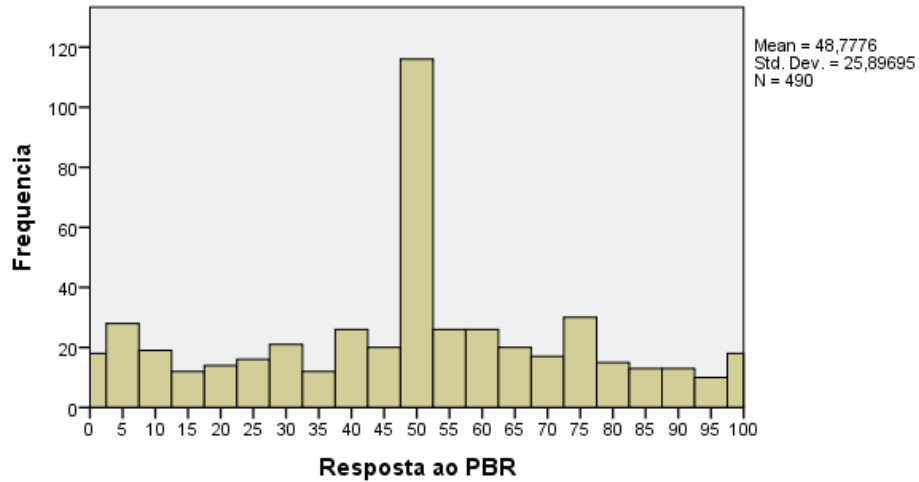
Gráfico 45: Frequência de respostas ao PBR na condição 25%.



Neste gráfico podemos observar que, embora haja um pico de respostas em torno do valor 25% também existe um pico de respostas em torno do valor 50% e um terceiro pico de respostas em torno do valor 70% e 75%.

O gráfico 46 apresenta a frequência de respostas ao PBR na condição 50%.

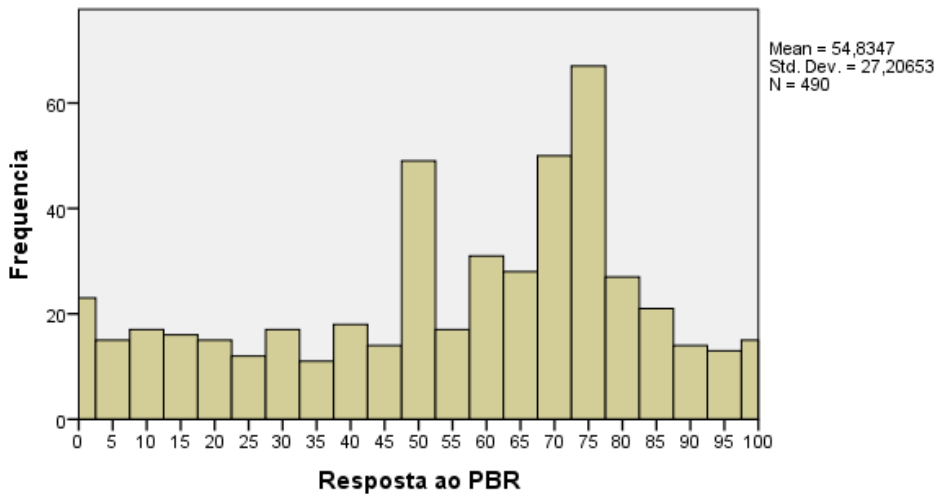
Gráfico 46: Frequência de respostas ao PBR na condição 50%.



Neste gráfico podemos observar que, há um pico de respostas em torno do valor 50% e que as outras respostas se distribuem homogeneamente pela escala de respostas.

O gráfico 47 apresenta a frequência de respostas ao PBR na condição 50%.

Gráfico 47: Frequência de respostas ao PBR na condição 75%.

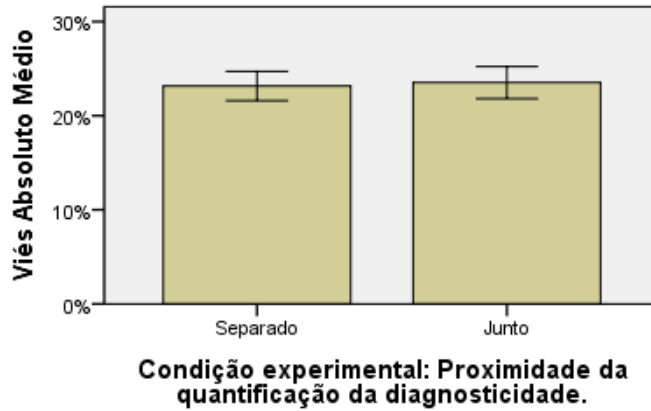


Neste gráfico podemos observar que, embora haja um pico de respostas em torno do valor 75% também existe um pico de respostas em torno do valor 50%.

Este padrão de respostas é semelhante ao da experiência 1 com a exceção de que em todas as condições parece haver um elevado número de participantes a responder próximo do valor de 50%, o que é um resultado inesperado.

De seguida analisamos o efeito da proximidade da quantificação da diagnosticidade no desempenho em PBR. O gráfico 48 representa o viés absoluto médio nas duas condições.

Gráfico 48: Desempenho nos PBR em função da condição experimental.



Não se encontraram diferenças significativas entre estas duas condições.

De seguida, correlacionamos o desempenho nos PBR com as medidas complementares de racionalidade (CRT), inteligência (MPR), estilo cognitivo (NFC e FI) e a média do secundário. A tabela 13 apresenta estas correlações.

Tabela 13: Correlação entre o Viés Absoluto Médio e medidas complementares:

		CRT	MPR	NFC	FI	MS
VAM	Correlação de Pearson	-,042	-,090	-,062	,013	-,157*
	Sig.	,511	,160	,330	,836	,014
	N	245	245	245	245	245

*. A correlação é significativa com $p < 0,05$.

Não encontramos correlações significativas entre estes índices, com a exceção da média do secundário que era uma variável de controlo.

Depois da remoção de outliers as correlações mantiveram-se iguais. A tabela 14 apresenta estas correlações após a remoção de outliers.

Tabela 14: Correlações após a remoção de outliers.

		CRT	MPR	NFC	FI	MS
VAM	Correlação de Pearson	-,026	-,076	,009	,037	-,162*
	Sig.	,712	,284	,903	,603	,022
	N	198	198	198	198	198

*. A correlação é significativa com $p < 0,05$.

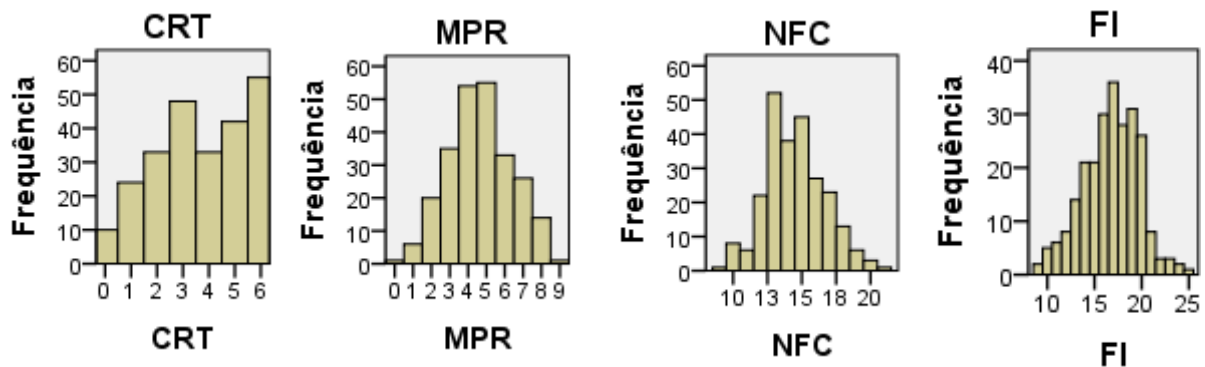
De seguida fizemos a mesma análise contrastando os participantes com bom desempenho em PBR (VAM<10%) com os participantes com mau desempenho em PBR (VAM>10%). No entanto, nesta análise também não encontramos correlações significativas. A tabela 15 apresenta a correlação de contraste de desempenho nos PBR com as medidas complementares.

Tabela 15: Correlação com contrastes: VAM < 10% VS VAM >= 10%

		CRT	MPR	NFC	FI	MS
VAM	Correlação de Pearson	-,101	-,087	,095	,039	,057
	Sig.	,115	,175	,138	,546	,371
	N	245	245	245	245	245

O gráfico 49 apresenta a distribuição das frequências das medidas complementares.

Gráfico 49: Distribuição das medidas de racionalidade, inteligência e disposição cognitiva.



Nestes quatro gráficos podemos ver que com a exceção de um ligeiro efeito de teto no CRT, todas as medidas têm uma distribuição simétrica. Por esta razão, um efeito de assimetria não poderá explicar a não existência de uma correlação entre as medidas.

Na tabela 16 podemos observar as correlações entre as medidas complementares.

Tabela 16: Correlações entre as medidas complementares.

		CRT	MPR	NFC	FI	MS
CRT	Correlação de Pearson	1	,390**	-,040	-,046	,164*
	Sig.		,000	,532	,475	,010
	N	245	245	245	245	245
MPR	Correlação de Pearson	,390**	1	-,039	-,087	,114
	Sig.	,000		,541	,176	,075
	N	245	245	245	245	245
NFC	Correlação de Pearson	-,040	-,039	1	,052	-,015
	Sig.	,532	,541		,420	,809
	N	245	245	245	245	245
FI	Correlação de Pearson	-,046	-,087	,052	1	,009
	Sig.	,475	,176	,420		,891
	N	245	245	245	245	245

** A correlação é significativa com $p < 0,01$. *A correlação é significativa com $p < 0,05$.

Nesta tabela podemos observar que o desempenho dos participantes no CRT e nas MPR está correlacionado significativamente. O desempenho dos participantes no CRT também está correlacionado significativamente com a média no secundário. Mais nenhum par de medidas complementares se encontrou significativamente correlacionado. Esta análise permite-nos confirmar que as medidas complementares estão correlacionadas e que a falta de correlação entre o desempenho em PBR e o desempenho com estas medidas se deve provavelmente à má qualidade dos dados obtidos no desempenho em PBR.

Experiência 4 - Base-Rates Adaptados (Replicação):

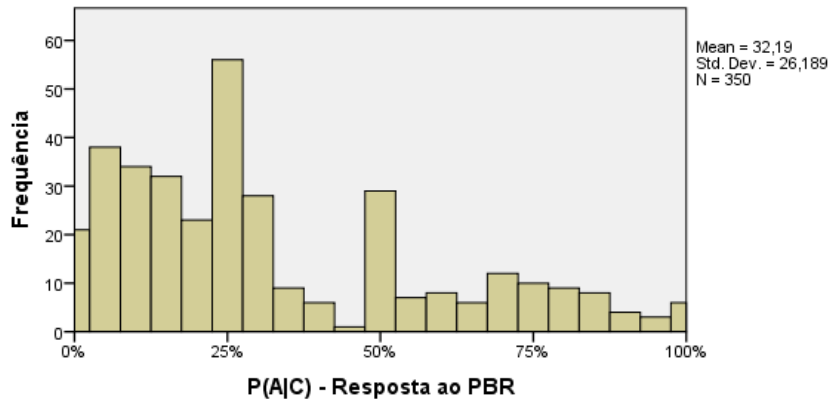
Na tabela 17 vemos a análise ANOVA da resposta ao PBR por condição experimental.

Tabela 17: ANOVA da Resposta ao PBR VS Condição experimental.

	Soma dos Quadrados	gl	Média Dos Quadrados	F	Sig.
Entre Grupos	86892	2	43446	71	,000
No Mesmo Grupo	648278	1054	615		
Total	735170	1056			

De seguida analisámos a frequência das respostas por cada condição experimental. O gráfico 50 apresenta a frequência de respostas ao PBR na condição 25%.

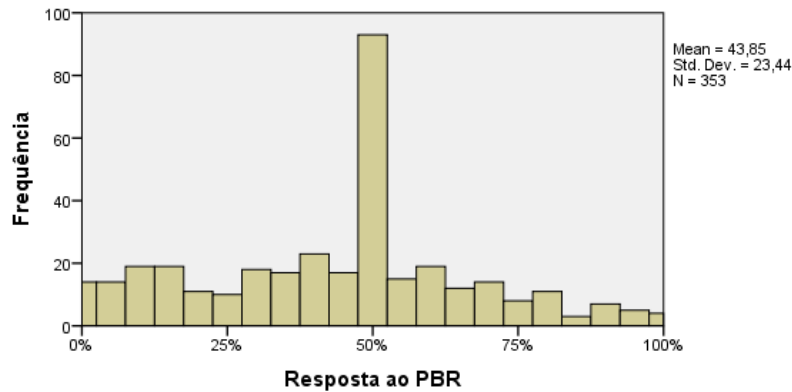
Gráfico 50: Frequência de respostas ao PBR na condição 25%.



Neste gráfico podemos observar que, embora haja um pico de respostas em torno do valor 25% também existe um pico de respostas em torno do valor 50% e algumas respostas em torno do valor 70%. Em comparação com a experiência 3 existem mais respostas abaixo dos 25% e menos respostas acima de 50%.

O gráfico 51 apresenta a frequência de respostas ao PBR na condição 50%.

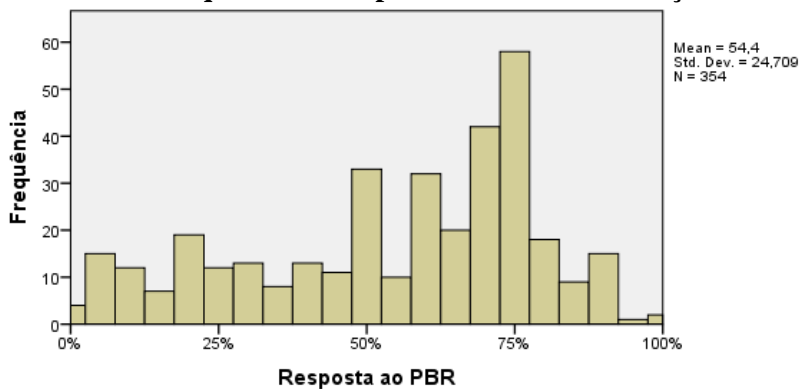
Gráfico 51: Frequência de respostas ao PBR na condição 50%.



Neste gráfico podemos observar que, há um pico de respostas em torno do valor 50% e que as restantes respostas se distribuem homogeneamente pela escala de respostas.

O gráfico 52 apresenta a frequência de respostas ao PBR na condição 75%.

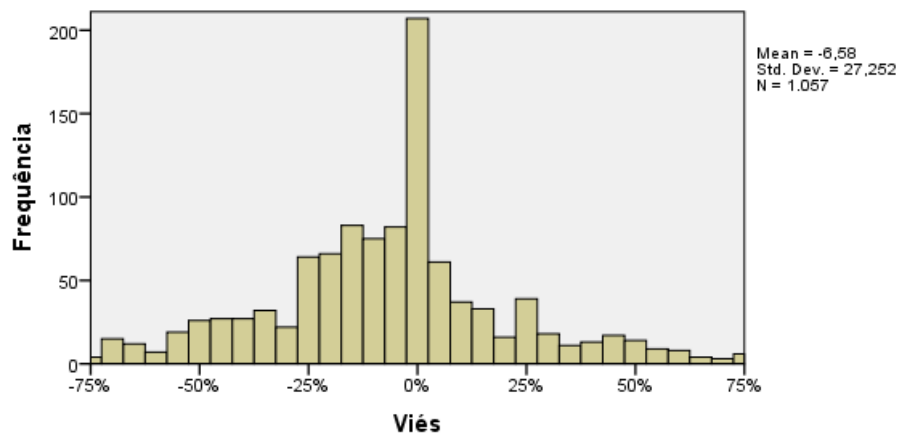
Gráfico 52: Frequência de respostas ao PBR na condição 75%.



Neste gráfico podemos observar que, embora haja um pico de respostas em torno do valor 75% também existe um pico de respostas em torno do valor 50%.

O gráfico 53 apresenta a frequência do viés à resposta bayesiana.

Gráfico 53: Frequência do viés à resposta bayesiana.



Neste gráfico podemos observar um pico de respostas em torno da resposta bayesiana, mas também podemos observar um elevado número de vieses negativos entre -25% e 0%.

Na tabela 18 estão as correlações entre o viés absoluto médio e as medidas complementares

Tabela 18: Correlação entre o Viés Absoluto Médio e as medidas complementares:

		CRT	MPR	NFC	FI	MS
VAM	Correlação de Pearson	-,169*	-,133	,025	,072	-,096
	Sig.	,023	,074	,737	,337	,206
	N	181	181	181	181	176

*. A correlação é significativa com $p < 0,05$.

Na tabela 23 podemos observar a correlação entre o viés absoluto médio e as medidas complementares após a remoção de outliers.

Tabela 18: Correlações após a remoção de outliers.

		CRT	MPR	NFC	FI	MS
Viés Absoluto Médio	Correlação de Pearson	-,136	-,171*	,029	,014	-,102
	Sig.	,111	,044	,734	,874	,235
	N	139	139	139	139	137

*. A correlação é significativa com $p < 0,05$.

Podemos observar que, após a remoção dos participantes considerados outliers, o desempenho em PBR se torna significativamente correlacionado com o desempenho nas MPR. No entanto a remoção destes participantes fez com que a correlação com o CRT deixasse de ser significativa. Este resultado não é surpreendente, pois o facto de retirarmos os participantes com pior desempenho nos PBR estamos também a retirar da amostra os participantes com pior desempenho no CRT, mas não nas MPR, o que explicaria este resultado. Para este fim analisámos a média do desempenho e variáveis complementares nos dois grupos.

Na tabela 18 podemos ver a média de todas as medidas para outliers e não outliers.

Tabela 18: Média de várias medidas para os grupos de outliers e não outliers.

Outlier?		VAM	CRT	MPR	NFC	FI	MS
Não	Média	17,4	3,3	4,3	14,4	15,2	15,4
	N	139	139	139	139	139	137
	Desvio Padrão	8	2	2	2	2	2
Sim	Mean	33,2	2,8	4,1	14,4	15,6	15,3
	N	42	42	42	42	42	39
	Desvio Padrão	9	2	2	3	2	1
Total	Mean	21,0	3,2	4,2	14,4	15,3	15,4
	N	181	181	181	181	181	176
	Desvio Padrão	10	2	2	2	2	2

Podemos observar que, tal como esperado, os participantes considerados outliers têm um desempenho no CRT pior do que os não são outliers e que esta diferença não é significativa no desempenho das MPR. Este resultado é natural pois os outliers foram retirados em função de critérios de resposta desviantes em PBR. Visto que esta medida se correlaciona mais fortemente com o CRT do que com as MPR, é natural que a remoção de outliers afete mais estas duas medidas do que as MPR.

A tabela 19 apresenta a correlação de contraste com as medidas complementares.

Tabela 19: Correlação com contrastes: $VAM < 10\%$ VS $VAM \geq 10\%$

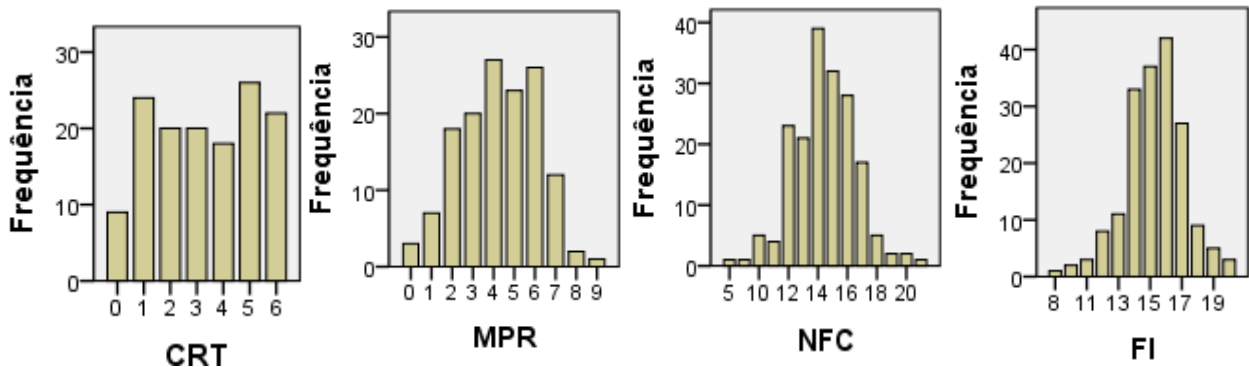
		CRT	MPR	NFC	FI	MS
VAM	Correlação de Pearson	,176*	,140	-,039	-,077	,111
	Sig.	,018	,060	,598	,300	,142
	N	181	181	181	181	176

*. A correlação é significativa com $p < 0,05$.

A análise de correlação com contrastes melhorou a significância de todas as correlações, no entanto as conclusões são semelhantes às análises principal da correlação linear. A melhoria de todas as correlações significa que o ponto de contraste onde o viés absoluto médio é 10% é um bom ponto de contraste em comparação com a utilização de toda a medida.

No gráfico 54 apresenta a distribuição das frequências das medidas complementares.

Gráfico 54: Distribuição das medidas de racionalidade, inteligência e disposição cognitiva.



Nestes quatro gráficos podemos ver que todas as medidas têm uma distribuição simétrica. Por esta razão, um efeito de assimetria não é uma boa explicação para a não existência de correlação entre as medidas.

Na tabela 20 podemos observar as correlações entre as medidas complementares.

Tabela 20: Correlações entre as medidas complementares.

		MPR	NFC	FI	MS
Cognitive Reflection Test	Correlação de Pearson	,283**	-,174*	-,188*	,017
	Sig.	,000	,019	,011	,826
	N	181	181	181	176
Matrizes Progressivas de Raven	Correlação de Pearson	1	-,091	-,102	,138
	Sig.		,223	,173	,068
	N	181	181	181	176
Need For Cognition	Correlação de Pearson	-,091	1	,827**	-,116
	Sig.	,223		,000	,126
	N	181	181	181	176
Faith In Intuition	Correlação de Pearson	-,102	,827**	1	-,085
	Sig.	,173	,000		,260
	N	181	181	181	176

** . A correlação é significativa com $p < 0,01$.

* . A correlação é significativa com $p < 0,05$.

Nesta tabela podemos observar que o desempenho dos participantes no CRT se correlaciona significativamente com o desempenho nas MPR, pontuação na escala NFC e FI, mas não com a média do secundário. Para além disso, neste estudo a escala NFC e FI estão significativamente correlacionadas. O desempenho dos participantes no CRT também está correlacionado significativamente com a média no secundário. Mais nenhum par de medidas complementares se encontrou significativamente correlacionado.

Na tabela 21 podemos observar o viés médio e absoluto médio em função da amostra.

Tabela 21: Viés Médio e Viés Absoluto Médio em função do tipo amostra.

Método de amostragem		Viés	Viés Absoluto
Laboratório	Média	-4,1	19,6
	N	452	452
	Desvio Padrão	26,7	18,6
Fora de laboratório	Média	-8,5	21,1
	N	605	605
	Desvio Padrão	27,5	19,6
Total	Média	-6,6	20,5
	N	1057	1057
	Desvio Padrão	27,3	19,2

Observámos que a amostra de laboratório teve um desempenho ligeiramente melhor e menos enviesado que a amostra recolhida em laboratório. Enquanto a melhoria no enviesamento é significativa ($p < 0,02$) a melhoria no desempenho não é significativa ($p > 0,2$).

Experiência 5:

Na tabela 22 podemos observar os grupos, a descrição e a diagnosticidade dos PBR.

Tabela 22: Problemas de Base-Rate: Grupos, Descrição e Diagnosticidade

Grupo A	Grupo B	Descrição	Diagnosticidade
Rapaz	Rapariga	Gosta de sair à noite com amigos, ouvir música e beber cerveja.	52
Psicoterapeuta	Oficial Militar	É amável e compreensivo.	58
Italiana	Americana	É bilingue, gosta de ler nos tempos livres e de cozinhar.	60
Brasileiro	Português	Gosta de dançar o samba.	65
Engenheiros	Advogados	Gosta de ler ficção científica e programar no computador.	68
20 anos	50 anos	Gosta de ouvir música techno e eletrónica, usa jeans e óculos escuros, tem um pequeno piercing no nariz e gostar de dançar.	70
Só comprar roupas em viagens ao estrangeiro	Só comprar roupas em supermercados	Vive no último andar de um dos melhores prédios da cidade e conduz um Porche.	73
Rapazes	Raparigas	Ter mais de 1,7 metros de altura	64
Budistas	Muçulmanos	Gosta de filosofia, tem aversão a bens materiais, usa habitualmente roupa em segunda mão.	63
Homem rural	Homem urbano	Usa boina ou bigode.	63

Tabela 23: Problemas de Base-Rate: Grupos, Resposta bayesiana e base-rate do grupo A.

Grupo A	Grupo B	Resposta Bayesiana	Base-Rate	
			PBR Inclusivo	PBR Exclusivo
Rapaz	Rapariga	45	53	43
Psicoterapeuta	Oficial Militar	65	57	28
Italiana	Americana	20	73	14
Brasileiro	Português	80	68	12
Engenheiros	Advogados	30	52	17
20 anos	50 anos	15	71	7
Só comprar roupas em viagens ao estrangeiro	Só comprar roupas em supermercados	45	68	6
Rapazes	Raparigas	65	X	6
Budistas	Muçulmanos	20	X	17
Homem rural	Homem urbano	80	X	28

NOTA: Para manter a resposta bayesiana igual na condição inclusiva e exclusiva, por vezes o grupo alvo do PBR era o grupo B em vez de ser o grupo A. Esta situação ocorreu na condição inclusiva para os problemas 1, 3, 5 e 6 e na condição exclusiva nos problemas 2, 4 e 7. Na fase de teste (PBR 8, 9 e 10) o alvo era sempre o grupo A.

Na tabela 23 podemos ver o output do teste T de Student que compara a média entre o viés absoluto médio da condição IS e IF na fase de aprendizagem.

Tabela 23: Diferença entre o desempenho na condição IS e IF: Fase de Aprendizagem

		Teste de Levene		Teste t de Student		
		F	Sig.	t	gl	Sig.
Viés Absoluto	Variâncias iguais	,067	,797	2,306	84	,024
Médio	Variâncias diferentes			2,303	83,041	,024

Na tabela 24 podemos ver o output do teste T de Student que compara a média entre o viés absoluto médio da condição ES e EF na fase de aprendizagem.

Tabela 24: Diferença entre o desempenho na condição ES e EF: Fase de Aprendizagem

		Teste de Levene		Teste t de Student		
		F	Sig.	t	gl	Sig.
Viés Absoluto	Variâncias iguais	,079	,779	1,311	83	,193
	Variâncias diferentes			1,312	83,000	,193

Na tabela 25 podemos ver o output do teste T de Student que compara a média entre o viés absoluto médio da condição IS e IF na fase de teste.

Tabela 25: Diferença entre o desempenho na condição IS e IF: Fase de Teste

		Teste de Levene		Teste t de Student		
		F	Sig.	t	gl	Sig.
Viés Absoluto	Variâncias iguais	,601	,440	,586	84	,560
	Variâncias diferentes			,584	81,434	,561

Na tabela 25 podemos ver o output do teste T de Student que compara a média entre o viés absoluto médio da condição ES e EF na fase de teste.

Tabela 26: Diferença entre o desempenho na condição ES e EF: Fase de Teste

		Teste de Levene		Teste t de Student		
		F	Sig.	t	gl	Sig.
Viés Absoluto	Variâncias iguais	,533	,467	1,767	83	,081
	Variâncias diferentes			1,765	81,596	,081

Na tabela 27 podemos ver o output do teste T de Student que compara a média entre o viés absoluto médio da condição ES e EF no PBR 10.

Tabela 27: Diferença entre o desempenho na condição ES e EF: PBR 10

		Teste de Levene		Teste t de Student		
		F	Sig.	t	gl	Sig.
Viés Absoluto	Variâncias iguais	,902	,345	2,389	83	,019
	Variâncias diferentes			2,383	79,118	,020