

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA**

HENRIQUE CASSOL LEAL

**RAZÕES EXIGEM DESEJOS? UMA ANÁLISE DA TEORIA DE RAZÕES PARA
AÇÃO DE BERNARD WILLIAMS**

PORTO ALEGRE

2020

HENRIQUE CASSOL LEAL

**RAZÕES EXIGEM DESEJOS? UMA ANÁLISE DA TEORIA DE RAZÕES PARA
AÇÃO DE BERNARD WILLIAMS**

Trabalho de Conclusão de Curso
apresentado para o Instituto de Filosofia e
Ciências Humanas da UFRGS para a
obtenção do título de Bacharel em Filosofia.

Orientador: Prof. Dr. David Horst

PORTO ALEGRE

2020

CIP - Catalogação na Publicação

Cassol Leal, Henrique

Razões exigem desejos? Uma análise da Teoria de
Razões para Ação de Bernard Williams RAZÕES
EXIGEM DESEJOS? / Henrique Cassol Leal. -- 2020.
50 f.

Orientador: David Horst.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto de
Filosofia e Ciências Humanas, Bacharelado em Filosofia, Porto
Alegre, BR-RS, 2020.

1. Bernard Williams. 2. John McDowell. 3. Christine
Korsgaard. 4. Internalismo de Razões. 5. Razões Internas e
Externas. I. Horst, David, orient. II.
Título.

HENRIQUE CASSOL LEAL

**RAZÕES EXIGEM DESEJOS? UMA ANÁLISE DA TEORIA DE RAZÕES PARA
AÇÃO DE BERNARD WILLIAMS**

Trabalho de Conclusão de Curso
apresentado para o Instituto de Filosofia e
Ciências Humanas da UFRGS para a
obtenção do título de Bacharel em Filosofia.

Porto Alegre, 13 de novembro de 2020.

BANCA EXAMINADORA:

Doutor André Nilo Klaudat
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS - UFRGS

Doutorando Rafael Graebin Vogelmann
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS – UFRGS

Agradecimentos:

Agradeço ao meu orientador, David Horst, por me auxiliar na criação e correção deste trabalho, responder meus e-mails longuíssimos em uma velocidade impressionante e ser um exemplo de filósofo para eu me espelhar.

Agradeço a minha família (em especial meus pais e irmão) por me oferecerem a oportunidade de pesquisar com tranquilidade e de realizar meus sonhos de buscar a profissão que escolhi. Devo agradecer, entre outras coisas, principalmente por oferecerem muito mais amor do que mereço.

Agradeço a minha namorada, Roberta, por sempre me escutar ao falar de assuntos de filosofia por vezes chatos (para os outros), por me incentivar a me organizar bem e por me acompanhar por mais de quatro anos.

Também agradeço aos vários alunos e professores que durante o curso de graduação pude conversar e aprender muito. Em especial, não posso deixar de fora a gratidão e admiração que tenho pelo professor Alfredo Storck, que também me acolheu como aluno de iniciação científica por mais de dois anos.

RESUMO:

O presente trabalho visa reconstruir e analisar os argumentos de Bernard Williams a favor do seu Internalismo de Razões. Grosso modo, podemos entendê-lo como defendendo que razões para ação não só causam uma motivação nova nos agentes a partir da deliberação correta e correção de crenças, como elas também dependem da motivação anterior deles (desejos) para poder motivá-los. Uma reconstrução da teoria de Williams será feita no capítulo 1, baseando-se principalmente em dois comentadores (Setiya e Paakkunainen). No capítulo 2, os principais argumentos e vantagens que Williams apresenta a favor de sua teoria serão reconstruídos e discutidos. No capítulo 3, algumas das principais inadequações extensionais da teoria williamsiana serão abordadas à luz de incompatibilidades com um dos pressupostos de Williams: a Restrição Deliberativa. Quando possível, essas inadequações serão respondidas e, caso contrário, pelo menos mitigadas.

Palavras-Chaves: Bernard Williams. Razões para ação. Normatividade. Internalismo de Razões. Razões internas e externas.

ABSTRACT

This work aims to reconstruct and analyze the arguments from Bernard Williams in favor of his Internalism about Reasons. We can roughly understand it as defending that reasons for action not only cause a new motivation in agents through sound deliberation and correction of beliefs, but they also depend on their prior motivation (desires) to motivate. A reconstruction of Williams' theory will be done in chapter 1, based mainly on two commentators (Setiya and Paakkunainen). In chapter 2, the main arguments and advantages Williams presents in favor of his theory will also be reconstructed and discussed. In chapter 3, some of the main extensional inadequacies of Williams' theory will be approached in the light of incompatibilities with a presupposition of Williams: The Deliberative Constraint. When possible, these inadequacies will be answered, and otherwise, at least mitigated.

Keywords: Bernard Williams. Reasons for action. Normativity. Reasons Internalism. Internal and External Reasons.

SUMÁRIO

Introdução	1
Capítulo 1: Apresentação inicial da Teoria de Razões para Ação de Bernard Williams	2
1.1 Formulação inicial do Internalismo Williamsiano	2
1.2 Williams a partir de Setiya, Restrição Explanatória e dois aspectos da crença.....	9
1.3 Interpretando a capacidade de ser motivado	14
1.4 Razões pro-tanto vs. razões decisivas; e motivar a φ vs. levar a φ	16
1.5 Comparando a reconstrução de TRW com a de Paakkunainen.....	20
Capítulo 2: Análise dos argumentos de Williams a favor de sua teoria	25
2.1 Restrição Explanatória (RE) e interpretação da capacidade explicativa	25
2.2 Vantagem de TRW': relação simples entre razões	31
2.3 Argumento do ônus de prova e comparação com Korsgaard	32
Capítulo 3: Inadequações extensionais e Restrição Deliberativa (RD)	35
3.1 Analisando objeções: otimismo deliberativo e falácia condicional.....	35
3.2 Contraexemplos a RD: James Bond, Soldado; e racionalidade atípica	38
3.3 Por RD: relação intuitiva e demandas autoritárias	44
Conclusões	47
Referências	50

INTRODUÇÃO

Em “Internal and External Reasons (1979), Bernard Williams lança uma teoria de razões para ação. Isto é, uma teoria que estabelece¹ quando temos ou não razão para fazer alguma coisa. Ao longo de nossas vidas, e mesmo cotidianamente, passamos por situações em que a incerteza sobre qual o melhor curso de ação nos consome. Ficamos indecisos sobre qual faculdade escolher. Ficamos indecisos sobre se devemos nos atrasar para nossos compromissos para que possamos socorrer alguém que está passando mal em nossa frente. Ficamos indecisos se devemos desistir de certos projetos quando eles parecem não funcionar, ou se deveríamos persistir neles – não importando o quão implausíveis eles sejam. Nessas e em várias outras situações buscamos encontrar razões para realizar certa ação no lugar de outras. Quando Williams estabelece sua Teoria de Razões para Ação, ele aponta que todas as razões que temos devem depender de certos estados internos nossos: de nossa motivação.

Um primeiro objetivo deste trabalho consiste em reconstruir essa teoria de Williams. Uma apresentação inicial dela, e desse objetivo, se dará no capítulo 1. Nele, o funcionamento interno da teoria de Williams será explorado, visando primeiramente responder porque quando buscamos determinar uma razão para ação algumas circunstâncias precisam ser satisfeitas: uma vez que a motivação é importante, nossas crenças devem ser corrigidas e devemos deliberar do melhor modo possível. Além disso, uma reconstrução inicial da teoria de Williams será feita com base nas duas teses identificadas por Setiya (2011, p.6). Poderemos afirmar a partir disso que o Internalismo de Razões de Williams defende, grosso modo, que 1) razões dependem de motivação; e 2) que essa motivação que surge através da razão só faz isso quando apoiada pela motivação anterior do agente. Uma vez que o entendimento de 2 é abrangentemente entendido como desejos, poderemos dizer que razões dependem não somente de motivação, mas também de desejos. Isso já aponta uma divergência de Williams e de outras teorias de razões para ação concorrentes a ele. Como Setiya (2011, p.6) ainda aponta, algumas teorias internalistas defenderão 1, mas não 2. Já outras teorias de razões para ação, não

¹ No início, Williams parecia entendê-la como estabelecendo condições suficientes e necessárias, mas como podemos perceber em “Internal Reasons and the Obscurity of Blame” (1995), ele passa a acreditar que seus argumentos só estabelecem condições necessárias.

aceitarão nenhuma dessas duas, podendo ser chamadas de externalistas (como a de McDowell). Outros detalhes interpretativos também serão neste capítulo explorados: qual exatamente é a importância das crenças para o Internalismo de Williams, de quais modos devemos pensar agentes em situações contrafactuais para que eles tenham uma razão para ação, se a deliberação sólida deve levar um agente a agir por razões ou a ser motivado por razões, entre outros. No fim deste capítulo, será feita uma comparação da reconstrução feita até o momento com a de Paakkunainen (2018, p.13). Por fim, será defendida uma reconstrução minha baseada nos pontos anteriores.

Já no capítulo 2, serão apresentados e analisados dois argumentos a favor da teoria de Williams e uma certa vantagem. Os argumentos são um argumento por melhor explicação para sua teoria e um argumento que aponta um ônus de prova para seus adversários – os externalistas. Essa vantagem comparativa será a da simplicidade que a teoria de Williams consegue propor entre as relações de razões normativas, explicativas e motivacionais.

Por fim, no capítulo 3, um dos pressupostos de Williams, a Restrição Deliberativa (RD), será analisado à luz de dois contraexemplos, uma suposta falácia (a falácia condicional), e de dois argumentos a seu favor. *James Bond* e *Soldado em um Guerra Justa* são contraexemplos à Restrição Deliberativa, ao passo que a relação intuitiva entre razões e deliberação e de razões e demandas autoritárias serão abordadas como argumentos por RD. Paakkunainen aponta como esse segundo argumento pode ser usado contra Williams (PAAKKUNAINEN, 2018, p.22). Apontarei que essa objeção se deve a uma outra intuição que se tem contra uma teoria como a de Williams: o Absolutismo Moral (AM). Já a falácia condicional é uma que tentarei apontar que podemos acusar McDowell, mas não Williams de cometer.

Capítulo 1: Apresentação Inicial da Teoria de Razões para Ação de Bernard Williams

1.1) Formulação inicial do Internalismo Williamsiano

Imagine a seguinte situação: Ana tem uma doença grave, o que a leva a precisar de certos remédios fortes. Esses remédios, entretanto, têm certos efeitos colaterais

extremamente indesejáveis para ela. Devido a isso, ela não pretende fazer uso deles – mesmo que eles possam salvar sua vida. Poderíamos nos perguntar: ainda assim, ela tem uma razão para tomá-los? Estará justificada em fazê-lo? Ou ainda: ela *deveria* fazer isso? A resposta a essas perguntas exige *uma teoria de razões para ação*. Bernard Williams é o responsável por uma teoria como essa que foi altamente influente². Para responder à Ana, ele poderia começar dando uma resposta geral para ela. Ele diria: para *determinar* se ela tem uma razão para fazer isso, é preciso saber se a motivação dela *pode* estar em uma relação apropriada com a proposição de razão (“Ana tem uma razão para tomar o remédio”), de modo que essa última seja considerada verdadeira³. Isto é, *devemos levar em conta a sua motivação para estabelecer se ela tem razão para tomar ou não seu remédio*. Assim, Williams defende que teorias de razão para ação devem ser *internalistas*. Ou seja, ele argumenta que elas devem fazer referência a estados internos de agentes (no seu caso, como veremos, motivação e desejos⁴). Se uma proposição de razão (como “Ana deve tomar o remédio”) tiver essa relação apropriada com os estados internos do agente e se seu conteúdo satisfizer ou promover algum desejo dele, poderemos então dizer que ela é verdadeira.

A seguir, Williams poderia continuar a explicar para Ana quais são as *circunstâncias dessa relação apropriada* – entre os estados internos de um agente, ou sua motivação, e a proposição de razão. Somente com tais circunstâncias poderemos *evitar erros* que levariam Ana, ou qualquer outro, a concluir algo de modo injustificado ou que mesmo a impediriam de chegar a conclusões razoavelmente. Mas antes disso, pretendo esclarecer o que Williams quer dizer com motivação e estados internos do agente. Por esses termos, ele se refere a todo elemento que constitui o *conjunto motivacional subjetivo* de um agente – também chamado de “S”. (Daqui em diante, chamarei apenas de “conjunto motivacional” ou

² “O texto de Williams renovou o interesse da filosofia da teoria da ação e da psicologia moral à ética; ele ajudou a iniciar uma investigação mais abrangente de razões não morais e instrumentais; e ele ofereceu uma perspectiva fresca sobre as questões da natureza da normatividade em geral.” (SETIYA, 2011, p.2, tradução minha)

³ “Basicamente, e por definição, qualquer modelo para uma interpretação internalista precisa demonstrar uma relatividade de proposições de razão para com o conjunto motivacional subjetivo do agente, que chamarei de S.” (WILLIAMS, 1979, p.102, tradução minha). Na verdade, no início desse texto de Williams, ele parece estar mais entendendo essas proposições de razão como “P é uma razão para A fazer ϕ ”. Essa distinção para a do exemplo de Ana não é relevante para este ponto, mas abordarei ela no final da seção 1.2.

⁴ Por ora, podemos entender motivação e desejos de modo intercambiável.

“S”). Esses elementos são a coleção de motivos, disposições de avaliações, padrões de reação emocional, lealdades pessoais, diversos projetos e mesmo – embora de modo indireto⁵ – crenças de um agente. Eles podem ser também entendidos como *desejos*. As proposições de razões internas se referem a eles. E quando elas os promovem ou os satisfazem, estando as outras circunstâncias atendidas, podemos dizer que o agente tem uma razão para agir. Por exemplo, se Ana tem o motivo ou desejo de sobreviver, é bem provável que aquele remédio promoverá elementos de seu S e assim a proposição “Ana tem uma razão para tomar o remédio” será verdadeira. Digo ‘bem provável’ porque aquelas circunstâncias precisam primeiramente ser atendidas para que a relação da proposição com o conjunto motivacional seja apropriada. Assim, voltemos a elas.

A *primeira dessas circunstâncias* é a de se ter posse de todas as crenças relevantes sobre fazer ou não certa ação ϕ – afinal, a ausência de alguma poderá levar o agente a *não ter a motivação* para agir. Parece intuitivo que se Ana não souber uma informação relevante, como a de que esse remédio prolongará sua vida por vários anos, ela poderá não estar motivada para tomá-lo. Assim, no seu conjunto motivacional, haverá uma ausência de elementos a serem promovidos ou satisfeitos pela proposição (devido à falta de uma informação). E, desse modo, Ana não poderá estar em uma posição adequada para dizer se ela tem ou não razão para fazer essa ação. A *segunda circunstância*, ligada à primeira, é a de que suas crenças nas informações relevantes devem ser verdadeiras⁶⁷. Afinal, se ela possuísse uma crença falsa que diria que há 50% de chance de se ter um ataque cardíaco a cada vez que ela tomasse esse remédio, sua motivação seria provavelmente a de não tomá-lo. E, nesse caso, seu julgamento sobre o assunto

⁵ Crenças fazem parte do conjunto motivacional na medida em que elas afetam os outros elementos. Se creio que há um copo de gim na minha frente, esse pode ser um motivo para desejar tomá-lo. (WILLIAMS, 1979, p.105) O ponto de como crenças afetam exatamente nossa motivação será abordado em especial no final da seção 1.2.

⁶ Pode-se pensar que podemos reduzir as duas circunstâncias a uma só ao falarmos em “fatos relevantes”, mas como ficará mais claro depois, pensar em termos de crenças é essencial para a teoria de Williams. Queremos que o agente tenha posse desse fato (sendo um link psicológico) para que seus estados internos possam ser influenciados – no lugar de esse ser um mero fato no mundo.

⁷ Com essa circunstância também se deve entender a exclusão de falsidade por associação, como no caso de uma crença depender de outras crenças falsas ou de até mesmo a crença ser verdadeira, mas o agente ter uma crença falsa quanto a importância daquela para se atingir certo fim. Podemos ver esses detalhes com a seguinte tese de Williams:“(ii) Um membro de S, D, fornece uma razão para ϕ , se a existência de D independe de crenças falsas ou se a crença do agente na relevância de ϕ para a satisfação de D for verdadeira”. (WILLIAMS, 1979, p.102, tradução minha).

seria tão extremamente prejudicado a ponto de poder não ser considerado razoável – e de modo similar para outras crenças menos ou mais influentes. Desta vez, a presença de um elemento injustificado (pois falso) é que impede que haja uma relação apropriada entre a proposição de razão e os estados internos do agente⁸.

Tais circunstâncias podem parecer demasiado exigentes, mas como já foi dito, elas são necessárias para se evitar qualquer erro de julgamento. Afinal, novamente, Williams pretende *garantir* as circunstâncias nas quais podemos dizer que certa proposição indica uma razão para agirmos ou não – e se a for, será uma proposição de razão para ação *verdadeira*. E como vimos, elas funcionam em *contextos contrafactuais* específicos: em contexto que poderíamos chamar de ideais. É somente com eles que podemos garantir *sem erro* que, por exemplo, dado que Ana consegue concluir que ela deve tomar o remédio, Ana tem uma razão para tomar seus medicamentos. (Ou qualquer outro caso de ação justificada). A primeira circunstância serviu para adicionar certos itens aos estados internos do agente que ele pode não ter acesso em um dado momento⁹. A segunda eliminou os itens dos estados internos do agente que poderiam falsificar a proposição de razão¹⁰ – ela ainda pode ser falsa, mas agora não mais por crenças falsas e sim por falta de motivação ou como veremos, uma insuficiência na deliberação. (Passarei a chamar essas duas circunstâncias, se juntas de “correção de crenças”).

Com a correção de crenças, ao modificarmos os itens que compõe o conjunto motivacional do agente, sua motivação poderá estar de um tal modo que as proposições de razão possam melhor ser estabelecidas como falsas ou verdadeiras. Essas circunstâncias oferecem as bases para uma relação apropriada entre os itens de seu conjunto motivacional e a proposição de razão. Entretanto, é somente com um

⁸ Junto com a circunstância anterior, crenças falsas, mas ainda assim irrelevantes, parecem não precisar afetar o possível julgamento sobre a proposição de razão. Por isso, pensamentos falsos como a posição da Terra em relação ao Sol em outros tempos ou meros medos infantis que não afetam o julgamento contrafactual podem ser desconsiderados.

⁹ Além do caso óbvio de o agente meramente não conhecer certa informação pela dificuldade de obtenção no momento (como saber se um banhista está meramente se afogando ou brincando dado a distância), essa circunstância também cobre outros. Por exemplo, é possível que o agente esteja alterado de uma tal maneira que ele não consegue absorver as informações novas de modo que sua motivação possa assim ser alterada. Esse pode ser o caso de quando por profunda irritação com um parente, ignoramos os fatos que eles podem nos trazer e como resultado não mudamos nossos pensamentos e motivações.

¹⁰ Além disso, como Williams especifica na proposição iii) (b), conhecimentos sobre nossos estados internos também podem ser adicionados – seja pela ignorância de um certo fato que levaria o agente a realizar a ação ou mesmo de certo elemento motivacional necessário para a conclusão deliberativa do agente dessa ação (WILLIAMS, 1979, p.103)

papel deliberativo do agente que ele – e nós – pode chegar a conclusões sobre isso. Assim, há ainda dois outros fatores que devemos levar em conta em tais circunstâncias contrafactuais. O primeiro desses é, na verdade, um *pressuposto para as circunstâncias* anteriores e também desse último. Para entendermos a imporância desse pressuposto podemos começar por nos perguntar pela raiz das circunstâncias anteriores. Por que pensamos que as duas primeiras circunstâncias – a adição de informação relevante e a exclusão de falsidades – são importantes? Por que pensar que devemos corrigir nossas crenças? O que isso interfere na determinação de razões?

Pensamos que as duas primeiras circunstâncias são importantes porque a partir delas podemos nos decidir melhor sobre o que fazer – e consideramos isso especialmente relevante em circunstâncias e decisões importantes (PAAKKUNAINEN, 2018, p. 19). Decidimos melhor sobre o que fazer, no geral, quando deliberamos bem¹¹ – à luz de fatos verdadeiros, considerações relevantes e boa deliberação. Assim, um princípio que rege as duas primeiras circunstâncias é a *Restrição Deliberativa (RD)*¹². Essa tese diz que se podemos expressar com verdade que “Ana tem uma razão para tomar um remédio”, então essa mesma proposição deve *poder* ser descoberta por Ana a partir de um exercício de deliberação. Isto é, a pessoa para qual a razão se aplica pode sempre descobrir suas razões – e se ela não o faz é por problema nas duas primeiras circunstâncias ou algum relacionado a deliberação. Assim, podemos dizer que *proposições de razões internas podem ser descobertas pelo raciocínio deliberativo*.

Devido a esse pressuposto, podemos nos perguntar “qual é o tipo de raciocínio correto?”, “com qual tipo de raciocínio Ana poderia concluir que ela tem razão para tomar seus remédios?”. Assim, é no tipo de deliberação a ser realizada pelo agente que estará nossa *última circunstância*: é necessário uma deliberação que também possa nos auxiliar a evitar o erro (por estabelecer as relações apropriadas entre a motivação e a proposição de razão) e concluir sempre que possível sobre se temos

¹¹ Williams aponta as circunstâncias de correção de crenças como relevantes por sua preocupação não apenas com a explicação das ações dos agentes, mas também com a *racionalidade*. (Creio que isso mostra uma proximidade da noção de racionalidade de Williams com a de Hume, também preocupada com a verdade). Penso que esse essa explicação de Paakkunainen também ajuda a explicitar o que Williams entende por esse termo. Por ele, devemos então entender, para Williams, a situação ideal de um agente que satisfaz as três circunstâncias aqui apresentadas (sendo que a terceira circunstância está no parágrafo seguinte).

¹² Deixarei este pressuposto para melhor análise no capítulo 3.

ou não uma razão para ação. Tal deliberação é cunhada por Williams como *deliberação abrangente*. Ela se refere não a um tipo único de deliberação, mas a vários processos deliberativos diferentes que podemos usar para concluir que temos razão para fazer ϕ . Williams sugere vários exemplos. Um primeiro tipo seria uma espécie de raciocínio instrumental. Podemos deliberar que devemos fazer ϕ , uma vez que esse é o modo mais conveniente, prazeroso, econômico, etc. de satisfazer certo elemento do conjunto emocional (S) do agente. (No caso de Ana, tomar seu remédio pode ser o melhor modo em todos esses sentidos). Também podemos citar raciocínios em que se combina o melhor modo de satisfazer os elementos de S, como em uma ordenação temporal – tomo banho para somente então sair com minha namorada. Ou no caso de conflitos, podemos pesar qual opção – e por consequência qual razão – é a mais importante. Outro exemplo seria o de soluções constitutivas: como deliberar que, por exemplo, passear no shopping é uma solução constitutiva para o meu desejo de aproveitar o sábado. É importante enfatizar que podemos também combinar os diferentes tipos de deliberação – ou não, podendo também usar somente um. E a partir dos mais variados modos de se deliberar abrangentemente o agente pode, como nas circunstâncias anteriores, terminar por eliminar ou adicionar certos elementos em seu conjunto motivacional (S).

Os exemplos de deliberação que forneci até agora se tratam apenas dos claros. Mas, como diz o nome, Williams pretende que sua deliberação seja realmente *abrangente*. Desse modo, além dessas, há principalmente as deliberações a partir de considerações advindas da imaginação. Dentre vários usos, com elas somos capazes de até mesmo adotar perspectivas em terceira pessoa que normalmente não teríamos. Assim, por exemplo, mesmo conselhos, encorajamentos e persuasões de outras pessoas também estariam aqui incluídos. É por isso que Williams diz sobre sua teoria:

Não é clara, e eu considero como uma característica fundamentalmente desejável de uma teoria de raciocínio prático que ela deva preservar e dar conta dessa falta de clareza. Há uma indeterminação essencial no que pode ser contado como um processo deliberativo racional. O raciocínio prático é um processo heurístico e imaginativo, e não há limites fixos no contínuo entre pensamento racional para inspiração e conversão. (WILLIAMS, 1979, p.110, tradução minha).

Agora que apresentei as circunstâncias pelas quais é possível haver uma relação apropriada entre a proposição de razão e a motivação do agente, voltemos

ao exemplo inicial. Williams poderia dizer à Ana que se ela estivesse nas circunstâncias de correção de crenças (circunstância 1 e 2) e deliberasse abrangentemente, ela poderia concluir que ela, *agora* (antes de estar nessas condições contrafactuais), de fato tem uma razão para tomar seu remédio – ou não. Se o resultado fosse positivo, ela estaria realizando uma *deliberação sólida* e antes de se aplicar as circunstâncias, já seria possível dizer com verdade que ela possuiria uma *rota deliberativa sólida*¹³. Isto é, uma rota de deliberação correta que a partir da nossa motivação¹⁴ atual e da correção de crenças vai em direção à conclusão que se tem razão para fazer algo. (Creio que Williams chama ela de “rota”, pois a deliberação parte de um lugar, a motivação pré-existente, e visa chegar a outro: a conclusão de que se tem uma razão para fazer algo, o que implica também a promoção ou satisfação dos elementos de S).

Até agora apresentei o funcionamento interno da teoria de Williams e algumas razões para ela – que no geral envolvem encontrar um melhor modo possível de satisfazer a Restrição Deliberativa. Williams pensava que seu modelo de teoria internalista era o único que poderia dar conta da *restrição explicativa* (que veremos em maior detalhe no capítulo 2, mas também no fim da próxima seção) e por isso sua teoria estaria justificada. Como consequência disso, em “Internal and External Reasons” (1979), Williams defende que *razões exigem desejos – desejos sendo os elementos do S*. (Nesse primeiro momento, Williams parece defender que seus argumentos fornecem condições suficientes e necessárias, ao passo que em “Internal Reasons and The Obscurity of Blame” (1995, p.35-36), ele *afirma estar tratando apenas de condições necessárias* de razões para ação). Teses externalistas, que acabam por recusar que razões dependem de desejos, encontrariam-se então em situação muito desfavorável. Mas desde o lançamento desse artigo, em 1979, uma vasta literatura surgiu sobre razões para ação em resposta a Williams. Logo em seguida, parte dessa literatura veio a defender o que Williams argumentou como impossível: uma posição externalista sobre razões para ação, como podemos ver com John McDowell (1995). Outra parte veio a defender versões diferentes do internalismo de razões, como podemos ver com Christine

¹³ Tal formulação pode ser vista no artigo “Internal Reasons and the Obscurity of Blame” (WILLIAMS, 1995, p.3). Nesse artigo Williams retoma e reexplica sua teoria do internalismo de razões.

¹⁴ Que como tentei apontar, não é estática, podendo mudar com cada uma das circunstâncias.

Korsgaard¹⁵. Pelo menos desde essa última autora há disputa sobre a noção de internalismo de razões para ação. Afinal, ela aponta que há uma tese ali embutida que não é exigida pelo internalismo de razões. Autores mais recentes, como Setiya e Paakkunainen reconstróem o argumento de Williams apontando assim pelo menos duas teses: o internalismo de razões – uma versão menos exigente que a de Williams – e um tipo de teoria humeana de motivação. Assim como Korsgaard, eles apontam que é possível – e preferível¹⁶ – descrever o internalismo de razões de modo diferente do qual Williams parecia exigir a princípio.

1.2) Williams a partir de Setiya, Restrição Explanatória e dois aspectos da crença

Setiya reconstrói a teoria de Williams em duas premissas, enquanto que Paakkunainen em mais. Em um primeiro momento, trarei as duas teses que para Setiya compõe a teoria de razões para ação de Williams. Vamos à primeira:

O *Internalismo de Razões (IR)* defende a posição que razões podem ser capazes de nos motivar ao acreditarmos em proposições de razão. Esse “ser capaz” se refere a certas circunstâncias em que necessariamente seremos motivados. (Voltarei mais de uma vez a como devemos entender essa noção). Vamos a um exemplo de Williams. *Owen* não tem qualquer motivação positiva para se alistar ao exército. Pelo contrário, ele odeia tudo sobre a vida militar e o que ela significa. Um internalista de razões, como Williams, diria que *se é impossível* a *Owen* ser motivado a se alistar no exército, então ele não tem uma razão para fazer isso. Se ele tivesse uma razão, ela o motivaria. Já um externalista diria que *independentemente* da motivação de *Owen*, ele pode ter uma razão para se alistar dado outras considerações, como morais e prudenciais. O internalista de razões considera a motivação relevante para que uma consideração seja uma razão para ação. O externalista não. O internalista, na versão de Williams, defende que a motivação é importante nas circunstâncias específicas mostradas anteriormente: na situação contrafactual de posse de todas as crenças verdadeiras e relevantes e de um certo

¹⁵ Setiya (2011, p.6) afirma que Williams, em “Internal Reasons and the Obscurity of Blame”, veio a admitir que há outras posições internalistas possíveis sobre razões para ação, entretanto, penso que ele duvida que a versão de Korsgaard esteja correta.

¹⁶ Essa distinção é metodologicamente justificada, pois ela permite um melhor mapeamento das posições de autores: alguns concordam com apenas uma das teses, outros com ambas e outros ainda com nenhuma delas, como já apontado inicialmente. Uma ligeira comparação da teoria de Williams com a de Korsgaard será feita na terceira seção do capítulo 2.

tipo de deliberação – no caso de Williams, a deliberação abrangente. Assim, um Internalista dirá que Owen tem razão para se alistar no exército se ele *puder* ser motivado nessa situação contrafactual. É preciso que seja *possível* que um agente seja motivado para uma certa ação na situação ideal para que ele tenha uma razão para fazê-la.

A segunda tese de Williams, uma *Versão da Teoria de Motivação Humeana (VTMH)*, diz que a motivação que surge através da deliberação abrangente e da correção de crenças depende do conjunto motivacional pré-existente do agente. Continuando a pensar no exemplo de Owen, o que essa tese diz é que: Owen virá a se motivar a se alistar no exército através da deliberação abrangente e da correção de crenças se ao fazê-lo, essa nova motivação puder *surgir de seu conjunto motivacional pré-existente*. A nova motivação em tais circunstâncias só surge a partir de uma motivação anterior. Se, então, o pai de Owen lembrá-lo que sua família tem uma tradição militar importante e Owen acredita que tradições familiares devem ser seguidas, ele poderia vir a ser motivado a se alistar ao deliberar corretamente na situação ideal. Como afirmei antes, a existência de uma possibilidade de motivação através da deliberação abrangente, partindo do do conjunto motivacional atual do agente quando suas crenças estão corrigidas é também chamada de *rota deliberativa sólida*. Isso é, uma rota que vai desde o conjunto motivacional existente do agente para a conclusão por deliberação abrangente quando plenamente informado¹⁷ de que se tem uma razão para fazer ϕ . Acreditar na VTMH é acreditar que o único modo de se criar motivação por deliberação é através de uma rota deliberativa sólida¹⁸. Como o agente passou a *acreditar* nesse fato, teremos aqui

¹⁷ Setiya, na verdade, considera desnecessária a *informação plena* das circunstâncias para essa tese (SETIYA, 2011, p.4). Como ele bem percebe, o argumento principal de Williams não depende disso. É por isso que na formulação dele dessa tese – ao contrário da minha – a noção de rota deliberativa sólida não exige a correção de crenças. Concordo com ele que isso é desnecessário para se chegar à conclusão que razões dependem do nosso conjunto motivacional, entretanto, isso permite facilitar a tese contraintuitiva para muitos de que podem não haver razões morais para certos agentes em certos casos. Quando enfatizamos as circunstâncias contrafactuais como fiz na seção anterior, além de melhor justificar a maior garantia de que elas nos dão para podemos encontrar razões (como tratarei depois ao abordar a restrição deliberativa), podemos melhor entender o quão mais improvável seria para que ocorresse o caso de que um agente não teria razões para ação. Por fim, o próprio Williams esclarece: “A afirmação de que alguém pode chegar na conclusão de que ele devia fazer ϕ (ou, a conclusão para ϕ) a partir de *uma rota deliberativa sólida envolve, em minha visão, pelo menos a correção de erros de fato e raciocínio* envolvidos na perspectiva do agente sobre o assunto”. (WILLIAMS, 1995, p.35, tradução minha, itálico meu). Há, na verdade, outras razões para exigirmos e explicitarmos essas circunstâncias e que ficarão mais claras ao longo deste capítulo – como o fato de a explicação de ações por razões ocorrer devido ao link psicológico da crença – e não de meros fatos no mundo.

¹⁸ Frisarei o ponto de que alguns termos recorrentes na literatura, como “rota deliberativa sólida”, ou

também uma ligação entre o mundo e o conjunto motivacional dele, estabelecendo um *link psicológico*. Essa crença vai alterar os elementos de seus S e assim indiretamente motivá-lo¹⁹. Outro modo de se referir a essa tese é dizer que a motivação surgida pelo raciocínio correto (abrangente) exige desejos (um conjunto motivacional possivelmente favorável a ϕ)²⁰.

Como vimos, Owen odeia não só o exército, mas tudo relacionado a ele. Assim, se mesmo à luz da informação plena e da deliberação abrangente, ele não deseja se alistar ao exército, um internalista que aceita a VTMH, dirá que Owen, *de fato*, não tem razão para isso. Não há rota deliberativa sólida para ele. Afinal, se para um agente ter razão para ϕ ele precisa ser capaz de ser motivado e se, previamente não tendo tal desejo (elemento de S), só lhe é possível criar um novo desejo de seu conjunto motivacional pré-existente (uma vez que esse é o único modo de poder ser motivado para a crença de que ϕ lhe é uma proposição de razão), então lhe é impossível ter razão para ϕ . Esse internalista, que une ambas as teses, defende a teoria de razões internas de Williams, que passarei a denominar de *Teoria de Razões de Williams (TRW)*. Grosso modo, o que Williams defende é o seguinte: se razões dependem de motivação (ter uma razão é poder ser motivado por ela na situação ideal) (IR), e se motivação que surge da deliberação abrangente e da correção de crenças depende de desejos (só surge deles) (VTMH), então razões dependem de desejos – dependem do nosso conjunto motivacional atual. Com maiores sutilezas, podemos formular essa teoria e suas teses como Setiya (2011, p.6) o faz:

Pr 1: Internalismo de razões (IR): o fato que p é uma razão para A fazer ϕ se A é capaz de ser motivado por ϕ ao creditar que p.

Pr 2: Ver. da Teoria de motivação humeana (VTMH): A é capaz de ser motivado a fazer ϕ pela crença que P apenas se há uma rota instrumental através da sua

“deliberação correta” tem interpretações ligeiramente diferentes em diferentes autores. Não obstante, estou usando o entendimento especificado neste parágrafo.

¹⁹ Novamente, creio que é apenas neste sentido que Williams quer dizer que as crenças podem fazer parte do conjunto motivacional dos agentes.

²⁰ E ao pensarmos nos termos de “rota deliberativa sólida” parece ficar mais claro que o tipo de raciocínio que Williams considera funciona de maneira instrumental. Afinal, com ela ocorre uma passagem da motivação anterior a partir da deliberação para a conclusão de que se tem uma razão e de que há uma nova motivação. Mas, novamente, essa noção será tratada no capítulo 2.

correção de crenças e da deliberação abrangente a partir do conjunto motivacional subjetivo do agente, junto com a crença que P, para o desejo de fazer ϕ .^{21 22}

Conc: Teoria de razões de Williams (TRW): o fato que p é uma razão para A fazer ϕ se há uma rota deliberativa sólida através da deliberação abrangente e da correção de crenças a partir do conjunto motivacional do agente, e da crença que P, para o desejo de fazer ϕ .

Como podemos ver em IR e VTMH, há também a menção ao termo “*crença*”, que não foi suficientemente abordado até agora. Explicá-lo exige explicar um outro pressuposto importantíssimo para Williams – talvez o mais importante. Podemos chamá-lo de *restrição explicativa (RE)* ou dimensão explanatória. Penso que podemos entendê-la como dizendo que: uma teoria de razões para ação precisa ser capaz de explicar a ação de um agente por razões²³(WILLIAMS, 1979, P.102). Se Owen vier a se alistar no exército e se Ana vier a tomar seu remédio, não queremos meramente dizer que “eles tinham uma razão para isso”, mas sendo proposições de razão para ação, queremos que elas expliquem o comportamento dos agentes ao agirem por razões. Quando pensamos que razões podem nos influenciar, também pensamos que nosso comportamento poderá ser explicado por elas – quando agimos de acordo com elas. Uma teoria de razões para ação deve ter uma tal capacidade explicativa. E para Williams é possível que essas teorias tenham essa dimensão. Porém, elas precisam ser internas. É preciso que a razão esteja *ligada* à motivação do agente. Afinal, como diz Williams, “nada pode explicar as ações (intencionais) de um agente exceto alguma coisa que o motive a agir assim” (WILLIAMS, 1979, p.107, tradução minha). Isto é, *é a motivação, no fim das contas, que pode explicar o comportamento de um agente*. Mas como exatamente a razão se liga à motivação? Quando falamos de uma razão, ela parece se tratar de um fato ou uma consideração

²¹ Não se deve deixar de notar que o conjunto motivacional de um agente, ou seus desejos, deve ser entendido como sua motivação atual (ou anterior à situação contrafactual) e não de alguma que ele possa vir a ter.

²² Ao contrário da premissa 1, a premissa 2 aqui tem uma formulação diferente daquela de Setiya. Como disse na nota 19, Setiya desconsidera a correção de crenças – ou talvez entenda-a dentro da noção de “rota deliberativa”. Devido a isso, a formulação dele é a de que “A é capaz de ser motivado para ϕ pela crença que p quando há uma rota instrumental *de suas crenças*, junto com o conjunto motivacional subjetivo e a crença que P, para o desejo para fazer ϕ ” (SETIYA, 2011, p.6, tradução minha, *ênfase* minha).

²³ “Se há razões para ação, deve ser o caso que pessoas às vezes agem por essas razões, e se elas o fazem, as suas razões devem aparecer em alguma explicação correta da ação” (WILLIAMS, 1979, tradução minha, p.102).

no mundo – como o de que um remédio pode salvar a vida de Ana. Mas como isso se relaciona com o agente?

Para responder isso, considere dois aspectos de proposições de razão. Por um lado, temos o *conteúdo da crença* que – se verdadeira – é um fato. Por outro lado, temos também a própria *atitude da crença*²⁴. Essa atitude não se refere a algo no mundo, mas a algo em nós. Assim, a crença oferece esse *link psicológico* entre o fato e a motivação devido a esses dois aspectos. E é a partir desse link que pode ser fornecida a explicação da ação de um agente por razões. Ou seja, esse *link é a crença*. Quando um agente tem uma razão para ação, ele acredita que ela é verdadeira e assim ele passa a ser motivado a agir assim. E é por ser motivado por essa razão – pelo conteúdo verdadeiro dela – que podemos explicar suas ações, dado que elas indiretamente afetam seu conjunto motivacional. Desse modo é possível melhorar as explicações dos exemplos acima. Owen poderia se alistar por *acreditar* que é verdadeira a proposição de que “é importante manter uma tradição de honra militar” e Ana poderia tomar o remédio caso ela *acreditasse* que, por exemplo “só vivendo por mais tempo é possível ver meus netos crescerem”. Tal crença causa indiretamente²⁵ um desejo, uma disposição para agir: aprovação, sentimento, reflexão emocional positiva, etc. Logo, uma crença do agente numa proposição de razão aponta um desejo do agente sobre ela, um conjunto motivacional favorável a essa consideração, e assim *explicamos* o porquê o agente agiu de tal modo. Logo, *acreditar é o que explica a aquisição de uma nova motivação por deliberação*. E é essa *motivação ligada à razão que explica o comportamento do agente por razões*. (E essa capacidade de satisfazer a restrição explicativa é também uma das maiores vantagens de teorias internalistas em relação às externalistas, como abordarei melhor no capítulo 2).

Como Setiya nota, além de dar conta de RE, *o link psicológico da crença também evita dois extremos* (SETIYA, 2011, p.5). Por um lado, evitamos que a teoria se torne mais implausível, por ser forte demais, ao retirarmos a possibilidade de um agente ser motivado a agir por razões que desconhece. (Já que se ele acredita na proposição de razão, sendo isso o que o levou a ser motivado, ele evidentemente a

²⁴ Essa distinção que fiz é baseada na distinção de Setiya entre “crenças cujo conteúdo são razões” e “crenças sobre razões”, que abordarei no próximo parágrafo. (SETIYA, 2011, p.5)

²⁵ “Indiretamente”, pois a crença não motivará por si mesma, mas afetará a motivação anterior do agente, seu S, e assim uma nova motivação (um desejo) poderá surgir.

conhece). Por outro lado, ele aponta que implausibilidade (por trivialidade) também é evitada, quando entendemos essa crença como uma crença cujo conteúdo é a *razão* e não meramente uma crença sobre razões. Assim, ele diz que devemos entender TRW como tratando apenas de crenças cujo conteúdo são razões (ibid.). *Discordo de Setiya* quando ele afirma que Williams trataria apenas desse segundo tipo e, na verdade, também penso que a distinção que usei antes é mais simples e consegue apontar as mesmas vantagens. Quando focamos na *atitude da crença*, também podemos dizer que o agente certamente conhecerá suas razões; já quando focamos no *conteúdo da crença (enquanto razão)*, várias supostas razões são eliminadas – e trivialidade é evitada.

Mas há algo interessante a ser dito sobre a distinção de Setiya? O que exatamente se quer dizer com “crenças sobre razões” e com “crenças cujo conteúdo são razões?”. Isso não é algo que ele especifica, mas creio que essas são duas formas que podemos entender “proposições de razão” ou “frases de razão”. É possível interpretar a ideia de uma proposição de razão tanto como “A tem uma razão para fazer ϕ ” (uma *crença sobre razões* que pode falhar em tratar de razões, não especificando o objeto) quanto como “apenas vivendo por mais tempo é possível ver meus netos crescerem” (uma *crença cujo conteúdo é uma razão*, sendo essa razão explícita). Um resultado disso é de que crenças sobre razões são *dependentes* de crenças cujo conteúdo são razões: a verdade de se Ana tem alguma razão para tomar seu remédio depende, entre outras coisas, da verdade de se é preciso viver por mais tempo para ela ver seus netos crescerem. Assim, Williams deve aceitar, na verdade, ambos os lados dessa distinção²⁶.

1.3) Interpretando a capacidade de ser motivado

Tanto VTMH quanto IR usa a expressão “ser capaz de ser motivado”²⁷. Mas podemos distinguir entre *dois sentidos diferentes de “ser capaz de ser motivado” – ou movido*²⁸. Mesmo que se diga que um agente está na situação ideal (com correção

²⁶ É interessante notar que Williams usa proposições de razão no sentido de crenças sobre razão principalmente no início de seu artigo de 1979, quando ele especifica seus objetivos e quando ainda lhe parece possível que seus argumentos a favor de TRW forneçam condições suficientes e necessárias para TRW (WILLIAMS, 1979, p.101). Já em “Internal Reasons and The Obscurity of Blame” (1995, p.35-36), podemos notar que aquela construção não só deixa de ser usada, como ele diz estar apenas tratando de que fornecem condições necessárias para TRW.

²⁷ Nesta seção estou seguindo e ligeiramente expandindo uma análise de Paakkunainen (2018, p.7-10) sobre aquela expressão.

²⁸ A diferença entre ser “motivado” e ser movido” será abordada na seção seguinte.

de crenças e deliberação sólida), então ele será capaz de concluir que tem uma razão para agir, podemos, na verdade, ainda assim entender essa situação contrafactual de dois modos. O primeiro é um que pensa em *mudar as circunstâncias contrafactuais* para que, em algum momento, cheguemos em circunstâncias nas quais aquela crença será verdadeira. Essa é a *interpretação de possibilidade*. Se existirem mundos possíveis em que o agente pode ser motivado por uma razão, então o agente é capaz de ser motivado. Se entendermos esses mundos possíveis como abrangendo também mundos possíveis distantes, em que incluímos mesmo situações de uma cirurgia cerebral²⁹ dos agentes (que muda a motivação do agente de modo a ele ser motivado a fazer algo, seja qual for a motivação anterior dele), teremos uma versão dessa *interpretação como uma de mera possibilidade*. Com ela, interpretaremos o “capaz de ser motivado” do Internalismo de Razões (IR), acima, como informando que o agente é capaz de ser motivado por uma razão se houver um mundo possível para tal. (E não há falta de opções para satisfazermos isso).

Porém, isso é bastante problemático. Para que a segunda tese, a VTMH, fale da mesma capacidade, é preciso que o antecedente dela, como IR, afirme que há um mundo possível em que o agente consiga ser motivado por certa consideração. O resto dessa última tese, entretanto, restringe quando o agente é capaz de ser motivado para que a sua motivação anterior importe – e restringe de modo a contradizer a primeira parte. Nesse caso, *essa segunda tese será simplesmente falsa*. O início da tese inclui exemplos como o de cirurgias cerebrais, que certamente não dependem da motivação anterior, ao passo que o consequente dela diz que mesmo esses casos dependem disso³⁰.

Mas há uma outra versão, mais exigente, para entendermos o “ser capaz de ser motivado” *ainda* como uma mera possibilidade. Podemos entender isso como se restringindo à possibilidade dos *mundos possíveis próximos, em que o conjunto motivacional do agente é como o atual*. Nesse caso, um outro problema – que na verdade também se aplica à versão anterior – ocorre: cometeremos *petição de princípio*. A primeira tese dirá que um agente é capaz de ser motivado a fazer certa ação se, em mundos possíveis em que ele é plenamente informado e delibera

²⁹ Paakkunainen (2018, p.7) se refere a este exemplo de David Sobel (2001;222-3).

³⁰ E como Paakkunainen nota, tentar usar o internalismo de razões como um argumento único (dado que a noção de capacidade de mera possibilidade já estaria ali) é usar ou um argumento inválido ou pelo menos, cometer petição de princípio.

abrangentemente, ele é levado a fazer essa ação por causa de uma certa consideração. Essa tese já parece, porém, pressupor a Teoria de Razões de Williams (TRW).

A outra leitura, mais promissora, entende o “ser capaz de ser motivado” da premissa do Internalismo de Razões como uma *capacidade motivacional do agente*. Essa capacidade motivacional seria a de poder ser motivado para fazer ϕ . Ela funciona do mesmo modo que podemos pensar outras habilidades permanentes do agente, tal como “nadar”. Eu posso não estar nadando agora, posso ter uma cãibra que me impede de nadar, posso estar longe de um corpo de água neste momento, mas nada disso muda o fato de que tenho essa habilidade: sou capaz de nadar. Para ter essa capacidade, basta ser verdade que eu *poderia exercer* essa capacidade se as circunstâncias fossem diferentes. Tudo o que é exigido é que existam mundos próximos ao atual nos quais eu consiga nadar e eu de fato o faça. Ou seja, basta existir a possibilidade nesses vários mundos próximos, de eu estar em frente a um corpo de água – juntamente com algumas outras circunstâncias, como ter a pretensão sincera de nadar – e eu conseguir fazer isso. E para isso, a minha capacidade é suficiente. Assim, de modo similar devemos entender o “ser capaz de ser motivado” como uma capacidade do agente: *nos mundos próximos em que tenho informação plena e delibero abrangentemente, serei motivado a agir. Minha capacidade será exercida*. “Ser capaz de ser motivado” é então apenas “ser capaz de ser motivado nos mundos próximos em que o agente está nessa situação contrafactual ideal”³¹. E com tal interpretação não haverá petição de princípio do caso anterior: o que se exige pelas premissas (uma capacidade motivacional) é diferente do que se quer provar (motivações e razões) – ao passo que antes, a possibilidade da motivação precisou ser restringida às mesmas circunstâncias que se quer provar.

1.4) razões pro-tanto vs. decisivas; e motivar a ϕ vs. levar a ϕ

Quando pensamos, como antes, na frase “A tem uma razão para ϕ ” (uma crença cujo conteúdo são razões), em oposição a algo específico, somos mais convidados a pensar *em qualquer razão* que o agente possa ter para fazer aquela ação. (Podendo ser até mais de uma). Isso toca numa distinção importante: a *distinção entre razões pro-tanto* (ou por certa consideração) e *razões decisivas* (“razões

³¹ Essa tese deverá ser melhor explorada por teorias motivacionais.

considerando-se-todas as coisas”³². Razões pro-tanto são considerações que contam a favor de algo apenas em alguma medida ou aspecto e cujo peso não é necessariamente determinante. Por exemplo, o fato de que minha avó está sozinha agora pode ser uma razão para visitá-la, mas o fato de que é quarta-feira também indica uma razão pro-tanto contrária – já que sei que esse é o dia que meu irmão sempre a visita. Já razões decisivas são, como o nome diz, categóricas – e únicas, as que aparentemente me levam à ação). Agora, se for segunda-feira e minha avó estiver sozinha, essas podem ser considerações que me levarão a decidir que o fato de que ela está sozinha é uma razão decisiva para eu visitá-la. Williams parece buscar tratar em sua teoria de ambos os casos de razões e não apenas de razões decisivas. Porém, Paakkunnainen pensa que podemos simplificar as pro-tanto em decisivas (PAAKKUNAINEN, 2018, p.7). Por causa disso, ela reconstrói³³ as teses de Williams em termos de *nos levar à ação* e não em termos de *nos motivar à ação*. O conflito dessa última distinção consiste no seguinte: como razões pro-tanto podem ser muitas para uma mesma ação, elas parecem apoiar a interpretação de *motivação à ação*, ao passo que razões decisivas (por serem únicas e assim gerarem uma única motivação) parecem apoiar a interpretação de *levar à ação*.

Vamos ver o que Paakkunnainen diz mais de perto. Ela começa por apontar que A) “[a] deliberação sólida certamente *não levaria alguém a agir* de acordo com *todas* as razões conflitantes em uma situação, nos modos (conflitantes) que elas apoiam” (ibid, *italico meu*, tradução minha). Ela diz que seria por isso até que Williams entende – como fiz acima – que o resultado da deliberação é de que *seremos motivados a agir* e não simplesmente agiremos (ibid). Porém, ela também aponta um outro problema: B) “é estranho esperar que a deliberação sólida apareça em múltiplas motivações conflitantes, sempre que há razões para ações conflitantes” (ibid, tradução minha). Devido a isso, ela faz a seguinte sugestão. Devemos pensar TRW enquanto tratando de razões conflitantes de algum outro modo – que *não nos motive à ação* – e quando TRW trata de razões decisivas, deveríamos pensar que o resultado da deliberação sólida *nos leva à ação* (ibid.). Por fim, ela afirma que C) entender a Teoria de Razões de Williams como tratando apenas de razões decisivas é uma assunção simplificadora que o próprio Williams usa às vezes (ibid). E, assim, seria preferível entender que a deliberação sólida nos leva a agir por razões decisivas.

³² Tradução literal minha de “all-things-considered-reasons”.

³³ A reconstrução de Paakkunnainen da teoria de Williams será abordada na próxima seção.

A nos aponta uma razão para entendermos as razões que vêm da deliberação sólida como *nos motivando* – seria ruim razões pro-tanto nos *levarem* a agir. B nos aponta que esse entendimento deve ser como o *de nos levando à ação* – já que a alternativa motivacional geraria motivações desnecessárias. Por fim, C resolve o conflito e apoia as sugestões de Paakkunainen: podemos entender razões decisivas como razões pro-tanto. E assim, podemos sempre pensar que *razões decisivas nos levam à ação*. Essa sugestão dela está correta?

Penso que B *não nos aponta um bom motivo* para pensarmos que o resultado da deliberação deverá ser a efetivação da ação. *Primeiro*, a noção de deliberação abrangente³⁴, apesar de oferecer diversos modos diferentes para concluirmos se temos uma razão, *não exige* a utilização desses vários raciocínios ou se quer mais de um. O mesmo pode ser dito da correção de crenças. *Tudo o que é exigido nos mundos possíveis próximos é que o agente tenha alguma rota deliberativa sólida* para concluir que ele tem alguma razão para fazer algo – e não que ela seja efetiva. Por exemplo, para determinar se tenho alguma razão (pro-tanto) para visitar minha avó, pode bastar que *em um mundo possível próximo*, eu saiba que ela está sozinha e que eu faça um tipo de raciocínio instrumental. Mas não é necessário que essa correção de crenças e deliberação abrangente seja atualmente efetivada. Do mesmo modo, não é necessário que a motivação pela razão seja atual para que ela seja uma razão para mim. Assim, o que noções de razão exigem são, novamente, certa capacidade motivacional *contrafactual*. *Segundo*, isso também implica que *podemos montar rotas deliberativas independentemente de outras*, de modo que cada uma *pode* nos motivar separadamente. No lugar de montar uma rota deliberativa a partir apenas do fato que minha avó está sozinha (uma crença minha que foi corrigida), indicando que devo visitá-la; também posso montar uma rota deliberativa indicando uma razão contrária: passo a saber que meu irmão está indo visitá-la, o que também me dá uma razão pro-tanto para visitá-la em outro dia.

Terceiro, penso que Paakkunainen toma B como problemático porque ela está imaginando um certo *procedimento diferente* do qual estou propondo. Imagino que ela está propondo que eu devo, em um mundo possível próximo, poder chegar à conclusão de que tenho uma razão *decisiva* para visitar minha avó – ou não. E quando chego nesta razão decisiva, isso implica que já considerei e pesei devidamente *todas*

³⁴ A Restrição Deliberativa (RD), o pressuposto que a deliberação abrangente se baseia, de modo similar só exige que agente seja capaz de descobrir suas razões.

as minhas razões pro-tanto – incluindo a de que meu irmão está indo visitá-la. Em outras palavras ela está pensando em uma *rota deliberativa específica*, uma rota deliberativa que *envolve* uma *rota deliberativa decisiva*. Porém, *uma rota deliberativa simples* (não sendo decisiva), como aquela considerando *apenas* o fato de que minha avó está sozinha – mas não que meu irmão está indo visitá-la – pode já ser suficiente para eu ter uma razão pro-tanto para visitá-la. Perceba que *interpretar* as deliberações daquele modo específico (como buscando razões decisivas) exige também interpretar a *correção de crenças* como tratando de *todas* as crenças verdadeiras e relevantes. (Ou seja, qualquer crença que minimamente influencie no julgamento do agente quanto a se ele tem certa razão é relevante). Mas como tentei apontar, esse não é o único modo: para encontrarmos *razões pro-tanto*, podem bastar as crenças *necessárias* para se concluir essa razão – como no exemplo da avó, em que meramente saber que ela está sozinha já indica que tenho uma razão para vê-la³⁵.

Quarto, podemos entender que o resultado da deliberação sólida ser a motivação à ação é compatível com os dois tipos de razões para ação. Quando estamos tratando de razões pro-tanto, a motivação é apenas em um pequeno grau, ao passo que quando estamos tratando de razões decisivas, a motivação que surge é uma em um grau tão grande que *pode*, nas circunstâncias ideais, de fato nos levar

³⁵ Não penso que a exclusão de rotas deliberativas simples é necessariamente problemática para outros internalistas. Afinal, poderíamos pensar que esse tipo de rota simples está injustificadamente excuída. Penso que esse é o caso para Williams (que permite esse tipo de rota), mas não para Korsgaard. Uma das intuições que temos com razões pro-tanto, é, por definição, que só parte dos aspectos relevantes são levados em conta. Uma *rota deliberativa simples* como a de que somente sei que minha avó está sozinha parece ser exatamente uma que nos levaria a ter uma razão pro-tanto para visitá-la. Nesse caso, interpretamos *todas as considerações relevantes* como apenas a de que ela está sozinha e um certo raciocínio instrumental, por exemplo, mas não outros, também seria necessário. Poderíamos pensar que isso tem consequências insolúveis para outras teorias internalistas concorrentes, como a de Korsgaard, que exige também certa compatibilidade com a moralidade. Como a teoria dela exige esse outro aspecto (mesmo em razões pro-tanto), ela seria impedida de exercer esse tipo de raciocínio mais simples – porque nele não se considera a moralidade. Porém, ela poderia responder que uma razão só é uma razão caso ela seja também compatível com a moralidade – ou seja, uma consideração moral é sempre relevante. Isso, entretanto, não mudaria o fato de que um agente para Korsgaard poderia concluir uma razão pro-tanto compatível com a moralidade (como a de que minha avó está sozinha ser uma razão pro-tanto para visitá-la) e ainda assim não saber que chegou numa conclusão de que é uma razão (porque as considerações morais, embora compatíveis, não foram usadas. A compatibilidade é necessária, mas não o uso de considerações e deliberações morais). Assim, essa razão parece ser compatível com a moralidade, mas o agente na situação contrafactual não sabe que é compatível. Porém, como apontei no primeiro ponto, Korsgaard precisaria afirmar nesse caso que razões exigem que exista em mundos possíveis próximos *alguma* rota deliberativa sólida para ele concluir que tem uma razão – assim, ela pode excluir aquela rota simples. Ademais, podemos pensar aquela *rota deliberativa simples* juntamente *com* considerações e deliberações visando a moralidade. Isso já será um caso de rota deliberativa sólida para aquela razão pro-tanto. Logo, apesar de ela ter de excluir esse caso de rota deliberativa simples (sempre exigindo um raciocínio puro, envolvendo considerações morais), isso não interfere em sua teoria.

à ação. *Quinto*, além de, como A aponta, ser problemático razões pro-tanto nos levarem à ação (uma vez que temos razões pro-tanto conflitantes), também razões decisivas várias vezes não nos levam à ação. Às vezes, temos razões para um certo curso de ação, mas que somos impedidos de fazer por outras circunstâncias – como o medo. Enfim, apesar de podermos simplificar razões pro-tanto em razões decisivas (como apontei no terceiro ponto), seria preferível ainda assim pensar que o Internalismo de Razões (IR) e a Versão da Teoria Humeana de Motivação (VTMH) *nos motivam* a agir.

Agora que boa parte dos detalhes interpretativos já foram tratados, podemos continuar a tentar reconstruir TRW no formato de um argumento de modo melhor. Para isso, vamos ver a reconstrução de Paakkunainen de TRW na próxima seção.

1.5) Comparando a reconstrução de TRW com a de Paakkunainen

Paakkunainen tem uma outra reconstrução do argumento de Williams (Paakkunainen, 2018, *passim*). Já que é ela é a responsável por aquela interpretação de capacidade – que opus a de possibilidade – da noção de “poder ser motivado”, sua reconstrução é voltada a isso. Ela é a seguinte:

(1C) O fato que p é uma razão normativa para A fazer φ apenas se A tem a capacidade atual de fazer φ porque P. (RE, versão de capacidade)

(2C) A tem a capacidade atual de fazer φ porque p apenas se A tem alguma motivação contingente em seu S atual que poderia levá-lo a fazer φ porque P através da deliberação abrangente instrumental sólida. (Aproximadamente Teoria de Motivação Humeana, versão de capacidade)

(3C) Então, o fato que p é uma razão normativa para A fazer φ apenas se A tem uma motivação contingente em seu S atual que podem levá-lo a fazer φ porque p através da deliberação abrangente instrumental sólida. (INT-W)

Um primeiro fato interessante da reconstrução de Paakkunainen é que ela usa o termo “capacidade atual” e não meramente “capacidade”. Com isso, ela quer dizer, como apontei antes, uma capacidade ligada a mundos possíveis próximos. Assim como um agente é capaz de nadar *agora* porque em mundos possíveis próximos ao atual ele tem as condições necessárias para isso (um corpo de água próximo, falta de estados físicos debilitantes e etc.) e de fato nada, o mesmo pode ser dito da

capacidade de um agente de fazer ações por razões. E como podemos perceber, essa leitura se torna mais clara com essa reconstrução da primeira premissa. Além disso, há ainda duas outras diferenças importantes entre ela e IR. Vamos ver IR novamente:

Pr 1: Internalismo de razões (IR): o fato que p é uma razão para A fazer ϕ se A é capaz de ser motivado por ϕ ao creditar que p .

A primeira diferença importante é a seguinte. Enquanto que (1C) restringe a razão à *capacidade do agente agir* de acordo com ela, IR restringe a razão à *capacidade do agente ser motivado* a fazer a ação correspondente. Como tratei na seção anterior, penso não haver desvantagens em entendermos que o resultado da deliberação é a de que o agente será motivado a agir – ao passo que a outra interpretação apresenta certos problemas.

A segunda diferença de IR com (1C) é a seguinte: Paakkunainen chama (1C) de *restrição explanatória* (RE) e não de internalismo de razões. Ele não explica *porque* exatamente ele faz isso, mas como a noção de RE é diferente da apresentada por Williams – que basicamente afirma que uma teoria de razões para ação deve ser capaz de explicar o comportamento de um agente por razões – parece em um primeiro momento preferível entender a primeira premissa como o Internalismo de Razões e RE como uma razão que o apoia. Creio que a razão para ele nomear a primeira premissa, entretanto, se baseia em seu *argumento que liga RE com RD* (a restrição deliberativa). (Como disse na primeira seção, RD é a suposição de que um agente é capaz de descobrir suas razões através da deliberação).

Esse argumento de Paakkunainen³⁶ aponta que versões plausíveis do Internalismo de Razões que aceitam RE também devem aceitar RD – como é o caso de Williams. Não queremos explicar apenas casos em que agentes são diretamente motivados a fazer algo, mas também outros, como por exemplo casos instrumentais pelos quais um agente faz ψ como um modo de fazer ϕ . E o único modo de pensar na capacidade dos agentes para além desse caso simples de motivação direta (em que o agente é motivado a fazer ϕ como uma solução constitutiva para seu desejo de ϕ), é através da deliberação. Assim, quem defende

³⁶PAKKUNAINEN, 2018, p.12.

(1C) deve defender também uma versão da Restrição Deliberativa³⁷. *Entretanto*, dado que a principal distinção de teoria de razões para ação é a de razões internas e externas, e não a de teorias com ou sem a Restrição Explanatória – o que é salientado por existir a possibilidade de se defender teses internalistas sem se defender RE –, então penso ser melhor continuar chamando (1C), ou suas outras versões (como IR), de *internalismo de razões*. Depois disso, o fato de que teoria X é justificada ou não por RE pode ser apontada e assim a relação de RE com RD. (E isso já nos mostra que será importante ver argumentos a favor de RE e RD, já que o internalismo de Williams é dependente deles).

E quanto à segunda premissa da reconstrução de Paakkunainen, (2C)? Quais as diferenças entre ela e VTMH? Vamos compará-las:

Ver. da Teoria de motivação humeana (VTMH): A é capaz de ser motivado a fazer ϕ pela crença que P apenas se há uma rota instrumental através da sua correção de crenças e da deliberação abrangente a partir do conjunto motivacional subjetivo do agente, junto com a crença que P, para o desejo de fazer ϕ .

(2C) A tem a capacidade atual de fazer ϕ porque p apenas se A tem alguma motivação contingente em seu S atual que poderia levá-lo a fazer ϕ porque P através da deliberação abrangente instrumental sólida. (Aproximadamente Teoria de Motivação Humeana, versão de capacidade)

As diferenças delas se tratam do seguinte. A premissa (2C), assim como a formulação anterior de Setiya, ignora a correção de crenças – a versão VTMH, que não faz isso, é na verdade uma versão levemente alterada da premissa de Setiya, como informei na nota 20. Como apontei antes, Williams (1995, p. 35) afirma que a correção de crenças faz parte da rota deliberativa sólida. No caso de Paakkunainen, entretanto, ao contrário do de Setiya, a necessidade da verdade das crenças relevantes está pressuposta em (2C). Além disso, uma outra diferença consiste na utilização do termo “através” no lugar de apontar a existência em mundos possíveis de uma rota deliberativa sólida. Em ambos os casos (na explicitação da correção de crenças e na utilização, como Paakkunainen, de “através”), essa diferença serve

³⁷ A relevância desse argumento neste ponto é somente para a reconstrução da tese de Williams de modo que possamos melhor comprar TRW com outras teorias internalistas. Porém, ele certamente já serve como objeção a teorias que, ao contrário da de Williams, aceitem RE sem aceitar RD.

para um mesmo propósito³⁸ (apesar de no segundo caso a autora não ter esclarecido isso): tornar mais claro o modo pelo qual o agente na situação contrafactual de outros mundos possíveis se trata dele mesmo e não de uma outra versão dele.

O que se pretende afastar aqui é uma confusão entre o mesmo agente, embora em outros mundos possíveis próximos, e uma outra versão “melhorada” dele e que não pode sê-lo, mas que por ser perfeitamente informado e racional pode *aconselhá-lo*. Essa última versão é uma versão contrafactual – e que não envolve a noção de capacidade – do agente segundo o modelo de conselheiro. (SMITH, 1995). A diferença entre o agente de Williams e esse conselheiro consiste principalmente no seguinte. Quando pensamos o agente no caso de Williams, deve haver uma rota deliberativa para ele *passar a ser motivado*, há um certo processo que o agente, mesmo em outros mundos possíveis próximos, *precisa passar* para saber se ele tem ou não razão para fazer algo. Já o agente do modelo de conselheiro de Smith é “mais” do que o agente nesse mundo possível próximo: ele não precisa passar por qualquer processo, pois já está completamente informado e motivado. Ele é perfeitamente – ou, como diz a literatura, “plenamente” – informado. Tudo o que lhe resta é informar ao agente atual o que ele tem ou não razão para fazer. Parece que ao não enfatizarmos a noção de *correção de crenças* e a ideia de *processo do agente* a partir de palavras como “através”, podemos mais facilmente pensar em termos dessa outra teoria – o que seria um erro.

Essa ênfase na noção de correção de crenças também nos leva a outra questão: o processo pelo qual podemos dizer que o agente nesses mundos possíveis é capaz de ser motivado consiste 1) apenas de deliberação e através dela suas crenças serão corrigidas ou 2) consiste, de modo separado, em deliberação e correção de crenças (dois processos diferentes)?³⁹ (Essa é muito próxima da distinção de Paakkunainen entre a capacidade do agente, *dado fatos*

³⁸ Na verdade, isso também enfatiza a correção de crenças que, como apontado no terceiro ponto da seção anterior, também permite melhor percebermos que existem diferentes rotas deliberativas possíveis – das mais simples às decisivas.

³⁹ O fato de que Williams afirma que o processo deliberativo é obscuro – e isso seria até mesmo uma vantagem de sua teoria –, como citei na primeira seção, nos desencoraja a especificar isso (WILLIAMS, 1979, p.110). Afinal, não há limites fixos no contínuo do pensamento racional para a inspiração e a conversão (ibid). Porém, creio que clarificar certos aspectos – não fornecendo uma fórmula – pode ser frutífero. Podemos também meramente pensar em casos de deliberação mais claros.

relevantes fornecedores de razão e a capacidade do agente de *passar* a acreditar em fatos relevantes fornecedores de razão)⁴⁰. Se consistir em 2, então seria preferível entender, como venho fazendo, a deliberação sólida e a correção de crenças de modo separado em VTMH. É isso o que Paakkunainen defende – uma vez que ela defende também “b”⁴¹. E, de fato, o processo não pode consistir em 1. Se ele consistisse em “1”, o agente do mundo possível próximo deveria realizar uma série de deliberações que sempre o permitiria adquirir as informações necessárias – a deliberação sólida também corrigiria as nossas crenças relevantes. Nesse caso, crenças relevantes de mais difícil acesso, como relacionadas aos problemas de saúde advindos do fumo antes de 1700, seriam misteriosamente descobertas pela deliberação do agente. Assim, é preferível *não pressupor* a correção de crenças na segunda premissa de Williams – ou omiti-la.

Mas além das premissas (1C) e (2C), Paakkunainen (2018, p.12-13) considera uma outra premissa e uma certa correção de (2C). Essa outra premissa e (2C) corrigido se trata do seguinte:

(2C*) A tem a capacidade atual de fazer ϕ porque p apenas se A tem alguma motivação contingente em seu S atual que poderia levá-lo a fazer ϕ porque P através da deliberação abrangente instrumental, seja sólida ou não. (Aproximadamente Teoria de Motivação Humeana, versão de capacidade revisada)

(2C)** Se p não é apenas a razão motivacional de A para fazer ϕ , mas também uma razão normativa para fazer ϕ , então a capacidade atual presente de fazer ϕ porque p precisa ser uma capacidade não apenas de deliberar a partir do S de A para

⁴⁰ Nas palavras de Paakkunainen (2018, p.17, tradução minha): “ (a) a capacidade de A, dado crenças em fatos fornecedores de razões relevantes, levá-los em consideração na deliberação e ser movido por eles para fazer o que eles apoiam, e (b) a capacidade de A de passar a acreditar nos fatos fornecedores de razão relevantes”. Tanto “a” quanto “2” *separa o processo* que o agente contrafactual deve ser capaz de passar em se tratando apenas de um processo deliberativo, na medida em que a correção de suas crenças já é algo pressuposto do agente nesse mundo possível. Enquanto isso, “b” e “1” imaginam a correção de crenças como uma capacidade que o agente pode atingir nesse mundo possível próximo. Penso ser preferível, na verdade, uma *terceira possibilidade*: que seja possível que os processos sejam independentes, mas não necessariamente. Desse modo, poderíamos tanto pressupor mundos em que o agente já tem certas crenças (“a”), quando falamos de crenças impossíveis ou de acesso muito difícil, quanto poderíamos ainda permitir que o próprio agente descubra informações novas em situações mais simples (“b”). (E, desse modo, a opção seria 3: o processo pode ser independente, mas não necessariamente, o que já exigiria uma maior ênfase na correção de crenças – ela teria de ser explicitada na tese). Mas essa é uma posição que não tenho espaço para defender aqui e, de todo modo, o ponto é que mesmo a interpretação preferida de Paakkunainen, “a”, já pressupõe aqueles dois processos como independentes – e assim que a correção de crenças deverá ser enfatizada. Tudo o que é necessário é rejeitar “b” e “1”.

⁴¹ Ver nota anterior.

fazer ϕ porque p , mas uma capacidade de deliberar solidamente de seu S para fazer ϕ porque p . (Isso é assim porque a Restrição Deliberativa é verdadeira.)

A diferença entre (2C) e sua versão melhorada, (2C*), é o fato de esta última abrir a possibilidade de o agente agir por um fato através de uma deliberação não sólida. E é isso o que torna (2C) falso: não é verdade que um agente só pode agir de acordo com um fato que o motive através da deliberação sólida. Apesar de eu defender que a deliberação abrangente deve nos motivar a agir, penso que essa objeção de Paakkunainen se mantém. Posso ser motivado a agir – e até ser levado a agir – de acordo com um fato também quando erro em minha deliberação⁴². Um argumento similar pode ser feito contra a noção de correção de crenças em VTMH. Por isso, ela deve sair dali. E assim, (2C**) será exigido para que o raciocínio seja garantido como *sólido – como Paakkunainen aponta –, mas também como uma deliberação que depende de crenças verdadeiras e relevantes*. Um último fato interessante dessa última tese, (2C**), é que ela nos aponta que p deve ser também uma razão motivacional – fato que abordarei no próximo capítulo. Devido a todas as considerações acima, a reconstrução do argumento de Williams será a seguinte:

Pr 1: Internalismo de razões (IR): o fato que p é uma razão para A fazer ϕ se A é atualmente capaz de ser motivado por ϕ ao creditar que p .

Pr 2: VTMH': A tem a capacidade atual de fazer ϕ porque p apenas se A tem alguma motivação contingente em seu S atual que poderia motivá-lo a fazer ϕ porque P a partir da deliberação abrangente instrumental. (aproximadamente Teoria de Motivação Humeana, versão de capacidade revisada sem correção de crenças)

Pr 3: VTMH'': Se p não é apenas a razão motivacional de A para fazer ϕ , mas também uma razão normativa para fazer ϕ , então a capacidade atual presente de A fazer ϕ porque p precisa ser uma capacidade não apenas de deliberar a partir do S de A para ser motivado a fazer ϕ porque p , mas uma capacidade de deliberar *solidamente e com a correção de crenças* de seu S para ser motivado a fazer ϕ porque

⁴² Neste ponto Paakkunainen (2018, p.12) usa o exemplo de Williams de um agente que não tem uma razão para tomar um copo de gasolina que ele erroneamente pensa ser de gim e tônica. Considero esse um erro, já que esse exemplo aponta que as crenças devem ser corrigidas e a consideração, que é uma razão, deve se tratar de um fato. Enquanto isso, a razão que Paakkunainen tem para usar (2C*) é puramente quanto a um erro de raciocínio.

p. (Isso é assim porque a Restrição Deliberativa é verdadeira).⁴³

Conc: TRW': Assim, o fato que p é uma razão normativa para A fazer ϕ apenas se A tem alguma motivação contingente em seu S atual que o torna capaz de ser motivado a fazer ϕ porque p através da deliberação instrumental⁴⁴ sólida e da correção de crenças.

Capítulo 2: análise dos argumentos de Williams a favor de sua teoria

2.1) Restrição Explanatória (RE) e interpretação de “capacidade explicativa”

Pretendo, agora, fazer uma avaliação crítica da teoria de razões para ação de Bernard Williams. Para fazer isso, começarei *reconstruindo o principal argumento de Williams a favor do Internalismo de Razões*⁴⁵. Esse argumento parte de uma tese que já abordei antes e apontei ser uma razão para o internalismo de razões. Essa tese supõe a necessidade de uma teoria de razões para ação precisar explicar a ação do agente por razões – a dimensão explanatória ou restrição explanatória (RE). Além dela, é defendido a tese de que a única forma de explicar a ação de um agente por razões é através de sua motivação (UE) – se ele é *capaz* de ser motivado pela razão.

I) Uma teoria de razões para ação precisa ser capaz de explicar a ação de um agente por razões. (*Restrição Explicativa, RE*)

II) A única forma de explicar a ação de um agente por razões é através de sua motivação – se ele é *capaz* de ser motivado pela razão. (Única ou melhor possibilidade de explicação, *UE*)

III) *Assim*, concluímos o Internalismo de Razões (que diz que razões normativas dependem de motivação).

Em “Internal and External Reasons” (1979), Williams acreditava que com essas duas premissas poderíamos concluir o seu internalismo: TRW'. Entretanto, daquelas duas premissas (RE e UE) só podemos concluir o Internalismo de Razões

⁴³ Chamarei VTMH' e VTMH", quando me referindo às duas, de VTMHs.

⁴⁴ A noção de *instrumentalidade* ainda não foi abordada, mas será vista no próximo capítulo.

⁴⁵ Como diz Paakkunainen, há controvérsia em saber se esse é o único argumento de Williams ou o melhor e até como interpretar o internalismo de Williams. (PAAKKUNAINEN, 2018, p.2)

e não alguma versão da Teoria de Motivação Humeana. Afinal, do fato de que uma teoria de razões precisa poder explicar a ação de um agente por razões e de que a motivação é o único modo de explicá-la, se segue que essa teoria deve ser interna, que razões podem ser capazes de motivar os agentes para os quais elas são razões, garantindo o Internalismo de Razões⁴⁶. Mas disso não se segue que a nova motivação que surge através da deliberação sólida precise estar baseada nos desejos pré-existentes do agente (a segunda e terceira tese, as VTMHs). É possível, por exemplo, que a razão sozinha ou até uma disposição crie a nova motivação – falsificando a premissa dois. (Isso já aponta para a *possibilidade* de outras teorias motivacionais envolvidas, que estariam presentes em outras teorias de razões para ação). E assim, concluímos com aquelas premissas apenas o Internalismo de Razões (IR).

Quanto a esse argumento, ainda é preciso *qualificar* o que significa “*ser capaz de explicar*”. Setiya distinguiu três interpretações da dimensão explanatória – implicando *três possibilidades de leitura* para aquela frase⁴⁷ (SETIYA, 2011, p.10). A *primeira é uma interpretação mínima*, ou fraca, pois entende a capacidade de explicação do comportamento de um agente por razões somente por apelo à categoria metafísica que ela pressupõe ser correta: os fatos. Já a segunda e a terceira entendem a capacidade de explicação por apelo à motivação. Isto é, como dito anteriormente, razões são capazes de explicar o comportamento de um agente na medida em que elas provocam, a partir da crença, sua motivação para ação – pois é a motivação, em última medida, que pode explicar o comportamento do agente. E essas duas últimas leituras se diferenciam na medida em que para a *segunda* um fato ou consideração será uma razão desde que seja *humanamente possível ser motivado por ela*, ao passo que para a *terceira*, uma razão não será uma razão para todos, mas apenas *para agentes particulares* – para os agentes que podem ser motivados por essas considerações ou fatos. Portanto, a segunda é uma interpretação geral (dado que razões são motivações para algum agente humanamente possível), enquanto que a terceira e última (a de Williams) é uma leitura particularista, pois as razões

⁴⁶ Novamente: devemos entender o “poder motivar” pelo menos no seguinte sentido: a capacidade contrafactual das razões de motivar agentes para os quais elas são razões.

⁴⁷ E cada uma dessas leituras também vai implicar em diferentes teorias de razão para ação, dado que elas vêem diferentemente as relações entre razões, motivação e explicação. Setiya na verdade não explora em detalhes essas possibilidades de leitura ou suas respostas e por isso tomarei a liberdade de colocar um nome em cada uma dessas interpretações diferentes.

dependem do agente para serem razões. Setiya aponta que com as duas primeiras leituras seria possível termos uma teoria de razões para ação que explique a ação dos agentes e que não implique o Internalismo de Razões (IR). Apesar de dizer isso, esse autor não cita isso como uma objeção a teoria de Williams. Entretanto, não vejo como esse fato, se for verdadeiro, não diluirá a força dos argumentos de Williams – em especial o da dimensão explanatória. Assim, avaliarei a plausibilidade dessas outras interpretações da dimensão explanatória em comparação com a de Williams.

A *primeira interpretação, a mínima*, diz que razões para agir nos fornecem a explicação somente no sentido de fundamentação da ação por fatos. Por exemplo, supondo que eu tenha que fazer uma prova amanhã, esse fato justificaria a minha ação de estudar para a prova que farei no dia seguinte. Podemos formulá-la na proposição "A fez ϕ porque F" – sendo "F" um fato que pressupomos. Esse parece até, *prima facie*, o modo como explicamos o comportamento de outras pessoas. Nesse sentido fraco ou mínimo de "explicar", basta para as proposições de razão indicarem esse tipo correto de categoria metafísica para serem razões normativas. Isto é, basta que as razões sejam fatos⁴⁸. Uma das vantagens em uma explicação nesse sentido é a de que por razões serem fatos, não há o comportamento de um agente baseado em falsidades⁴⁹. Entretanto, o poder explicativo dessa interpretação é tão reduzido que ela se torna bastante preterível. Apontarei *dois argumentos* que creio serem suficientes para *desconsiderarmos essa interpretação* como plausível.

O *primeiro* é o seguinte: esse tipo de explicação parece, na verdade, a que usamos em outros contextos, os contextos que não possuem agência. Por exemplo, ao dizer que "o teto (A) caiu ('ação' X) porque as paredes não eram fortes o suficiente (fato que explica)". Esse tipo de explicação não seria adequado para razões para ação, pois seria um tipo de explicação que usamos em contextos não intencionais – sendo assim um erro categorial. Afinal, tal explicação permite explicações meramente causais entre um fato e uma ação, sem que haja um vínculo intencional⁵⁰. E é

⁴⁸ Há quase consenso na literatura que razões para ação devem ser fatos ou proposições. Mas qual delas ainda é um problema. E desse modo, estou usando "fatos" e "considerações" de modo intercambiável, exceto quando estou me referindo a essa primeira interpretação.

⁴⁹ Note que isso é diferente da segunda circunstância de correção de crenças. Naquela, as crenças do agente precisam ser moldadas aos fatos – relevantes – para que o agente possa melhor deliberar, ao passo que aqui a deliberação – e assim RD – se quer é necessária.

⁵⁰ Creio que esse argumento não seria aceito por um 'naturalista hardcore'. Ele pode pensar que a explicação causal das paredes (sobre objetos inanimados) deve ser a mesma que usamos em contextos de agência. Isto é, um 'naturalista hardcore' pode objetar ao primeiro argumento ao dizer

exatamente a falta de um vínculo entre os fatos e o agente que nos aponta para um *segundo argumento*. A falta do *link psicológico* torna essa uma *explicação meramente parcial* do comportamento do agente. Afinal, com essa explicação por fatos (supondo que razões para ação sejam desse tipo ontológico) *não se tem a explicação* de o porquê eu passei a estudar ou porque as ações de Owen ou Ana foram as que eles tomaram. Essa leitura supõe a ação do agente de acordo com um fato, mas não há uma justificativa para se pensar assim, já que esses *fatos podem estar no mundo sem que o agente sequer saiba ou acredite neles*. (E era isso que Williams pretendia evitar ao dizer que o agente, na situação contrafactual, deveria estar sob a posse dos fatos relevantes).

Em suma, os fatos que explicam as ações não possuem o *link psicológico* que os permitiriam poder influenciar em nossas ações. Retomando o exemplo anterior, o fato de que tenho prova amanhã pode não explicar minha ação. Por exemplo, eu poderia estar estudando para a prova porque penso que ela ocorrerá na outra semana e assim estou tentando me adiantar. Ou ainda, posso estar estudando para a prova independentemente do fato de que ela ocorrerá amanhã: estudo porque tenho uma compulsão a isso. Assim, está permanentemente aberta a possibilidade de exemplos em que os fatos não causam ações – pois tudo que é necessário é que os fatos sejam fatos: verdadeiros. Além desses dois argumentos, devemos nos lembrar do argumento de Paakkunainen sobre a relação entre RE e RD: se o agente não puder ser capaz de deliberar para a conclusão de que tem uma razão, então tudo o que se pode explicar são casos de soluções constitutivas – em que o agente é diretamente motivado a fazer algo. Mesmo se um modo de responder aos dois primeiros argumentos for encontrado⁵¹, não poderemos explicar, por exemplo, quando um agente faz ψ como um modo de fazer ϕ .

A *segunda interpretação* é a *da humanidade ou geral*. Ela diz que se uma consideração é uma razão para uma ação, *alguma* pessoa pode ser motivada⁵² a

que uma explicação por fatos não seria um erro categorial. Mas creio que ele teria que admitir o segundo argumento que apresento a seguir.

⁵¹ O que creio ser extremamente implausível no segundo argumento, já que pensamos nossas ações como pelo menos parcialmente um resultado de nossos estados internos, exigindo assim o link psicológico. Mesmo que seja verdade que o fato de que a prova ocorrerá amanhã causou a minha ação de estudar, esse fato só provocou esse comportamento em mim uma vez que eu estava de posse dele.

⁵² Quando Setiya formula essa interpretação ele usa o termo “movido”, indicando que a razão deve levar o agente à ação e não meramente motivá-lo (SETIYA, 2011, p. 10). Isso já foi discutido no capítulo anterior.

fazer algo por *acreditar* em tal consideração⁵³. Com a adição do termo “acreditar”, temos finalmente o link psicológico entre esse fato ou consideração e a motivação do agente. Assim, *parece* que podemos dizer que o comportamento do agente pode ser explicado totalmente – como dito antes, é por acreditar em uma proposição de razão que Owen e Ana se motivarão a fazer certa ação e poderemos explicar suas ações. Porém, isso não implica o Internalismo de razões. Por “*alguma pessoa*”, quer se dizer que algum ser humano, em algum estado, precisa ser capaz de ser movido ou motivado pela consideração que é uma razão. Deve ser *humanamente possível* de sermos motivados por uma certa consideração para que ela seja uma razão. O internalismo de Razões não defende que *algum* agente ou outro pode ser motivado, mas que o agente para o qual a consideração é uma razão poderá ser motivado por ela.

Em contraste, a referência na tese de Williams é a um *agente particular*, a pessoa para a qual uma consideração é uma razão é a pessoa *capaz* de ser motivada por ela. Assim, a *terceira interpretação, a particular (e que Williams defende)*, diz que se um fato é uma razão para ação, o agente para o qual isso é um fato *pode ser motivado* por ele – e novamente, há um vínculo psicológico, pois é um fato para ele, ele acredita em tal fato.

Temos mais razão para defender uma interpretação do que outra? Apresentarei *um argumento* para dizer que a interpretação de Williams é *preferível*. Ele é tão breve que devo reconstruí-lo ainda neste mesmo parágrafo: embora com a segunda interpretação exista um vínculo entre um fato e a motivação pela crença, tal *vínculo erra o alvo (!)*. Não queremos saber se um frade dominicano do século XV tem a motivação que explica a minha ação de estudar para a prova amanhã ou se um importante líder do PCUS tem a motivação de não tomar os remédios de Ana logo após 1990. Queremos saber se *Eu* e se *Ana* temos essas razões e assim *podemos* ser motivados a fazê-las. E *quando* agimos por elas, *teremos uma explicação* para isso. O fato de ser humanamente possível que alguém outro seja motivado por uma consideração explicará a ação desse alguém outro, não a de quem nos referimos. Assim, a referência precisa ser particular, tal como Williams

⁵³ Ao contrário da interpretação anterior, agora não é necessário ver razões como fatos – embora Setiya pareça o fazer (loc. cit.).

prevê.

Logo, essa *terceira interpretação é preferível*. Somente com ela podemos bem explicar quando um agente age por razões. Para Williams, essa explicação consiste em um estado conativo do S do agente causar a ação por razões. A consideração que é uma razão – junto com a correção de crenças e a deliberação abrangente – foi a responsável por trazer esse estado de coisas. Porém, apenas IR pode ser concluído da Restrição Explanatória e não VTMH' e VTMH''. O fato é que basta para a razão ser ligada por um link psicológico (a crença) para que possamos explicar o comportamento do agente quando ele age por razões. (E para que um maior potencial explanatório seja preservado, é preciso que o agente seja capaz de descobrir essa razão por deliberação – e ainda melhor se for sólida). Mas disso não se segue que a motivação que surge *precisa* surgir de uma deliberação abrangente, que funciona de modo instrumental – o que implica que ela dependeria da motivação anterior. Porém, isso poderia se seguir se esse fosse o único ou o melhor modo de explicar a motivação dos agentes (como UE). O único modo certamente não é: há várias outras teorias que explicam diferentemente como o agente passa a estar motivado ao deliberar solidamente. Podemos ter outras teorias motivacionais, como veremos, como a de Korsgaard que diz que a motivação que surge da deliberação sólida não depende da motivação anterior – a razão por si só pode causá-la.

2.2) Vantagem de TRW': relação simples entre razões

Uma *vantagem explicativa* pode ser apontada *quanto às relações de razões normativas, explanatórias e motivacionais*⁵⁴. Como razões normativas para Williams dependem dos desejos do agente, ele as identifica com razões motivacionais. Fazendo isso, na verdade, também estamos explicando a relação entre *razões normativas e razões explanatórias*. Williams, além disso, diz que razões normativas são potencias razões explanatórias. Afinal, como apontado anteriormente, se dissermos que Owen tem uma razão para se alistar no exército devido ao fato de que é importante seguir as tradições militares de sua família, então isso será o mesmo que dizer que ele *acredita* que é importante seguir as tradições de sua

⁵⁴ Respectivamente, uma razão que conta a favor de se fazer algo (uma razão para agir), uma razão pela qual alguém age de certo modo, e uma razão que explica a ação do agente – sem necessariamente justificá-la (ALVAREZ, 2017).

família. Essa segunda proposição é uma razão explanatória. Não é o mero fato de que é importante, mas de que *é importante para ele*. Por esse ser o caso, teremos uma explicação caso ele de fato venha a se alistar. Aquela proposição consegue explicar na medida em que foi criado um vínculo entre razões normativas e o conjunto motivacional do agente – entre fatos e a motivação do agente – já apontado anteriormente como um link psicológico: a crença. Essa crença (na proposição de razão) *causa* uma nova motivação no agente e assim seu comportamento é explicado.

Como podemos perceber, a motivação que surge é, também uma *razão motivacional* (como expresso por VTMH’). Se o agente agir por uma razão, essa será a razão pela qual ele assim o faz. E a razão normativa pode ser entendida como essas outras duas se formulada de modo correto. Assim, Williams não diz apenas que Razões normativas são potenciais razões explanatórias, mas ele *explica a relação entre cada uma dessas razões por uma mesma teoria*. Logo, pode-se apresentar como uma vantagem da teoria de Williams a sua capacidade de explicar a relação entre todas essas razões, principalmente de um modo tão simples. Porém, há diversas outras interpretações sobre a relação entre razões normativas e explicativas – e possivelmente também com as motivacionais⁵⁵. Por outro lado, essa, novamente, é uma das poucas explicações que conseguem explicar todos esses fenômenos – as relações entre esses diferentes tipos de razões.

2.3) Argumento do ônus de prova e comparação com Korsgaard

A outra justificativa que Williams fornece para a sua teoria – e assim o seu segundo principal argumento – é uma que pode também fornecer uma defesa das VTMHs (VTMH’ e VTMH’). (E é um argumento que, na verdade, depende do argumento anterior – um argumento por melhor explicação que usa UE (Única ou melhor explicação) e RE (Restrição Explicativa). Creio que esse segundo argumento se trata de uma resposta a uma possível objeção apresentada por um externalista de razões. Podemos resumi-lo do seguinte modo:

⁵⁵ Citarei alguns exemplos de Setiya (2011, p.11): 1) Razões precisam ser normativas, exceto quando falsas; ou que isso forma um "ideal regulativo" pela qual razões explanatórias se aproximam; 2) Razões para ação são considerações que tomamos como verdadeiras para o que estamos fazendo. (Essa é uma posição compartilhada por muitos e que parece refletir a interpretação fraca da dimensão explanatória); 3) Razões normativas ou "boas" podem ser pensadas como a fundamentação pela qual seria bom agir.

OP) A teoria contrária não dá conta da dimensão explanatória e se tentar terá um *ônus de prova*: o que além da *motivação* pode motivar ou até mover o agente de modo a explicar sua ação?

Penso que esse *argumento funciona de duas formas: I) como um ataque a posições externalistas e II) como um ataque (dependente de uma hipótese) a posições contrárias a algo próximo de uma Teoria Humeana de Motivação*. (Devemos dividi-lo assim, novamente, pois não estava tão claro para Williams que sua teoria dependia de algum tipo de teoria de motivação humeana). Um externalista poderia afirmar que uma proposição externa é verdadeira e que ela dá conta da RE⁵⁶. Nesse caso, esse externalista estaria afirmando que um agente poderia deliberar solidamente e, não importando sua motivação, ele chegaria à conclusão de que tem uma razão para fazer ϕ . Já Williams poderia afirmar que, quando esse agente *passa a acreditar* que possui uma razão para ϕ , ele também passa a estar motivado a fazer essa ação – sua crença de algum modo (como afetando o S do agente) produz uma nova motivação e assim poderemos explicar seu comportamento por razões. A crença serviu como um link psicológico entre o fato e sua nova motivação. Não é isso o que acontece com o externalista. Crenças cujo conteúdo são razões para o agente não precisam motivá-lo, diria o externalista. Mas de que modo podemos entender essa proposição de razão externa para que ela não dependa da motivação do agente, de algum modo advinda da razão, mas ainda assim possamos explicar sua ação por razões? Que sentido, preservando RE, proposições de razão podem ter? O que exatamente é implicado ao acreditar nela?⁵⁷ De onde ou como poderia surgir essa motivação do agente? Isso é algo *realmente misterioso*. *Esse é o ônus de prova que Williams aponta ao externalista*. Assim, proposições externalistas não conseguem dar conta de proposições de razão com sentido e explicar a restrição explanatória. O único modo de entendermos uma proposição de razão com sentido (UE), explicando também as ações do agente (RE), é se essa proposição for interna. Isto é, se a razão puder gerar, de algum modo (o que para Williams é a partir da influência que a crença causa no S anterior

⁵⁶ Para esse argumento se supõe a interpretação três, a particularista, da “capacidade de explicar”.

⁵⁷ Essa pergunta é baseada numa das formulações de Williams centrais desse argumento é a seguinte: “O que é que o agente passa a acreditar, quando ele passa a acreditar que há uma razão para ele fazer ϕ , se não for a proposição, ou algo que implique a proposição, de que se ele deliberou racionalmente, ele estaria motivado a agir apropriadamente?” (WILLIAMS, 1979, p. 109, tradução minha).

do agente), uma nova motivação. Essa seria *uma versão de OP que poderia ser compartilhada por outros internalistas*.

Agora, vamos ver a *versão original de Williams*, que *critica* não só externalistas, mas *também aqueles que não defendem versões da teoria de motivação humeana*. (Para o texto original de Williams, esses eram também externalistas). Novamente, quando um agente passa a acreditar em uma proposição de razão, uma nova motivação deve surgir para que ele aja de acordo e seu comportamento possa ser explicado. A única diferença dessa versão do argumento para a anterior é que o modo pelo qual a crença produz a motivação é *através* da motivação anterior do agente. Isto é, a crença interfere nos elementos de S já existentes e muda-os de modo que o S fique conativo à ação. Em outras palavras, a versão I de OP entende “motivação” apenas como a motivação nova (como em IR), ao passo que II a entende tanto como a motivação nova como a motivação pré-existente (os elementos de S, os desejos). Ainda assim, esse argumento parece convincente. Se chego à conclusão de que tenho uma razão para visitar minha avó, isso parece se dever em parte pelo meu carinho a ela – um conjunto motivacional subjetivo conativo a essa ação. De todo modo, essa versão do argumento aponta: de onde vem essa nova motivação se não a partir da minha motivação pré-existente? Esse é um ônus de prova aos “externalistas”.

“*De certas crenças*” poderia facilmente responder *Korsgaard*. Em “Skepticism About Practical Reason” ela oferece uma resposta a essa segunda versão de OP. (KORSGAARD, 1996). Primeiro, ela aponta que Williams compartilha de um ceticismo motivacional quanto a razões pra ação (como Hume): há dúvidas de a capacidade da razão de poder motivar por si mesma – Williams defende que a razão e seus raciocínios precisam de motivação anterior. Ela aponta que Williams (e Hume) só pode partilhar desse ceticismo porque, na verdade, ele supõe um outro tipo de ceticismo: o ceticismo de conteúdo de razões para ação. Esse último é um que afirma que as *deliberações para ação funcionam apenas de modo instrumental*. Por “*instrumental*”⁵⁸ penso que ela *corretamente* quer dizer, como aponta Setiya, *raciocínios* – como os de Williams – *que são “governados por e visam desejos”* (SETIYA, 2011, p.5). Já o raciocínio que a deliberação abrangente de Williams

⁵⁸ Esse, na verdade, é um termo importante pressuposto nas reconstruções de Williams, mas que só estou abordando agora.

exclui, o raciocínio da razão pura⁵⁹, é um que não é nem governado por (não depende de motivação anterior) e nem visa desejos (não é um tipo de raciocínio que visa promover ou satisfazer os elementos de S)⁶⁰. Assim, Korsgaard aponta que OP (na formulação II) só funciona com a seguinte hipótese: não existem razões puras⁶¹. Logo, se elas existirem, Korsgaard pode responder a Williams sobre a pergunta “de onde vem essa motivação?” do seguinte modo: de crenças sobre razões para ação que são produto de deliberações envolvendo o imperativo categórico. Essas, ela diria, podem nos motivar por si mesmas. Além disso, outras explicações de ações por razões são preservadas: Korsgaard ainda poderá dizer que visitei minha avó pelo carinho que sinto por ela. Raciocínios instrumentais são permitidos. Tudo o que Korsgaard sugere é que nem todos são instrumentais – e que todos sejam compatíveis com a moralidade ao serem compatíveis com o imperativo categórico.

O resultado disso é que essa objeção de Korsgaard enfraquece bastante a segunda versão de OP - enquanto que a primeira é um argumento sólido. Como apontei, isso depende, no entanto de uma hipótese: da hipótese da existência de razões puras e de raciocínios puros – razões que são produtos apenas da razão e raciocínios que não dependem de qualquer motivação anterior. Essa é uma tese que, por outro lado, muitos não aceitam. Além disso, todos os argumentos apresentados aqui, a favor da teoria de Williams, dependem da Restrição Explicativa e da Restrição Deliberativa⁶². E uma depende da outra, como mostra o argumento de Paakkunainen da seção 5 do capítulo anterior. Assim, uma última empresa neste trabalho será a de avaliar algumas objeções a Williams baseando-se nessa noção (RD), além de um argumento a favor dela apontado por Paakkunainen.

⁵⁹ As razões puras e seus raciocínios são os que funcionam como parte de uma deliberação por um princípio como o imperativo categórico. Devido a isso, elas exercem justificação última e devem ser capazes de motivar por elas mesmas a todos na medida em que são racionais. Assim, elas são universais, pois os agentes são capazes de serem motivados por elas enquanto são racionais (KORSGAARD, 1996).

⁶⁰ Entre parênteses está a minha interpretação daqueles dois sentidos de instrumentalidade apontado por Setiya – embora o próprio Setiya não os nomeie assim. Estou usando também uma interpretação de Korsgaard que entende que o que existe em todo S, como apontado por razões puras e o que nos torna motivado por elas, é uma disposição e não um outro mero elemento de S. Essa interpretação e essa distinção entre elemento de S e disposição de S vem de Setiya (2011, p.6).

⁶¹ Na verdade, há uma outra hipótese que Korsgaard (1996, p.24) aponta: a de que razões puras podem, de fato, nos motivar por si mesmas. Se minha interpretação estiver correta, ela aponta que Kant já demonstrou isso.

⁶² Creio que, mesmo que em menor medida, o argumento sobre *como* são as relações entre razões também depende de RD, uma vez que um tipo delas são as razões explicativas, que dependerão de deliberação, como o argumento de Paakkunainen na seção 5 do capítulo 1 demonstrou.

Capítulo 3: Inadequações extensionais e Restrição Deliberativa

3.1) analisando objeções: otimismo deliberativo e falácia condicional

Nesta seção avaliarei uma série de “inadequações extensionais”: contraexemplos e consequências indesejáveis da teoria de Williams. Começarei com uma consequência indesejável. Uma das principais críticas à Restrição Deliberativa é que ela supõe um *otimismo injustificado sobre nossas capacidades deliberativas*. Como diz Setiya:

“Para o internalista, cada um de nós, não importa o quão deficiente ou de maus hábitos, tem a capacidade de ser movido por qualquer razão para a qual ele é sujeito. Se nós somos capazes de ser movidos por razões em proporção aos seus pesos, a consequência é mais dramática: que aqueles que conseguem agir por razões consegue fazê-lo perfeitamente. Por que acreditar nisso? Por que acreditar que nosso potencial é tão sublime, que não conseguimos ser sujeito a razões que não podem nos mover? O que há sobre a natureza da agência, ou a metafísica e a epistemologia de razões, que torna tal incapacidade impossível?” (SETIYA, 2011, p. 11, tradução minha).

Algumas *preferências interpretativas* pelas quais argumentei ao longo desse trabalho nos ajudam a responder à essa preocupação quando pensamos na Teoria de Razões de Williams. Primeiramente, a TRW’ abrange razões pro-tanto e razões decisivas. Mas essa capacidade ainda parece forte demais: a pergunta pode ser reformulada para “por que pensar que razões decisivas podem sempre nos mover?”. E então uma outra escolha interpretativa se torna importante: não há necessidade de essas razões, mesmo decisivas, nos moverem. Na verdade, o que as razões precisam fazer é *nos motivar*. Essa motivação pode levar à ação, mas não necessariamente. Diversas vezes, apesar de sabermos que devemos agir diante da injustiça, deixamos de agir por, por exemplo, medo de retaliação. E quanto a alguém deficiente, como Setiya sugere? Deixarei esse caso para o fim da próxima seção, quando tratar de racionalidades atípicas. Assim, penso que essa preocupação quanto a esse otimismo está, por hora, mitigada⁶³.

⁶³ É interessante notar que Korsgaard tem um otimismo ainda maior do que o de Williams: ela pensa que os agentes não apenas são capazes de ser motivados a agir por razões, mas, quando eles

Um *segundo problema* para RD são preocupações advindas da *falácia condicional*. McDowell, que compartilha de RD apesar de ser um externalista, enfrenta dificuldades quanto a isso. Ele pensa que a deliberação deve ser feita, não exatamente pelo próprio agente (McDowell, 1995), mas por uma outra versão sua que delibera corretamente por ser virtuoso⁶⁴. Williams afirma que os não virtuosos podem ter razões para ação que os virtuosos não tem (PAAKKUNAINEN, 2018, p.14). Por exemplo, alguém intemperado pode ter boas razões para evitar certas tentações. Assim, a falácia consiste no fato de que “a condição necessária proposta sobre razões colocaria a pessoa totalmente virtuosa em circunstâncias em que ela não poderia estar – nas circunstâncias de ausência de virtude – compativelmente com a sua suposta virtude.” (PAAKKUNAINEN, 2018, p.15).

Paakkunainen (2018, p.15 apud SOBEL 2001; JOHNSON, 1999) aponta que *essa mesma falácia poderia ser levantada contra Williams*. Como vimos, TRW’ afirma que uma das condições necessárias para um agente ter uma razão é a de que ele tenha a capacidade atual de, em mundos possíveis próximos em que ele está na situação em que crenças são corrigidas e que ele delibera de modo abrangente e sólido, chegar à conclusão de que ele tem uma razão para uma ação. Porém, será que esse agente ideal tem, de fato, as mesmas razões que o agente atual? Quando suas crenças são corrigidas, ele não deixará de ter razões para, *exatamente*, corrigi-las? Parece que o agente na situação ideal deixará de ter razões para se informar mais adequadamente sobre se tem razão para ϕ , uma vez que *ele já tem todas as crenças relevantes e já foram excluídas todas as falsas sobre isso*. (Considerações como essa é que levaram alguns a abraçar o modelo de conselheiro de conselheiro). (PAAKKUNAINEN, 2018, p. 15).

Compartilho da dúvida de Paakkunainen sobre a capacidade dessa falácia ser uma que se aplica a versões de RD que concluem IR – como a de Williams (ibid, p.16). Como já afirmei, quando se diz que o agente deve ser capaz de deliberar de modo a chegar na conclusão de que ele tem uma razão para fazer algo, não é

deliberam corretamente, essa motivação nunca implica – em qualquer grau – em nos motivar a ter ações imorais (PAAKKUNAINEN, 2018, p.12, usei aqui “motivar à ação” no lugar de “levar à ação”).

⁶⁴ McDowell rejeita IR por acreditar que agentes não virtuosos várias vezes não são capazes de deliberar de modo a saber se tem uma razão para ação. Razões sempre motivarão *apenas* agentes virtuosos. Assim, McDowell rejeita IR ao pensar que a transição para sermos movidos por razões é uma *transição que envolve sermos virtuosos* e não meramente uma transição *efetuada* pela deliberação (PAAKKUNAINEN, 2018, p. 11). Em outras palavras, para sabermos as razões que temos, seria necessário um perfil psicológico específico, o de um *phronimos*.

necessário pensar essa mesma deliberação para todas as razões que ele possui. É preciso apenas que haja uma rota deliberativa possível para essa ação – não importa se essa rota deliberativa exclui ou não outras razões quando estamos pensando nela. É possível *pensar assim em cada rota deliberativa separadamente*: a deliberação para saber se tenho *razões para adquirir informações* sobre visitar minha avó é independente da deliberação para saber se tenho uma *razão para visitá-la*⁶⁵. Posso concluir que devo adquirir essas informações em um mundo possível e em outro concluir que devo visitá-la. Assim, essa falácia é respondida. Penso que é, no fim das contas, devido a esse aspecto independente das rotas deliberativas sólidas que Paakkunainen pode também dar sua resposta à falácia condicional.

A resposta de Paakkunainen (2018, p.16) é a seguinte. Ela diz que deve ser possível que um *agente delibere solidamente para concluir uma razão enquanto ele mantém todos os fatos relevantes de suas circunstâncias atuais (mesmo suas imperfeições)* que são relevantes para se chegar nessa razão (ibid). Ela aponta que tudo o que é restringido é que se um agente visa, por exemplo, corrigir suas informações, ele deve poder fazê-lo sem que seja necessária uma rota deliberativa na qual para eliminar suas crenças falsas, seus passos deliberativos dependeriam de ele já não ter essas crenças falsas. Assim, se descobro que minhas crenças sobre a história de Siracusa estão repletas de erros (talvez por falhar em uma prova), posso deliberar de modo a tomar passos para removê-los sem que eu já os tenha removido: posso estudar certos livros, perguntar a especialistas, etc. De modo similar, posso adquirir novas crenças sem que minha rota deliberativa de algum modo dependa delas: se tenho razões para aprender sobre a história de Siracusa, posso deliberar sobre como aprender mais sem que eu já tenha adquirido as informações que busco.

3.2) Contraexemplos a RD: James Bond, Soldado; e racionalidade atípica

Vamos ver alguns outros contra-exemplos a RD e a Williams. Paakkunainen

⁶⁵ E, como já apontei, a possível independência de rotas deliberativas depende de não usarmos uma interpretação específica de quais são as crenças relevantes. Podemos, por exemplo, ter como crenças relevantes *somente* as crenças *necessárias* para concluir uma razão – tendo assim uma rota simples. Assim, creio que uma resposta ainda melhor seria dizer que não só podemos ter rotas deliberativas independentes, como podemos, na verdade, *variar* as deliberações e as crenças relevantes. Podemos ter deliberações que concluem razões decisivas ou razões que concluem razões pro-tanto. E por causa disso também variamos o que é relevante que o agente saiba para mais ou para menos.

(2018, p.17) traz esses dois:

James Bond: O fato, p , que eu tenho uma crença ilusória que sou James Bond, é uma razão para eu buscar ajuda psiquiátrica. Mas p é um fato, e assim uma razão, apenas se eu não acreditar nele, e assim apenas se eu não tomá-lo em consideração na deliberação. Pois, se eu acreditasse que eu tenho uma crença ilusória de que sou James Bond – isto é, se eu pensasse a minha crença que sou James Bond *como* ilusória – isso terminaria minha ilusão. (ibid., apud MARKOVITS 2014, p.41, discutindo um caso de JOHNSON, 2003, p.575, tradução minha)

Soldado em uma Guerra Justa: o fato, p , que os habitantes do estado inimigo compartilham uma humanidade em comum, é uma razão para um soldado lutar em uma guerra justa. Mas o soldado não deveria pensar sobre esse fato ou ser movido por ele no meio da batalha para que ele não perca a sua coragem e assim batalhe (talvez perigosamente) de modo ineficaz. Assim, parece que não seria parte da deliberação sólida agir com base em razões nesse caso. (ibid., apud MARKOVITS, 2014, p.47, tradução minha)

Vamos analisar primeiramente o caso do *Soldado*. A partir dele se tenta mostrar que é melhor ignorar o fato que fornece razões, p , para que o desempenho do soldado não seja prejudicado. Afinal, é plausível que na situação de combate o soldado deixe de ser capaz de ter a conclusão de que deve participar da guerra: o medo da morte pode torná-lo em um combatente ineficiente. Esse caso apenas mostra que a razão falha, em certa medida, a *levá-lo a lutar eficientemente*. Mas 1)⁶⁶ *participar da guerra é uma ação diferente de lutar eficientemente* (como seguir uma tática específica) e, mais importante, 2) *a deliberação deve apenas motivá-lo em certa medida* (na minha versão de RD⁶⁷), não sendo preciso especificar o grau da motivação que deve ser produzida – como uma que leva alguém a agir sem temor. Como diz Paakkunainen (ibid., p.18), 1 nos aponta que a própria situação do combate é uma em que considerações sobre a humanidade em comum do inimigo são possivelmente irrelevantes: isso não traz uma razão para seguir uma tática específica em comparação à outra. Por outro lado, há situações em que essa

⁶⁶ Paakkunainen (2018, p.18) aponta isso.

⁶⁷ Não deixei explícito, mas minha versão da Restrição Deliberativa deve ser uma em que, como apontei anteriormente, o agente deve ser capaz de deliberar para as conclusões e então ser motivado pela conclusão. Afinal, se todo agente é capaz de descobrir suas razões, ele também deve ser capaz de ser motivado por elas – e entendendo “ser capaz de ser motivado” como uma capacidade atual.

consideração pode ser relevante – como para evitar brutalizar os habitantes. Penso que se o objetivo de uma deliberação for para saber a razão de *participar na guerra*, não haverá qualquer impedimento. Já se o objetivo for achar razões para *lutar eficientemente*, creio que o soldado pode ainda achar razões e motivações, embora possa não as exercer tão bem quanto gostaria. Assim, esse é um exemplo que não apresenta problemas para RD ou TRW’.

Agora, o caso *James Bond* tenta apontar que um certo agente na situação atual tem uma razão para visitar um psiquiatra, mas não na situação contrafactual. Ele *atualmente* pensa ser James Bond, mas esse *autoengano será corrigido na situação contrafactual* e assim isso *não pode mais funcionar como uma razão* para ele visitar o psiquiatra. Paakkunainen (2018, p.18) aponta que mesmo se supormos que não podemos ser motivados pela crença de que “estou sendo iludido de que sou James Bond” enquanto continuamos a acreditar ser James Bond, esse fato deve ser um que indica que devo ir buscar um psiquiatra *em algum sentido*. Ela diz que não é claro se *devemos interpretar* essa situação como o caso sugere. Isto é, 1) com *p* sendo uma razão para eu – “James Bond” – buscar ajuda. Ela faz algumas outras sugestões interpretativas. 2) Pode ser que uma crença geral próxima, como a de que eu tenho *uma* crença ilusória, já possa ser uma razão para eu buscar um psiquiatra. 3) Pode ser que se eu estivesse em um consultório, seria *bom* ou *bom para mim* tornar evidente a natureza das minhas ilusões. 4) Pode ser que esse fato seja uma razão normativa *para outros* buscarem ajuda para mim ou me exortar a ir. *Paakkunainen pensa que não é possível escolher entre a melhor interpretação* desse caso com justificações puramente pré-teóricas. Ela oferece então *duas alternativas* para James Bond e casos similares: podemos fornecer argumentos sólidos para a Restrição Deliberativa e então encontrar leituras compatíveis com RD ou tratar os contraexemplos como casos atípicos.

Discordo de Paakkunainen na ideia de que não é possível escolher a melhor interpretação do caso com justificações pré-teóricas. Creio que vários casos “problemáticos”, como James Bond, podem ser compatibilizados com RD. Vamos olhar as interpretações possíveis. 2, 3 e 4 parecem todos verdadeiros. E, de fato, provavelmente são razões para “James Bond” e pessoas próximas dele. Mas elas tratam ou de outras razões que não *p* ou de outras pessoas. Como RD se liga com RE (pelo argumento de Paakkunainen) e como a melhor interpretação de “capaz de

explicar” (uma consequência de RE por melhor explicação) é a *particularista* (a terceira), temos de entender o caso James Bond no sentido 1. Isso, porém, parece ir contra reações intuitivas que temos, como diz Paakkunainen (2018, p.18). Parece que *p* deveria ser uma razão para o agente visitar um psiquiatra. Tentarei mostrar que temos razões ainda mais intuitivas para pensar o contrário.

Primeiro, devo enfatizar que estou tratando de se *p* (a crença de que o agente é James Bond) é uma razão para o agente – e não as outras interpretações sugeridas. *Segundo*, discordo que as crenças precisem funcionar desse modo bipolar. Se a família dele conversar com ele e ele perceber várias incongruências em sua autoilusão, isso não significa que ele passará a *acreditar* que tem uma autoilusão e assim ele estará convencido de que deve ir ao psiquiatra. Ele pode passar a *suspeitar* que tem uma autoilusão⁶⁸. Essa suspeita pode ser suficiente para compor o link entre o fato e a motivação do agente (S e a nova motivação). Ele será motivado por *p* ao mesmo tempo em que crê ser James Bond. Acredito que isso já é suficiente para responder ao contraexemplo.

Ainda assim, vamos tentar supor que não existe essa gradação – ou até que “James Bond” é alguém que é incapaz de ter uma gradação como essa. Desse modo, deve ser dito que “*James Bond*” é *certamente uma pessoa com uma patologia* ou algo próximo – já que pensamos que ele deve ter uma razão para ir a um psiquiatra. (Se o autoengano dele fosse não patológico, como por alguma circunstância fantástica em que todos os seus vizinhos decidem pregar uma peça nele por anos para convencê-lo de que ele é James Bond, isso não indica uma razão para ele ir a um psiquiatra, mas para mudar de cidade). Isso nos mostra que suas razões poderão ser diferentes das nossas, uma vez que sua capacidade de deliberar para razões é em alguns sentidos prejudicada. E isso, eu acredito, indica que *I) James Bond deve poder ter crenças contraditórias e II) devem existir tipos de racionalidade atípicos* – embora ainda assim compatíveis com RD.

Vamos tratar primeiro de I. Podemos nos esquecer que se o agente possui alguma patologia que faz ele pensar ser James Bond, uma mera crença de que ele

⁶⁸ Não sei se devemos entender a atitude de *suspeitar* como um grau mais fraco de crença ou como um outro tipo de atitude, mas para ser compatível com TRW', deveríamos entender do primeiro modo. Porém, mesmo se suspeitar fosse um tipo de pro-atitude diferente de *crer*, penso que essa seria uma mudança tão mínima que poderíamos ainda salvar TRW', uma vez que isso está longe de ser o aspecto mais importante dessa teoria.

está se autoiludindo não deve ser capaz de mudá-la. Assim como alguém com esquizofrenia paranoide tem alucinações que não se pode deixar de perceber (sendo evidências para ele), mesmo sabendo que elas são falsas, o mesmo deve ocorrer com um delírio como o de ser James Bond (na verdade, delírios como o de acreditar ser outra pessoa são até mesmo um outro sintoma dessa doença⁶⁹). O que podemos concluir com isso é que “*James Bond*”, por ter uma patologia, pode certamente ter crenças contraditórias⁷⁰. Vou tratar de *II* agora: mas não é intuitivo pensar que *p* é uma razão para “James Bond” visitar o psiquiatra? Penso que essa intuição está errada, mas pode ser bem explicada. *É o fato de que não estamos levando em conta sua patologia que explica isso*. Por ser doente, “James Bond” é incapaz de ter essa razão, assim como um cadeirante é incapaz de ter como razão a crença de que “correr é o modo mais rápido para eu chegar à parada de ônibus”. *Essa razão não se aplica a ele*. E isso nos aponta que sua racionalidade é atípica neste aspecto. Isto é, que existem outros tipos de racionalidade não típicos, para os quais os agentes possuem razões diferentes. Penso que essa noção pode nos ajudar a explicar esse e outros casos similares – principalmente de uma perspectiva Williamsiana.

Além disso, essa noção de *atipicidade* nos ajuda a fortalecer o argumento de Paakkunainen sobre casos como esse. Ela indica que não é claro se há um custo teórico intuitivo de considerarmos casos como James Bond (ou outros que não conseguimos dar conta) enquanto “atípicos”, ao passo que negar a conexão entre razões e deliberação é bem alta. (PAAKKUNAINEN, 2018, p.19). *Primeiramente*, note que a noção dela de “atípico” é diferente da minha. *A que ela sugere é produto de uma certa incapacidade da teoria*, mas que não possuiria custo teórico intuitivo, ao passo que *a noção que estou sugerindo é, na verdade, produto de explicação*. *Segundo*, penso que o custo de a possibilidade de *p* não ser uma razão para “James Bond” pode não ser tão claro, mas *parece* indicar uma certa intuição e assim um custo. O que estou sugerindo diante disso, com essa outra noção de *atipicidade*, é que essa intuição é infundada. “James Bond” tem uma doença psicológica e por

⁶⁹ PIMENTA (2019).

⁷⁰ É certamente possível que alguns casos de “James Bond” possam, mesmo com essa contradição de crer que são e não são JB, serem motivados a ir ao psiquiatra por isso – afinal, elas teriam também a crença adicional e implícita de que uma contradição como essa é patológica. Mas não são esses os casos atípicos que estou tratando. Estou tratando de uma pessoa que não consegue *perceber* plenamente essa contradição *ou* que meramente não se motiva a ir ao psiquiatra por ela.

causa disso sua racionalidade quanto a p é atípica: essa não é uma razão para ele.

Mas por que pensar que existem outros tipos de racionalidade – as atípicas? Como tentei apontar com a esquizofrenia (que também penso ser uma das melhores, se não a melhor, entre as interpretações de James Bond), para certos tipos de pessoas e até em certos momentos de nossas vidas, possuímos circunstâncias em que as razões parecem se aplicar de um modo diferente do normal. Não me refiro apenas a casos patológicos (o que inclui casos de amoralidade) ou de deficiências físicas, mas também casos de bebês ou de certos impedimentos atuais. Por esse último tipo me refiro a casos em que até mesmo um pequeno fator pode tornar nossa racionalidade em atípica, ou melhor, *atípica para aquela ocasião*. Suponha que o caso James Bond é um que não há qualquer patologia envolvida, se não a do delírio do agente pensar ser James Bond (se é que esse é um caso humanamente possível⁷¹). Já poderíamos entender esse como um caso atípico, em que o agente possui razões diferentes das normais, pois ele não pode ser motivado pela consideração p enquanto uma razão sem deixar de ser atípico.

Creio que uma noção de *racionalidades atípicas* ou diferentes das normais não é algo tão atípico ou original quanto possa a princípio parecer. Wong, por exemplo, realiza um argumento baseado em uma pressuposição similar quando ele argumenta que o corpo é uma condição constitutiva da ação corpórea (WONG, 2020). Quando olhamos para casos de agentes deafferentivos⁷², o que fica evidente é que o modo de eles agirem é diferente do nosso. Certos aspectos de suas ações devem funcionar de modo diferente – eles precisam ser suplementados por uma vigilância constante e ainda assim agem sem fluidez. Por causa disso, o modo de explicar a agência normal é a partir do corpo como uma condição constitutiva, que nos auxilia a explicar ações. De modo similar, e com exemplos mais comuns, pessoas com certas patologias terão aspectos ausentes de uma racionalidade típica, e por isso devem ter aspectos como a sujeição a razões para eles próprios

⁷¹ Imagino que certas drogas aliadas de hipnose (ou somente hipnose) podem induzir certas pessoas a esse estado, mas não é isso o que queremos. O que o contraexemplo trata, novamente, é de p : que o agente pense que tem uma razão específica de ir ao psiquiatra, isto é, a de ter pensamento de ser James Bond. Ademais, não estou considerando deliberações ruins aqui, como a de que *um mero pensamento passageiro* de “eu ser James Bond” ser uma razão, por si só, para “James Bond” visitar o psiquiatra. Como apontarei depois, deliberações ruins não são bons candidatos para RD.

⁷² “tradução minha de “deafferented agents”, ou seja, aqueles com perda de propriocepção ou toque ou até outros problemas no sistema vestibular.

diferentemente.

Acredito que considerações como essas nos ajudam também a mitigar outras supostas inadequações extensionais de TRW'. Uma das principais objeções a ela é que ela vai contra o *Absolutismo Moral (AM)*. Essa tese diz que algumas ações são moralmente erradas para qualquer agente, independentemente dos estados internos deles. Por exemplo, se alguém vê um banhista se afogando e pode salvá-lo, ele deve fazê-lo. Uma outra objeção, menos exigente, é a de que às vezes agentes não terão razões normativas, apesar de termos certas intuições para isso (um certo *pessimismo*). Um exemplo famoso de Parfit (2011, p. 73-74) aponta que parece muito contraintuitivo que agentes possam vir a não ter razões decisivas para *evitar a agonia futura*.

Penso que quando enfatizamos que o agente já está *na situação ideal*, por já ter deliberado solidamente e com todas as suas crenças corrigidas, parte dessas intuições é respondida. A parte “nova” que sugiro aqui é a de que um sujeito que não tem razão para agir em casos de inadequação extensional, como esses dois últimos, pode não possuir essas razões por sua racionalidade *ser atípica*. Um agente que não tem razões para salvar alguém (mesmo sendo algo fácil e sem custo para ele) é provavelmente um *verdadeiro psicopata*⁷³. Esse seria um caso muito *mais raro* de indivíduos, que, *em certas circunstâncias*, podem não ter razões para agir de modo moral⁷⁴ - ou pelo menos, que não tenham uma razão específica *p* (*seja o que for neste caso*) como uma razão para isso. O mesmo pode ser dito de um caso parecido com o agente de Parfit – que também é patológico. Algo como uma depressão profunda – e realmente sem esperança⁷⁵ – ou outras doenças aliadas a

⁷³ Com “verdadeiro psicopata” não estou me referindo ao tipo de psicopata comum, que por vezes são nossos colegas de trabalho, vizinhos e outros. Esse agente que não tem uma razão para salvar o banhista se afogando deve ser um com uma altíssima pontuação na Escala PCL-R. Não meramente a partir do ponto de corte de 30, mas consideravelmente acima disso e provavelmente chegando próximo do limite de 40 pontos (como sugerido pelos manuais PCL-R/SV; ver Skeem et al. 2011). São esses os assassinos em série, pedófilos e estupradores várias vezes reincidentes – entre outros.

⁷⁴ Isso não significa que seus atos não são imorais, mas que por não terem a capacidade de serem motivados em certos casos, a agir de modo moral, eles não terão certas razões também.

⁷⁵ Do modo que Parfit constrói o caso, a agonia não pode ser usada de modo instrumental (PARFIT, 2011, p.73-74). Devido a isso, agentes em estado tão depressivo que preferam sentir agonia a ter de se manter em seu estado de desinteresse e desesperança não são possíveis. Mas penso que não são apenas agentes em estado tão desesperador que até a agonia futura lhe é preferível que teriam um desejo por uma agonia futura. Há teoricamente também casos de agentes que, por seja qual for o motivo – como uma depressão profunda e irreversível, uma doença neurológica ou até uma cirurgia maliciosa –, vêem a agonia como boa em si mesma – ou pelo menos neutra. (Sendo neutra, o agente já não teria razões para evitar a agonia futura). Para uma outra, e mais completa, resposta a esse argumento, ver Sobel (2009).

circunstâncias desesperadoras podem-lhe ser apontadas e desse modo também uma racionalidade atípica. Precisamos dar conta desses agentes e vê-los enquanto casos atípicos de racionalidade ajuda a fortalecer RE (e assim RD). RD também é fortalecido diretamente por poder lidar com James Bond – a intuição que temos para aquele caso será infundada. Mas além disso, essa noção fortalece a versão de razões para ação de Williams, que consegue explicar as ações deles *exatamente* pela referência ao S anterior.

3.3) Por RD: relação intuitiva e demandas autoritárias

Como afirmei antes, Paakkunainen (2018, p.19) diz que o custo teórico de rejeitar RD é alto, enquanto que não há custo intuitivo de considerar certos casos como James Bond enquanto atípicos. Mas por que pensar que ele é bem alto? Que vantagem a Restrição Deliberativa nos oferece? Paakkunainen aponta dois argumentos a favor dessa noção. O *primeiro* aponta *uma relação muito intuitiva de razões e nosso descobrimento cotidiano delas pela deliberação*. Quando estamos diante de decisões importantes, tendemos a pensar que faz sentido entender os fatos relevantes a elas e tomar decisões com base nesses fatos. Tentamos pesar razões favoráveis e contrárias a certos cursos de ação e pensamos não apenas que pela deliberação é possível encontrá-las, mas também que deliberando bem (melhores informados) e em boas condições – não estando incapacitados, como bêbados e etc. –, nossa chance de encontrá-las será ainda maior. (Na verdade, ela se refere aqui somente a razões decisivas, mas também podemos expandir esse argumento para razões pro-tanto).

O *segundo argumento de Paakkunainen a favor de RD* é mais complexo. Ele aponta que a Restrição Deliberativa é um modo plausível de *entendemos razões como impondo demandas autoritárias*. Quando dizemos que “se A tem uma razão decisiva para ϕ , então A deve fazer ϕ ” esse “deve” não é no sentido de uma mera regra de um jogo ou de um clube a ser seguida, mas é *um “deve” robustamente normativo ou autoritário*⁷⁶. Mas por que pensar que a autoridade de razões, como

⁷⁶ É essa ideia de razões como autoritárias que, por exemplo, estrutura debates sobre a autoridade da moralidade: eles normalmente pressupõe que a autoridade da moralidade sobre nós *depende* de termos razões para sermos morais. E essas razões morais devem ser autoritárias no sentido de que não devemos facilmente deixar ou escapar da moralidade: a moralidade impõe uma demanda incondicional. Ela deve nos ser inescapável. Assim, se há razões para agirmos moralmente bem, tais

decisivas – e em especial as morais – dependem da Restrição Deliberativa?

Perceba que há uma relação entre exigências autoritárias e expectativas razoáveis. Essa noção é parcialmente preditiva e parcialmente normativa. Se é exigido de A que em certa situação ele faça ϕ , então podemos esperar que exista uma condição possível X de modo que o permita realizar essa ação. A *parte preditiva* se refere ao fato de *esperarmos* que o agente faça a ação nessa condição X – pelo menos na ausência de outras condições que o escuse. Parece que se não existir uma tal condição X, a exigência que se coloca em A não é apropriada. Já a *parte normativa* diz que é razoável esperarmos de A que ele faça isso e que o *critiquemos* se ele falhar em fazê-lo. E A é *criticável* apenas na medida em que ele *age ou deixa de agir por si mesmo*. E para agir por si mesmo é preciso que ele *exerça suas capacidades de agência racional*: suas capacidades de considerar informações relevantes e agir à sua luz. (Afinal, assim como não criticamos pedras por seu comportamento, também não criticamos agentes por seus espasmos não-rationais e involuntários). Logo, a condição X precisa ser uma na qual A age a partir de considerações e escolhendo baseado nelas. Isso mostra que X é uma condição deliberativa. E sendo uma condição deliberativa, o tipo que melhor satisfaz é o de deliberação sólida⁷⁷, o que envolve considerar os fatos que impõe as exigências relevantes e ser movido com base nelas. Enfim, “razões conseguem impor demandas para A apenas se há uma condição deliberativa possível dele que o permita satisfazer essas demandas [...]” (PAAKKUNAINEN, 2018, p.20, tradução minha).

Paakkunainen faz duas sugestões baseada nessas relações. A primeira é que ela diz que esse argumento também sugere versões do Internalismo de Razões que são apoiadas por RE⁷⁸. A condição X deve ser uma encontrada em mundos possíveis próximos ligados à capacidade do agente, em que podemos razoavelmente esperar certas ações como salvar alguém se afogando, em oposição a pular por cima do Empire State por acreditar que um evento quântico raro lhe permitirá fazê-lo em um salto. Essa suposta demanda é totalmente irrazoável pelo fato de o agente não ter essa capacidade em mundos possíveis próximos – seria

razões devem ser autoritárias. (PAAKKUNAINEN, 2018, p.19)

⁷⁷ É difícil de pensar que uma deliberação ruim ou levar em conta fatos irrelevantes sobre fazer possa levar agentes a fazer o que lhes é exigido.

⁷⁸ Já que podem existir versões de IR que não são apoiadas por RE (PAAKKUNAINEN, 2018, p.4).

incrivelmente difícil *fazer essa ação*. De modo similar podemos afirmar sobre nossas capacidades motivacionais: certos agentes, como não maduros ou com graus variantes de saúde podem ter demandas diferentes das quais temos. E assim as razões e obrigações que temos são, de algum modo, *restringidas* por nossas capacidades motivacionais. Se esse modo é restringido por RD e RE dependerá de como resolvemos questões sobre o *aspecto normativo de expectativas razoáveis*.

Esse argumento de Paakkunainen serve para mostrar como podemos dar conta do aspecto autoritário de razões – o que exige RD. Esse aspecto parece indicar que para o agente satisfazê-lo, ele deverá ser levado à ação – a versão preferida da autora quanto ao resultado da deliberação correta em situações ideais. Entretanto, assim como já argumentei anteriormente, não penso que razões devem necessariamente *nos levar à ação*. Isso seria extremamente problemático para razões pro-tanto e isso exigiria um certo S específico do agente para que ele sempre fosse levado a agir de acordo com razões decisivas – o que se aproximaria de um externalismo como o de McDowell. Afinal, novamente, podemos deixar de agir não porque não estamos convencidos do curso de ação correto, mas porque o medo nos impede. Porém, não podemos tão bem criticar o agente que *meramente* não passou a estar motivado. *Parece que a crítica deveria se referir a suas ações*. Mas ainda assim, dado que quando a motivação gerada pela razão (seja como dependendo ou não do S) pode nos levar à ação, podemos de fato criticá-los. Logo, penso que essa é uma *relação mais indireta* do que a que Paakkunainen quer defender. E, partindo dela, esse segundo argumento ainda pode funcionar.

E quanto à *segunda sugestão* ligada às demandas autoritárias? Como vimos, um dos resultados da TRW' é que certos agentes suficientemente asquerosos não terão razões morais, por exemplo (isso é parte do *pessimismo* mencionado antes). Williams sugere que podemos ainda assim criticá-los sob outros aspectos: não é possível dizer que eles deviam ou tinham uma razão para fazer aquela ação, mas ainda podemos dizer que eles são egoístas, brutos, ingratos e vários outros adjetivos desfavoráveis. A segunda sugestão de Paakkunainen consiste em afirmar então que *a crítica possível de Williams ao agentes imorais, por causa da importância normativa de razões (como impondo demandas autoritárias), é difícil de aceitar* (PAAKKUNAINEN, 2018, p.22). Isso implicará que agentes suficientemente asquerosos não terão razões morais.

Ao trazer essa objeção, Paakkunainen, no fim, se apega ao Absolutismo Moral. Essa objeção dela depende desse pressuposto que Williams já não aceitava. E Williams não a aceitava porque isso depende da hipótese da existência de razões puras. Para podermos responder a essa intuição, teremos que supor ou defender essa hipótese. Penso que isso nos deixa com uma escolha. Na filosofia, por vezes nos agarramos a certas intuições, ao passo que abrimos mão de outras (como ser cético quanto à existência de razões puras). Preferir se agarrar ou não a AM ao custo de passar a aceitar que temos uma faculdade da razão pura parece ser mais uma delas – até porque muitos não vêem essa hipótese como custosa, senão o contrário.

Conclusão:

Farei duas formulações quanto às conclusões deste trabalho, ressaltando o que foi feito para assim mostrar o que foi atingido. Primeiramente, formularei de modo mais geral, para só então abordar as conclusões de modo mais detalhado.

Neste trabalho, como a indicação por capítulos mostra, busquei *I*) apresentar de um modo melhor a Teoria de Razões para Ação de Bernard Williams, *II*) analisar os principais argumentos dele a favor dela (juntamente com outros menores), *III*) analisar o principal pressuposto da teoria de razões de Williams, RD, visando principalmente argumentos e objeções que podem ser levantados na versão de Williams dessa tese.

No capítulo 1, aponte que a reconstrução de Setiya de TRW (em duas teses) já nos ajuda a perceber que o Internalismo de Razões não é equivalente a TRW. A divisão de TRW em pelo menos duas teses nos auxilia a mapear e melhor comparar diferentes posições sobre razões para ação – e podemos escusar Williams disso, já que foi ele quem primeiro estabeleceu essa distinção. Mas também foi apontado que seria melhor dividi-la, na verdade, em três teses para que uma delas não incorra em falsidades (já que motivação por deliberação podem ocorrer de modo sólido ou não). Devido a isso, TRW (agora TRW') deve ser entendida pelo Internalismo de Razões (IR) e pelas duas metades de uma Versão de uma Teoria de Motivação Humeana (VTMH' e VTMH''). IR aponta que razões precisam ser atualmente capazes de nos motivar. VTMH' aponta em que condições essa capacidade motivacional por um fato (que pode ser uma razão) ao deliberarmos deve ser exercida: quando possuímos rotas deliberativas *instrumentais* e abrangentes (sólidas ou não), de nosso S para um

desejo. VTMH” mostra que esse fato será uma razão quando pudermos não apenas deliberar, mas deliberar de modo sólido e baseado em correção de crenças para esse desejo. O resultado, TRW’, afirma que um fato é uma razão se podemos deliberar *abrangevolmente*, de modo sólido, com as crenças relevantes corrigidas e nos baseando nesse fato para a conclusão que temos razão para fazer algo. E por “abrangevolmente” me refiro aos diversos raciocínios incluídos por Williams, mas que somente funcionam de modo instrumental – o que implica tanto um raciocínio dependente da motivação anterior, o S, quanto um que promove ou satisfaz seus elementos, produzindo uma nova motivação.

Para Williams é a motivação que explica nosso comportamento por razões. E essa motivação surge com as razões devido aos dois aspectos da crença. Seu conteúdo trata dos fatos, enquanto que atitude pode influenciar nosso S anterior – e assim gerar um novo desejo. *Crenças sobre razões* também são resultados de se determinar *crenças cujo conteúdo são razões* – e assim Williams pode usar os dois sentidos em sua teoria. Um outro detalhe importante visto aqui foi a escolha da reconstrução da “capacidade” de Williams não em termos de mera possibilidade, mas de *capacidade motivacional* em mundos possíveis próximos (em última medida para evitar uma petição de princípio). Também defendi que IR não deve ser o mesmo que a Restrição Explanatória (dado que a distinção entre IR e Externalismo de Razões é mais relevante na teoria de razões para ação do que a de teorias com ou sem RE).

E um último detalhe a ser notado (que leva a outros): argumentei que *devemos entender IR como nos motivando* – com graus motivacionais variáveis – e *não nos levando à ação*. Razões pro-tanto, por exemplo, com frequência não nos levam a agir – e mesmo razões decisivas, como em casos de medo. Não penso que a objeção levantada nesse ponto (B: que a deliberação sólida gera múltiplas motivações conflitantes quando temos razões para ação conflitantes) torna isso problemático, principalmente por três considerações. Primeiro, podemos ter rotas deliberativas independentes ao permitirmos também rotas não decisivas. Segundo, razões podem nos motivar em graus diferentes (de modo que nem todas precisam nos levar a agir). Terceiro (que também possibilita a primeira): podemos variar o que entendemos como “relevante” na correção de crenças, nos permitindo ter diferentes processos para razões – das razões mais fracas até as mais fortes. O primeiro e o terceiro ponto indica que devemos *explicitar a correção de crenças*, o que também nos ajuda a evitar ler o

agente como o modelo de conselheiro, que já seria uma versão perfeitamente informada do agente e que não precisa passar por qualquer processo.

No capítulo 2, reconstruí os dois principais argumentos de Williams a favor de TRW'. Para isso, defendi que o melhor modo de interpretarmos o “ser capaz de explicar” de RE é enquanto uma tese particularista e como havia apontado de início, mostro que seu principal argumento parece só garantir IR – como a literatura vem argumentando. Esse primeiro consiste em apontar que o único ou melhor modo de se *explicar ações por razões é através da motivação* gerada – de algum modo ou outro – *por razões*. Apontei também uma certa *vantagem de TRW'*: ela pode melhor explicar a relação entre as diferentes razões: normativas, explicativas e motivacionais. Penso que a simplicidade de pensar que razões normativas são potenciais razões motivacionais – e neste sentido são também potenciais razões explanatórias – é no que consiste o apelo desta vantagem (afinal existem outras formas de se relacionar essas razões). O *segundo principal argumento* de Williams foi reconstruído em *duas versões*: uma visando defender somente IR e outra visando defender IR e VTMHs. Na primeira, “motivação” é sugerido como a única explicação possível de comportamentos por razões (assim como o primeiro argumento) e *se põe em dúvida outras possibilidades*. Essa versão pode ser utilizada por todos os internalistas que compartilham de RE. A segunda versão do argumento é uma mais fraca (e menos convincente), dependendo pelo menos da hipótese de não existirem razões puras. Essa entende “motivação” como a motivação anterior e a nova do agente (como mostrado pelos dois sentidos de instrumentalidade). (Novamente, em outras teorias internalistas a motivação se refere somente à nova motivação gerada direta ou indiretamente pela razão).

No capítulo 3, avalio algumas inadequações extensionais da teoria de Williams diante de ataques a RD (como Soldado e James Bond). Defendo que TRW' pode não só bem se defender deles, como também *não precisa* pensar alguns deles como atípicos no sentido de casos em que a teoria não pode dar conta, mas que é preferível pensá-los assim (como exceções que fogem à regra) para não incorrerem em um grave custo teórico. Argumento que podemos bem entendê-los com uma outra noção de *atipicidade* pela qual eles são casos de exceção *por* apresentarem alguma incapacidade de ver um fato como uma razão para eles. (Ou seja, se negarmos Absolutismo Moral (AM), podemos ter uma noção de Restrição Deliberativa (RD) que

abrange eles). Também é mostrado como a falácia condicional não se aplica a Williams – mas somente a McDowell: podemos tanto pensar em rotas deliberativas independentes umas das outras, quanto, como aponta Paakkunainen, podemos elaborar rotas que não dependem de mudarmos fatos relevantes das nossas circunstâncias atuais – por exemplo, pensar em como remover crenças falsas sem que eu já as tenha removido.

Apresento então *dois argumentos de Paakkunainen a favor de RD*. O primeiro aponta que há uma relação muito intuitiva entre razões para ação e nosso descobrimento cotidiano delas a partir da deliberação. O segundo mostra que RD é necessário para termos um bom modo de compreendermos o aspecto autoritário de razões. Tento apontar que a sugestão de Paakkunainen de que as críticas possíveis a agentes amorais (entre outros) com TRW' é dependente de AM, o que Williams já não aceitava. Por outro lado, AM pode ser enfraquecido por uma série de aspectos que tentei apresentar até aqui. Ela é enfraquecida quando percebemos que a grande maioria dos agentes terá de fato razões morais: agentes moralmente asquerosos são mais raros do que pensamos. (Essas devem ser somente pessoas com pontuações próximas de 40 no PCLR-R). Penso que a ideia de que eles são *incapazes* (por serem atípicos) também ajuda neste sentido. O mesmo pode ser dito do fato de que podemos ver eles como inclusos (e explicados) em TRW'.

Todas essas vantagens, entretanto, não impedem que teorias contrárias sejam propostas para *exatamente* dar conta de AM – uma intuição que muitos preferem manter. Podemos supor que existem razões puras e fazer uma teoria a partir disso. Korsgaard, por exemplo, faz isso. Outras vantagens que apontei aqui também passam por isso: podemos montar *outras teorias* sobre relações entre razões e argumentar que ações imorais não precisam ser incluídas em teorias de razão para ação – o que indicaria que realmente não há qualquer sentido em que psicopatas poderiam se justificar em fazer o que fazem. Tudo isso aponta que a teoria de Williams e os argumentos que temos a favor dela (e até de RD e EC) não são decisivos. Mas penso que essa é uma teoria que ainda pode competir com outras no debate contemporâneo de razões para ação. O fato de que ela consegue muito bem explicar comportamentos por razões é sua principal vantagem. Isso também, novamente, amortiza a rejeição de AM – uma vez que podemos entender a atipicidade de agentes amorais e outros, como James Bond. E se algum progresso foi feito durante este trabalho, foi mostrar

como podemos melhor interpretar, reconstruir, defender e responder objeções à Teoria de Razões de Williams.

Referências bibliográficas

- ALVAREZ, Maria, **Reasons for Action: Justification, Motivation, Explanation**, *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>
- AMBIEL, Rodolfo Augusto Matteo. **Diagnóstico de psicopatia: a avaliação psicológica no âmbito judicial**. *Psico-USF*, v.11, n.2, p. 265-266, jul./dez.2006
- DUTILH NOVAES, C. **Reductio ad absurdum from a dialogical perspective**. *Philos Stud* 173, 2605–2628 (2016).
- FINLAY, Stephen and Mark Schroeder, **Reasons for Action: Internal vs. External**, *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/reasons-internal-external/>.
- MATOS, Cristiann Wissmann. **Internalismo de Razões e a Deflação de Sistematização Ética segundo Bernard Williams**. Dissertação (Mestrado em filosofia) - Instituto de Filosofia e Ciências Humanas, Universidade Federal de Pelotas, Pelotas, 2016.
- MCDOWELL, J. (1995). **Might There Be External Reasons?** In J. E. J. J. Altham and R. Harrinsons (eds), *World, Mind, and Ethics*, 68-85. Cambridge: Cambridge Unver
- KORSGAARD, C. (1996), **The Sources of Normativity**. Cambridge: Cambridge University Press.
- PAKKUNAINEN, H. (2018). **Internal and External Reasons**. in *The Oxford Handbook of Reasons and Normativity* Edited by Daniel Star
- PARFIT, D. (2011). "On What Matters". Oxford: Oxford University Press.
- PIMENTA, T. Esquizofrenia paranoide: causas e tratamento. **Blog Vittude**,

2020. 5 de maio de 2019. Disponível em:

<https://www.vittude.com/blog/esquizofrenia-paranoide/>. Acesso em: 2 de novembro de 2020.

SETIYA, K. (2011), **Introduction: Internal Reasons** in Internal Reasons: Contemporary Readings edited by Kieran Setiya (Editor) and Hille Paakkunainen. Milton Press

SKEEM, J. L.; POLASCHEK, D. L. L.; PATRICK, C. J.; LILIENFELD, S. O. Psychopathic personality: Bridging the gap between scientific evidence and public policy. **Psychological Science in the Public Interest**, v. 12, n. 3, p. 95-162, 2011.

SMITH, M. (1995). **Internal Reasons**. Philosophy and Phenomenological Research, 55(1), 109.

SOEIRO, Cristina; GONÇALVES, Rui Abrunhosa. **O estado da arte do conceito de psicopatia**. Análise Psicológica (2010), 1 (XXVIII): 227-240

WILLIAMS, B. (1979). **Internal and External Reasons**. Repr. In Moral Luck, 101-13 (Cambridge: Cambridge University Press, 1981).

WILLIAMS, B. (1995). **Internal reasons and the obscurity of blame**. In Making Sense of Humanity: And Other Philosophical Papers 1982–1993 (pp. 35-45). Cambridge: Cambridge University Press.

WONG, H.Y. (2017). **Embodied Agency**. Philosophy and Phenomenological Research, 97 (3), 584-612.

WONG, H.Y. (2020) **Agency and the Body**, no prelo.