

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Filosofia e Ciências Humanas
Neiva Gonçalves de Oliveira - CRB 8/6792

M368a Maruchi, Gabriel de Andrade Pagnozzi, 2000-
O argumento direto pelo incompatibilismo / Gabriel de Andrade Pagnozzi
Maruchi. – Campinas, SP : [s.n.], 2022.

Orientador: Marco Antonio Caron Ruffino.
Trabalho de Conclusão de Curso (graduação) – Universidade Estadual de
Campinas, Instituto de Filosofia e Ciências Humanas.

1. Livre arbítrio e determinismo. I. Ruffino, Marco, 1963-. II. Universidade
Estadual de Campinas. Instituto de Filosofia e Ciências Humanas. III. Título.

Informações adicionais, complementares

Título em outro idioma: The direct argument for incompatibilism

Palavras-chave em inglês:

Free will and determinism

Titulação: Bacharel em Filosofia

Banca examinadora:

Marco Antonio Caron Ruffino [Orientador]

Pedro Merluzzi

Raoni Wohnrath Arroyo

Data de entrega do trabalho definitivo: 17-11-2022

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS
IFCH/UNICAMP

DEPARTAMENTO DE FILOSOFIA

TRABALHO DE CONCLUSÃO DE CURSO

O Argumento Direto pelo Incompatibilismo

DISCIPLINA: HG880 - MONOGRAFIA II

Aluno: Gabriel de Andrade Pagnozzi Maruchi

Orientador: Marco Antonio Caron Ruffino

Campinas - São Paulo
15 de dezembro de 2022

Resumo

Nesta monografia, discutirei o Argumento Direto pelo incompatibilismo. Incompatibilismo é a tese de que responsabilidade moral é incompatível com o determinismo nomológico. De maneira simplificada, o argumento é o seguinte: Nós não somos responsáveis pelo passado e pelas leis da natureza. Se o determinismo for verdadeiro, nossas ações são consequência do passado e das leis da natureza. Portanto, se o determinismo for verdadeiro, não somos responsáveis por nossas ações. Defendo ao longo da monografia que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo – a posição oposta ao incompatibilismo –, já que os proponentes do argumento falham em motivar com sucesso o argumento. A monografia se divide em três capítulos. No primeiro, apresento a formulação do Argumento Direto proposta por Peter van Inwagen (1983). Sua formulação se vale de um operador modal de não-responsabilidade, para o qual van Inwagen define duas regras dedutivas. Nos dois capítulos subsequentes, discuto algumas das críticas encontradas na literatura contra essas regras e apresento minhas próprias críticas. Na conclusão, considero alguns modos pelo qual o debate pode avançar.

Agradecimentos

Dizem que estudar responsabilidade moral te faz perceber o quanto coisas pelas quais você se orgulha ocorreram por fatores fora do seu controle – por sorte basicamente –, e que isso te torna mais humilde. Eu não acho que tenha ficado mais humilde nesses últimos anos, mas eu certamente passei a apreciar mais a sorte que tenho em vários aspectos. Entre eles, a sorte de ter encontrado várias pessoas que enriqueceram minha vida pessoal e acadêmica.

Tive a sorte do Daniel Coimbra me convencer a estudar sobre livre-arbítrio e frequentar as aulas do Pedro Merlussi. Tive a sorte do Merlussi aceitar me orientar no tema e do professor Marco Ruffino aceitar ser o orientador. Tive a sorte do Raoni Arroyo aceitar compor a banca. Agradeço a eles por essas oportunidades, que certamente fizeram diferença no meu percurso acadêmico.

Tive a sorte de fazer várias amizades ao longo da graduação, que renderam algumas discussões filosóficas edificantes (e algumas nem tanto). Destaco aqui Alex Bispo, Ariston Sócrates Guimarães, Daniel Diniz, Gabriel Eugênio, João Felipe Rasi, Mahan Vaz, Mariana Fernandes, Rafael de Souza e Sofia Abelha. Agradeço a eles pela companhia ao longo desses anos. Tive o *azar* de conhecer Renato Valvassori; não agradeço aos comentários maldosos que ele teceu contra minha pessoa e meu trabalho ao longo desses anos de amizade. Agradeço à Mariana também pela companhia – ainda que sonolenta – durante boa parte do tempo em que eu estava efetivamente escrevendo esta monografia.

A pesquisa que resultou nesta monografia foi possível devido ao financiamento da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo projeto nº 2019/12207-5. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do autor e não necessariamente refletem a visão da FAPESP.

Agradeço aos meus pais por, depois de um tempo, desistirem da ideia de me fazer abandonar a filosofia e me apoiarem fortemente desde então. Por último, mas não menos importante, tive a sorte de existir no mesmo mundo possível que uma instância do Bar do Ademar; o período da graduação teria sido bem menos divertido sem sua existência.

*[...] but they should have a virtue to shape
their life, amid the powers and chances of
the world, beyond the Music of the Ainur,
which is as fate to all things else [...]*
J. R. R. Tolkien, The Silmarillion

Conteúdo

Resumo	iii
Agradecimentos	iv
Epígrafe	v
Introdução	1
1 O Argumento Direto de van Inwagen	10
2 Críticas à Regra A	14
2.1 Responsabilidade e Veridadores	14
2.2 Um Contra-exemplo à Regra A	16
2.3 E as Instâncias Confirmatórias?	17
2.4 O Argumento Direto Lógico sem a Regra A	19
2.5 Considerações Finais	22
3 Críticas à Regra B	23
3.1 Um contra-exemplo à Regra B	23
3.1.1 A resposta de McKenna	25
3.2 Uma defesa da Regra B	26
3.3 Novamente, Instâncias Confirmatórias	27
3.3.1 Causação e a Regra B	28
3.4 Considerações Finais	36
Conclusão	38
Bibliografia	41

Introdução

Esta monografia tem como tema o Argumento Direto, um argumento pelo incompatibilismo – a tese de que responsabilidade moral é incompatível com a verdade do determinismo nomológico. Eu apresentarei alguns dos debates em torno desse argumento e defenderei que não há uma versão cogente do argumento. Mais especificamente, eu defenderei que não há uma formulação do Argumento Direto que seja (pelo menos) *prima facie* uma ameaça ao compatibilismo. Farei isso mostrando que, no momento, não há nenhuma formulação suficientemente motivada do argumento, de modo que ele não apresenta uma ameaça à posição adversária. Terminarei a monografia apontando alguns caminhos pelos quais o debate sobre o Argumento Direto pode avançar.

Antes de mergulhar em discussões específicas ao Argumento Direto, é importante ter claro na mente a terminologia relevante e como esse argumento se encaixa na discussão sobre o problema do determinismo. Por causa disso, antes de explicar a estrutura do texto, discutirei – nas próximas duas seções desta introdução – o problema do determinismo e o lugar do Argumento Direto nessa discussão.

O Problema do Determinismo

O problema do determinismo é o problema filosófico de estabelecer se responsabilidade moral é compatível com o determinismo e, caso não sejam, se temos responsabilidade moral no final das contas. Para obtermos clareza nessa discussão, temos que elucidar o que queremos dizer com “responsabilidade moral” e “determinismo”. Na próxima subseção, trato da primeira noção. Na outra, apresento a noção de “determinismo” e explico porque o determinismo é *prima facie* incompatível com responsabilidade moral.

Responsabilidade Moral

A expressão “ser responsável por” tem diversos significados. Às vezes, ela atribui um papel causal a uma entidade: “O tráfego foi responsável pelo nosso atraso”. Em outros casos, usamos a expressão para atribuir certos deveres a um indivíduo: “João era responsável por nos deixar lá na hora certa”. Além desses usos, dizemos coisas como “João é responsável por nosso atraso”. Essa sentença parece ter a mesma atribuição de papel causal que a primeira, mas ela vai além

disso. Se julgamos que João é responsável pelo atraso, temos toda uma gama de atitudes em relação a ele – como exigir uma desculpa ou sentir ressentimento – que não temos em relação ao tráfego, mesmo ambos gerando atrasos similares. João é responsável pelo atraso de um modo que só agentes, i.e., serem capazes de realizar ações, podem ser. A esse sentido de responsabilidade, dá-se o nome de “responsabilidade moral”.¹ Esse será o sentido de responsabilidade discutido no restante da monografia.

Antes de começar a análise filosoficamente mais rigorosa desse tipo de responsabilidade, é importante notar quão pervasivo esse fenômeno é em nossas vidas. Nós sentimos orgulho de nossas conquistas na medida em que nos sentimos responsáveis por elas. Acreditamos que devemos condenar criminosos somente se eles são responsáveis por seus crimes. Sentimos tristeza ou raiva quando coisas ruins acontecem conosco, mas há toda uma gama de emoções e atitudes que costumamos ter somente quando alguém é responsável por essas coisas ruins. Ações das pessoas que nos são queridas faz com que nos sintamos amados apenas quando cremos que essas ações não são meros acidentes, mas expressões de seus sentimentos. Quando alguém falha conosco mas não é responsável por isso, tendemos a desculpar a pessoa; se alguém é responsável por nos causar mal entretanto, nós adaptaremos nossa relação com ela de acordo. Não foi um exagero quando Peter Strawson afirmou que “na ausência de qualquer forma dessas atitudes, é duvidável que teríamos algo que poderíamos achar inteligível enquanto um sistema de relações humanas”² (Strawson, 1974, 26).

O ensaio *Freedom and Resentment* de Strawson (1962) reformulou o debate contemporâneo sobre responsabilidade moral. Ele chamou atenção para os sentimentos e atitudes que constituem nossas práticas de responsabilização, os quais apelidou de *atitudes reativas*. Ele listou como exemplos destas “gratidão, ressentimento, perdão, amor e sentimentos feridos” (Strawson, 1974, 5). Mesmo filósofos e filósofas que não aceitam suas teses mais controversas aceitam a visão geral de responsabilidade moral proposta por ele.

Para os fins desta monografia, não é preciso uma teoria detalhada de responsabilidade moral. Podemos, aceitando a compreensão strawsiana de nossas práticas como um pressuposto de fundo, assumir a seguinte caracterização de responsabilidade moral:

Responsabilidade Moral O agente *S* é moralmente responsável por *X* se, e somente se, há uma razão básica *pro tanto* para responder a *S* com um subconjunto de atitudes reativas em relação a *X*.³

¹Nesta introdução ao conceito de responsabilidade moral, supôs-se que há um sentido único para o termo. Uma posição popular na área, porém, é que há mais de um tipo de responsabilidade moral e, portanto, mesmo o termo “responsabilidade moral” é ambíguo. Encontramos essa posição, por exemplo, em McKenna (2012), Shoemaker (2015) e Mason (2019). Caso haja mais de um tipo de responsabilidade moral, o termo “responsabilidade moral” neste projeto designa o tipo de responsabilidade discutido na literatura sobre o problema do determinismo.

²Todas as citações apresentadas ao longo da monografia são traduções livres.

³O uso de “há uma razão” pode dar a impressão que estou me comprometendo com externalismo sobre razões. Isso, porém, não é o caso. Um tratamento internalista de razões morais poderia ser aplicado aqui. Para a distinção entre internalismo e externalismo sobre razões, ver Tiberius (2014, Capítulo 4).

Passemos por cada parte dessa caracterização. Primeiro, qual o domínio da variável ‘X’ nesse bicondicional? Isto é, sobre que tipo de entidades nós podemos ser moralmente responsáveis por? Dizemos que pessoas são responsáveis por um grupo bem diverso de entidades. Presumivelmente, pessoas são responsáveis por: ações, eventos, omissões, estados de coisas, condicionais, traços de caráter, crenças, desejos, fatos. Para os fins desta monografia, o domínio de ‘X’ é composto apenas de ações. Isso porque o interesse no debate sobre o problema do determinismo é em responsabilidade moral *direta*.⁴ A visão padrão é que os únicos objetos que temos responsabilidade moral direta por são ações; quando agimos satisfazendo certas condições, nos tornamos responsáveis pela nossa ação. Somos responsáveis por outros tipos de entidades quando elas têm uma determinada relação com ações pelas quais somos diretamente responsáveis; nisso consiste a responsabilidade moral *indireta*. Logo, dado o foco da monografia em responsabilidade direta, considerarei apenas responsabilidade por ações.

Por que responsabilidade nos dá uma razão básica *pro tanto* em vez de uma razão *simpliciter*? Uma razão *pro tanto* é uma que pode ser derrotada por outras razões. Uma razão levando-tudo-em-consideração é uma que não pode ser derrotada. Por exemplo, numa interpretação padrão de Kant, a ação *A* ser *contar uma mentira* nos dá uma razão levando-tudo-em-consideração para não realizar *A*. Por outro lado, para William Ross (1930), a ação *A* ser *contar uma mentira* nos dá apenas uma razão *pro tanto* para não realizar *A*.⁵ Em alguns casos, outras razões podem fazer com que mentir seja a ação certa a se tomar; se mentir pode salvar uma vida, a razão *pro tanto* para não mentir é derrotada. A razão na caracterização de responsabilidade moral é *pro tanto* porque alguém pode ser moralmente responsável por algo, mas termos uma razão para não responder com as atitudes reativas usuais. E.g., se você me dá um soco, pode ser melhor que eu não proteste – uma atitude reativa de culpabilização – porque eu poderia tomar uma surra como resposta. Outro exemplo: se uma criança faz algo perigoso e acaba se machucando, os pais podem decidir não dar uma bronca nela; afinal de contas, ela já “aprendeu a lição”.

Responsabilidade moral nos dá uma razão *básica* no sentido de não ser derivada de outras considerações normativas. Suponha que nós façamos um acordo: eu doarei muito dinheiro para os Médicos Sem Fronteiras caso você vá para seu vizinho e o culpe pelo tempo estar tão quente. Esse acordo faz com que você tenha uma razão moral *pro tanto* para culpar o seu vizinho. Ainda assim, isso não significa que ele seja – colocando de lado aqui a possível responsabilidade dele pelo aquecimento global – responsável pelo tempo estar tão quente. Se ele fosse responsável pelo tempo estar quente, você teria uma razão para culpá-lo independente de nosso acordo (e outras considerações consequencialistas ou contratualistas); em outras palavras, você teria uma razão básica para culpá-lo.

Quando alguém é responsável, temos uma razão para responder com um *subconjunto* de

⁴Para uma discussão mais aprofundada sobre a distinção entre responsabilidade direta e indireta, ver Mele (2021)

⁵Ross, na realidade, usa o termo “*prima facie*” para o que eu estou chamando de “*pro tanto*”. Como eu usarei “*prima facie*” ao longo da monografia com um sentido epistêmico, dei preferência pelo termo “*pro tanto*”.

atitudes reativas. Por que apenas um subconjunto? Porque nem toda atitude reativa é justificada em uma dada situação. Se você tenta queimar minha casa, eu tenho uma razão para ressentir você, mas não para sentir gratidão.

Devo deixar claro que aceitar esse bicondicional não me compromete com uma análise específica de responsabilidade moral. Essa caracterização é neutra sobre se atitudes reativas são mais fundamentais que ser moralmente responsável ou não. Em outras palavras, não me comprometo com alguma direção explicativa nesse bicondicional. A caracterização também é neutra quanto a merecimento.

Assumirei essa caracterização de responsabilidade moral porque algo em sua vizinhança é amplamente aceito na literatura. Por exemplo, encontramos esse bicondicional (ou algo similar) em McKenna (2012), Shoemaker (2015), Mason (2019), Pereboom (2014), Fischer & Ravizza (1998).

Ainda que esta monografia tenha uma abordagem mais abstrata à responsabilidade moral, essa caracterização será útil. Ao longo do texto, afirmarei coisas como “é intuitivo que ela é responsável por isso” ou “claramente ninguém é responsável por isso”. Creio que esses apelos a intuições são justificados na medida em que temos um bom entendimento de quando atitudes reativas são apropriadas. Por causa disso que é importante explicitar a relação entre responsabilidade moral e atitudes reativas.

Talvez essa caracterização seja insatisfatória para alguns. Esses podem achar que conceitos como “dever” ou “merecimento” são essenciais para responsabilidade moral. Alternativamente, podem achar que responsabilidade moral é sobre consequências e não razões básicas. Infelizmente, uma caracterização mais geral sobre responsabilidade moral pode não estar disponível. É plausível que diferentes fenômenos são estudados sobre o rótulo de “responsabilidade moral” e, nesse caso, é preciso fazer certos recortes teóricos que deixarão alguns debates de lado (Shoemaker, 2020).

Nesta subseção, eu apresentei – tentando manter o mínimo de comprometimentos teóricos – o conceito de responsabilidade moral. Como vimos, responsabilidade moral está fortemente associado a atitudes reativas. Conforme o debate for se tornando mais abstrato ao longo da monografia, vale a pena manter em mente *o que* está em jogo: nosso sistema de responsabilização e todas as atitudes que o compõe.

A Ameaça do Determinismo

Na subseção anterior, apresentei o conceito de responsabilidade moral e frisei o papel que as práticas de responsabilização têm em nosso dia-a-dia. Essas práticas são tão basilares em nossas vidas pessoais e relacionamentos interpessoais que a tese de que nunca somos moralmente responsáveis por nossas ações parece, em um primeiro momento, extremamente revisionista. Dado isso, que motivos poderíamos ter para seriamente considerar que essa tese é o caso? Segundo algumas filólogas e filósofos, temos um montão de motivos: o fato de não sermos substâncias

não-físicas, as influências que nosso ambiente social têm em nosso caráter, a impossibilidade de escolhermos nosso caráter *ex nihilo*, a possibilidade de Deus saber o que faremos antes mesmo de agirmos, a existência de proposições verdadeiras sobre ações futuras... Entretanto, entre todas as ameaças levantadas contra nossa visão de que (às vezes) somos moralmente responsáveis, aquela que se destaca por sua centralidade na área de metafísica da responsabilidade é a do determinismo nomológico. Nesta subseção, eu explicarei o que é a tese do determinismo nomológico e porque, aparentemente, se ela é o caso, nós nunca somos moralmente responsáveis por nossas ações.

Vale a pena introduzir a noção de determinismo com um exemplo mundano. Alex é um exímio jogador de sinuca. Sabendo a distribuição das bolas na mesa, a qualidade da mesa, a força e a direção da tacada, ele sabe exatamente o estado que a mesa vai ficar depois da tacada. Isso não é possível unicamente por causa de sua habilidade; se toda a vez que as bolas trombassem houvesse 50% de chance delas não se moverem, Alex não conseguiria prever a configuração final da mesa. Ou seja, há um fato sobre sinuca – independente do Alex – que faz com que, dado um estado da mesa e uma determinada tacada, seja possível apenas um estado final da mesa. Diríamos então que o sistema da mesa de sinuca é *determinístico*. O universo como um todo pode ser um sistema determinístico; dado um estado do universo, as leis da natureza fixam um único estado futuro. É debatível no que exatamente consiste esse “fixar”. Uma definição mais precisa de determinismo e que é usual na literatura é a seguinte:

Determinismo Um mundo m é determinístico se, e somente se, para qualquer tempo t , uma descrição completa dos fatos não-relacionais sobre t em conjunção com uma descrição completa das leis da natureza acarreta qualquer fato sobre m .⁶

Antes de motivar a incompatibilidade entre determinismo e responsabilidade moral, algumas elucidaciones quanto a *Determinismo*. Primeiro, ainda que falemos de *leis* determinísticas, no debate sobre o problema do determinismo, estamos interessados em *mundos* determinísticos. As leis da natureza podem ser determinísticas, mas o mundo não — no caso, por exemplo, das leis não cobrirem todos os eventos ou permitirem exceções. Para que um mundo seja determinístico, é preciso que as leis (i) sejam determinísticas, (ii) cubram todos os eventos que ocorrem e (iii) não permitam exceções. Segundo, é importante que a descrição de m em t seja apenas de fatos *não-relacionais* sobre t . Isso não significa que fatos como <Maria é irmã de Ana> não estejam inclusos na descrição; esse é um fato relacional sobre Maria e sobre Ana, mas não sobre t . Os fatos relacionais excluídos da descrição são aqueles que relacionam t com outros instantes temporais; tal como <No futuro, Maria decide tomar café>. Se não fizessemos essa restrição, o problema do determinismo nomológico seria equivalente ao do determinismo lógico – o problema de saber se responsabilidade é compatível com a verdade de proposições

⁶Encontramos formulações de determinismo equivalentes a *Determinismo* em van Inwagen (1983, 65), Vihvelin (2013, 1) e McKenna (2008, 351). Definições similares também são encontradas em Russell (1913, 18) e Earman (1986, 13).

sobre o futuro –. Por fim, definimos acarretamento do seguinte modo: *A* acarreta *B* se, e somente se, necessariamente, *A* implica *B*.

É importante ter claro em mente como determinismo se relaciona com previsibilidade e inevitabilidade. A verdade do determinismo (e sua ameaça à nossa agência) é independente da possibilidade de previsibilidade de nossas ações. Mesmo se o determinismo for verdadeiro, pode não ser possível saber uma descrição completa de um instante temporal e derivar fatos futuros a partir dessa descrição e as leis da natureza. O determinismo também não implica que todos os acontecimentos são inevitáveis; pelo menos não no sentido de, *independente do que fizermos*, isso acontecerá. Ainda que o determinismo seja verdadeiro, há eventos que não ocorreriam se agíssemos de outro modo. O que é verdade dado o determinismo é que outros eventos não são possíveis *mantendo o mesmo passado e leis da natureza*. A tese de que todos os acontecimentos são inevitáveis, comumente chamada de fatalismo, não parece derivar do determinismo nomológico, mas sim do determinismo lógico, apresentado no parágrafo anterior (Rice, 2018).

Por fim, antes de passarmos ao problema do determinismo propriamente dito, vale a pena considerar uma objeção à empreitada de discutir esse problema. O debate é sobre as possíveis consequências do determinismo para nossa responsabilidade. Logo, se tivermos bons motivos para crer que o determinismo é falso, diz o objetor, não há porque ter esse debate.⁷ E, segundo o proponente dessa objeção, nós temos uma boa razão para crer que o determinismo é falso; a mecânica quântica é indeterminística. Mais precisamente, a objeção é a seguinte: a mecânica quântica padrão é uma teoria essencialmente probabilística; com o postulado do colapso, ela oferece uma dinâmica fundamentalmente indeterminista dos processos quânticos. Teríamos então uma razão científica para rejeitar a tese do determinismo e, portanto, o problema do determinismo não seria relevante para estabelecermos se temos ou não responsabilidade.

Contra essa objeção, é importante ressaltar dois pontos. Primeiro, não temos razões suficientes para aceitar que a mecânica quântica padrão é verdadeira. Embora seja uma teoria empiricamente adequada, há alternativas a ela que são compatíveis com o determinismo. Por exemplo, a teoria da onda piloto (Goldstein, 2021) é uma teoria quântica determinística e empiricamente adequada. Fora isso, a mecânica quântica padrão pode ser interpretada de maneira determinística, como na interpretação super-determinística (Andreoletti & Vervoort, 2022). Portanto, é uma questão em aberto se os fenômenos quânticos realmente nos obrigam a aceitar o indeterminismo. Segundo, mesmo se o mundo atual for indeterminístico, caso o indeterminismo da física quântica não gere indeterminismo no nível das ações humanas, a ameaça do determinismo à nossa responsabilidade persiste.

Há dois modos de desenvolver a aparente incompatibilidade entre determinismo e responsabilidade moral. Esses dois modos rastreiam duas compreensões distintas que podemos ter da

⁷Vale frisar que essa objeção defende que nossa responsabilidade está a salvo da ameaça do *determinismo*, não que temos responsabilidade no final das contas. O indeterminismo também oferece uma ameaça à agência responsável; mesmo se aceitarmos o indeterminismo, ainda é uma questão em aberto se somos responsáveis.

agência responsável (Timpe, 2016).

Por um lado, podemos focar na ideia de alternativa. Intuitivamente, uma pessoa só é responsável por agir de um modo se ela poderia agir de outro modo (mais sobre isso na próxima seção). Se nossas ações são consequência do passado e das leis da natureza – coisas fora do nosso controle –, nós não podemos agir de outro modo e, portanto, não somos responsáveis. De maneira rudimentar, a ideia aqui seria que não temos alternativas se o determinismo é verdadeiro porque o passado e as leis já “decidiram” nossa escolha.

Por outro lado, podemos focar na ideia de origem. Parece que somos responsáveis apenas pelas ações que se originam em nós. Se o determinismo for verdadeiro, porém, o passado e as leis da natureza são a origem de nossas ações e, portanto, não somos responsáveis. Quando cogitamos a verdade do determinismo, ficamos com a impressão de que nossas ações seriam meros desdobramentos do passado e das leis, aos quais nada contribuiríamos; não seríamos, portanto, origens de nossas ações.

Há duas posições principais nesse debate. O compatibilista crê que, apesar dessa incompatibilidade *prima facie*, o determinismo nomológico é compatível com sermos responsáveis por nossas ações. Já o incompatibilista crê que essa incompatibilidade não é mera aparência; o determinismo nomológico realmente é incompatível com sermos responsáveis por nossas ações. Entre os incompatibilistas, encontramos os libertistas, que creem que o determinismo é falso e temos responsabilidade, e os incompatibilistas radicais, que creem que o determinismo é incompatível com responsabilidade e que temos boas razões para achar que não somos responsáveis no final das contas.

Onde o Argumento Direto se encaixa nisso?

Na seção anterior foi apresentado superficialmente o problema do determinismo. Nesta seção, explicarei como o Argumento Direto se encaixa nesse debate.

Anteriormente, apresentei dois modos de desenvolver a incompatibilidade *prima facie* entre determinismo e responsabilidade moral. Uma focada na ideia de alternativas e outra na ideia de origem. Essas duas visões da agência responsável podem parecer complementares; para que realmente possamos escolher entre alternativas, devemos ser a origem dessa escolha, e para que possamos ser a origem de uma ação, nós precisamos selecioná-la entre alternativas. Se isso for o caso, o compatibilista precisaria mostrar que esses dois aspectos da agência responsável estão a salvo da aparente ameaça do determinismo.

Muitos membros do debate, porém, estão convencidos de que alternativas não são necessárias para a responsabilidade moral. A principal motivação para rejeitar a relevância de alternativas são os casos de Frankfurt. Primeiramente propostos por Henry Frankfurt (1969), esses casos nos apresentam um agente que intuitivamente não tem alternativa, mas ainda assim é responsável por sua ação. Segue um caso de Frankfurt:

Chip na Cabeça Gabriel está planejando matar Renato. João também deseja que Renato morra, mas preferiria não ter que sujar as próprias mãos; se Gabriel matá-lo no final das contas, melhor para João. Por causa disso, João, que é um grande neurocientista, instala um chip na cabeça de Gabriel de modo que, se ele hesitar em matar Renato, João vai apertar um botão que ativará o chip e fará Gabriel cometer o assassinato. No final das contas, porém, João não precisa intervir; Gabriel mata Renato por suas próprias razões.

Nesse cenário, Gabriel não tem alternativa além de matar Renato. Ainda assim, intuitivamente Gabriel é moralmente responsável pelo assassinato. Esse cenário, então, parece mostrar que alternativas não são necessárias para ser moralmente responsável. Os casos de Frankfurt geraram uma imensa literatura.⁸ Os pormenores dessa discussão não serão relevantes para o tema desta monografia; o que nos interessa é a cissão que os casos de Frankfurt provocaram na discussão sobre o problema do determinismo.

Aqueles que não foram convencidos pelos casos de Frankfurt continuaram a debater se o determinismo impede que tenhamos alternativas. As duas posições nessa discussão são o incompatibilismo de alternativas – alternativas são relevantes para responsabilidade e incompatíveis com o determinismo – e compatibilismo de alternativas – alternativas são relevantes para responsabilidade e compatíveis com o determinismo.

O principal argumento pelo incompatibilismo de alternativas é o Argumento da Consequência. A formulação mais famosa desse argumento é aquela apresentada por Peter van Inwagen (1983). Uma formulação simples do argumento – bem diferente das formulações mais técnicas em que ele é realmente discutido – é a seguinte:

Argumento da Consequência Nós não temos escolha pelo passado e pelas leis da natureza.

Se o determinismo for verdadeiro, nossas ações são consequência do passado e das leis da natureza. Portanto, se o determinismo for verdadeiro, não temos escolha por nossas ações.

O argumento, por si só, conclui apenas que não temos escolha se o determinismo for verdadeiro. Para disso concluirmos que não temos responsabilidade moral se o determinismo for verdadeiro, devemos aceitar que a noção de “escolha” – no sentido de alternativas – é relevante para responsabilidade moral. Aqueles que foram convencidos pelos casos de Frankfurt, porém, negam isso. Sendo assim, o *Argumento da Consequência* não apresenta uma ameaça ao *compatibilismo de origem*, a posição que defende que o relevante para a responsabilidade é ser a origem de sua ação e isso não é ameaçado pelo determinismo.

Aqui que o Argumento Direto entra em cena. Van Inwagen crê que a ideia central do *Argumento da Consequência* pode ser apresentada sem o intermédio da noção de alternativa; em outras palavras, diretamente. Teríamos então um argumento incompatibilista que se opõe ao compatibilismo de origem. Esse argumento é o Argumento Direto, que pode ser apresentado de maneira simples como se segue:

⁸Conferir Sartorio (2017) para uma introdução ao debate sobre os casos de Frankfurt.

Argumento Direto Nós não somos responsáveis pelo passado e pelas leis da natureza. Se o determinismo for verdadeiro, nossas ações são consequência do passado e das leis da natureza. Portanto, se o determinismo for verdadeiro, não somos responsáveis por nossas ações.

Ao longo desta monografia, será exposto o debate que surgiu em torno desse argumento (em uma formulação mais técnica proposta por van Inwagen). Nesta seção, foi exposto onde o argumento se encaixa na discussão do problema do determinismo; O Argumento Direto é um argumento pelo incompatibilismo que tem como adversário o compatibilista de origem.

Escopo e Percurso

Nesta seção, explicarei mais precisamente qual o foco e a estrutura da monografia. A literatura sobre o Argumento Direto é extensa. Há, pelo menos, seis formulações do argumento, cada uma com suas nuances próprias. Costuma-se dividir as formulações entre Argumentos Diretos Lógicos (entram aqui a formulação de van Inwagen, a de Ted Warfield (1996) e a de Justin Capes (2016)) e Argumentos Diretos Causais (as três apresentadas por Michael McKenna (2001; 2008)). Argumentos Diretos Lógicos têm esse nome por entenderem “consequência” na formulação do *Argumento Direto* como consequência lógica, já os Argumentos Diretos Causais entendem como consequência causal. A monografia focará na formulação de van Inwagen. Isso por dois motivos: (i) a maior parte das críticas apresentadas a ela também vale para outros Argumentos Diretos Lógicos e (ii) discutir os Argumentos Diretos Causais de maneira satisfatória exigiria discussões sobre causação que seriam extensas demais para esta monografia (mas aguardem minha dissertação de mestrado!). Ainda assim, meu argumento geral de que o Argumento Direto não é suficientemente motivado valerá para todas as formulações; diferentes críticas que apresentarei valerão para diferentes formulações – como apontarei nos momentos apropriados –, mas nenhuma das formulações na literatura escapa de todas as críticas.

No primeiro capítulo, eu reconstruirei o Argumento Direto proposto por Peter van Inwagen. O argumento se vale de duas regras dedutivas para um operador modal de não-responsabilidade. Os capítulos dois e três apresentam debates que surgiram em torno dessas duas regras. Eu defenderei que as objeções apresentadas contra o Argumento Direto nessas duas seções são suficientes para mostrar que o argumento é nem *prima facie* uma ameaça ao compatibilismo. Por fim, na conclusão, apontarei alguns caminhos pelos quais creio que o debate pode avançar de maneira proveitosa.

1. O Argumento Direto de van Inwagen

Na introdução, eu apresentei o Argumento Direto da seguinte maneira:

Argumento Direto Nós não somos responsáveis pelo passado e pelas leis da natureza. Se o determinismo for verdadeiro, nossas ações são consequência do passado e das leis da natureza. Portanto, se o determinismo for verdadeiro, não somos responsáveis por nossas ações.

Colocando de maneira esquemática, temos o seguinte argumento:

P1 Não somos responsáveis pelo passado e pelas leis da natureza; (Não-Responsabilidade pelo passado e leis)

P2 Nossas ações são consequência do passado e das leis da natureza; (Determinismo)

C1 Não somos responsáveis por nossas ações; (De P1 e P2)

Dois problemas saltam aos olhos nesse argumento. Primeiro, “consequência” aqui parece mais vago que o desejável; qual exatamente é a relação entre o passado, as leis e nossas ações dado a verdade do determinismo? Segundo, assumindo apenas as regras de inferência da lógica clássica, o argumento é inválido.

Dado esses problemas, o proponente do Argumento Direto precisa fornecer uma versão mais precisa do Argumento Direto e que tenhamos motivação para crer que é válida. Uma formulação do argumento que parece resolver esses problemas é aquela proposta por Peter van Inwagen (1983), sendo considerada a versão original do argumento. Neste capítulo, apresentarei essa formulação do Argumento Direto.

Van Inwagen, baseando-se em sua formulação modal do Argumento da Consequência, propõe uma formulação modal do Argumento Direto. Para formulá-lo, primeiro definimos um operador modal de não-responsabilidade. Seja P uma proposição, temos o seguinte operador:

(NR) $NRP \equiv P$ e ninguém é, nem nunca foi, nem parcialmente, moralmente responsável por P .

Para esse operador, van Inwagen propõe duas regras dedutivas. Sendo P e Q proposições e ‘ \Box ’ representando necessidade metafísica, as duas regras dedutivas para NR são as que se seguem:

Regra A $\Box P \vDash NRP$

Regra B $NRP, NR(P \supset Q) \vDash NRQ$

A Regra A estabelece que ninguém é, nem parcialmente, moralmente responsável por uma verdade necessária; a Regra B é uma regra dedutiva segundo a qual não-responsabilidade é fechada sob implicação material.

Com essas regras dedutivas, podemos formular formalmente o Argumento Direto. Seja P_0 uma proposição descrevendo o estado completo do mundo em um momento do passado remoto¹, L uma proposição descrevendo as leis da natureza e P qualquer proposição verdadeira, o Argumento Direto é formulado como se segue:

Argumento Direto Original		
P1	$\Box((P_0 \wedge L) \supset P)$	Determinismo
P2	NRP_0	Não-Responsabilidade pelo passado
P3	NRL	Não-Responsabilidade pelas leis
C1	$\Box(P_0 \supset (L \supset P))$	De P1
C2	$NR(P_0 \supset (L \supset P))$	De C1 e Regra A
C3	$NR(L \supset P)$	De C2, P2 e Regra B
C4	NRP	De C3, P3 e Regra B

Analisemos esse argumento por partes, começando com o operador NR . O operador é definido em termos de responsabilidade sobre proposições, mas *Responsabilidade Moral* está em termos de responsabilidade sobre ações. É preciso então de uma ligação entre responsabilidade sobre ações e responsabilidade sobre proposições. O próprio van Inwagen levanta essa questão ao apresentar o *Argumento da Consequência*:

Determinismo é uma tese sobre proposições, mas a tese do livre-arbítrio é uma tese sobre agentes. Se iremos investigar as relações conceituais sobre essas duas teses, faremos bem em formular a tese do livre-arbítrio como uma tese sobre agentes e proposições (van Inwagen, 1983, 65-66)

Ao discutir o *Argumento da Consequência*, van Inwagen faz essa ligação usando a noção técnica de *capacidade para falsear uma proposição* (66-68). Essa noção, porém, está intimamente ligada com a visão de alternativas da agência responsável. Tendo em mente que o Argumento Direto é um argumento contra o compatibilismo de origem, essa noção não servirá. No próximo capítulo, apresentarei uma tese que liga responsabilidade sobre ações com responsabilidade sobre proposições. Por enquanto, fiquemos apenas com a ideia intuitiva de responsabilidade sobre proposições.

¹“Passado remoto” é, nessa formulação, um termo técnico que significa um instante temporal fora do controle de qualquer agente. Isso é relevante para evitar algumas objeções apresentadas contra Argumento da Consequência – e que se estendem ao Argumento Direto – por Joseph Campbell (2007; 2010).

Feita essa discussão sobre o operador, passemos às premissas do argumento. P1 nada mais é que uma reformulação de *Determinismo*. Usando essa formulação para a tese do determinismo, van Inwagen especifica em que sentido nossas ações são consequência do passado e das leis caso o determinismo seja verdadeiro: uma descrição do passado em conjunção com uma descrição das leis acarreta todos os fatos verdadeiros sobre nossas ações.

P2 e P3 afirmam, respectivamente, que ninguém é, ou foi, nem parcialmente, moralmente responsável pelo passado remoto e que ninguém é, ou foi, nem parcialmente, moralmente responsável pelas leis da natureza. Essas premissas parecem difíceis de negar; como poderíamos culpar alguém pelo Big Bang ou pela segunda lei da termodinâmica? As premissas análogas presentes no *Argumento da Consequência* foram colocadas em xeque por diversos autores², mas seus contra-argumentos não foram estendidos ao Argumento Direto.

Dada a forte plausibilidade das premissas, parece que a esperança do compatibilista é argumentar contra a validade do argumento. Encontramos na literatura críticas tanto à Regra A quanto à Regra B, as quais serão apresentadas nos dois próximos capítulos. Antes de apresentar essas críticas, porém, é importante considerar as motivações para aceitar essas regras para começo de conversa.

Após apresentar a Regra A, van Inwagen diz:

A validade da Regra (A) me parece ser indisputável. Ninguém é responsável pelo fato que $49 \times 18 = 882$, pelo fato que a aritmética é essencialmente incompleta, ou, se Kripke está certo sobre verdade necessárias, pelo fato que o número atômico do ouro é 79. (184)

Como veremos no próximo capítulo, há quem não concorde com van Inwagen que a Regra A é indisputável. O mais importante dessa citação é a presença de putativas *instâncias confirmatórias* da Regra A. No debate sobre o Argumento Direto, uma instância confirmatória da regra X é uma inferência cuja validade intuitiva dá suporte à validade de X . Não faria sentido van Inwagen apresentar esses exemplos para argumentar indutivamente pela Regra A; uma indução a partir de três casos está longe de ser um bom argumento. Esses exemplos motivam a inferência da nossa não-responsabilidade por um fato a partir da necessidade desse fato. Em outras palavras, já que esses fatos são necessários, não somos responsáveis por eles. Se essas inferências são intuitivamente válidas, temos um motivo para aceitar que a Regra A é intuitivamente válida.

Instâncias confirmatórias são especialmente relevantes na discussão sobre a Regra B. Van Inwagen apresenta duas motivações a favor da Regra B. Primeiro, ele ressalta que a Regra B é válida se afirmar que $NR(P)$ é equivalente a afirmar que P é verdadeiro em uma classe de mundos possíveis. Em outras palavras, a Regra B é válida se NR funcionar como um operador de necessidade em uma lógica modal igual ou mais forte que K . Depois de apontar isso, van Inwagen diz que provavelmente deve haver tal classe de mundos possíveis (ainda que ele não

²Ver, por exemplo, Lewis (1981), Beebe & Mele (2002), Tognazzini & Fischer (2017)) e Esfeld (2019).

consiga delimitar qual é essa classe). Ele mesmo aponta, porém, que esse argumento é pouco convincente (186-187).

Resta, então, a segunda motivação a favor da Regra B: duas instâncias confirmatórias (187). As duas instâncias, apelidadas respectivamente de *Mordida de Cobra* e *Platão*, são as seguintes:

MC1 *NR*(John é mordido por uma cobra em seu aniversário de 30 anos);

MC2 *NR*(John é mordido por uma cobra em seu aniversário de 30 anos \supset John morre no seu aniversário de 30 anos);

MC3 *NR*(John morre no seu aniversário de 30 anos); (De SB1 e SB2 pela Regra B)

E

PT1 *NR*(Platão morreu na antiguidade);

PT2 *NR*(Platão morreu na antiguidade \supset Platão nunca conheceu Hume);

PT3 *NR*(Platão nunca conheceu Hume); (De PT1 e PT2 pela Regra B)

Segundo van Inwagen, essas duas inferências são intuitivamente verdadeiras, o que nos dá uma razão para crer que a Regra B é válida.

Apresentei, então, as motivações pelas premissas e pela validade do argumento. Se o argumento for sólido, ninguém é, nem nunca foi, mesmo parcialmente, moralmente responsável pela verdade de uma proposição verdadeira qualquer caso o determinismo seja verdadeiro. Generalizando esse argumento, se o determinismo for o caso, não somos responsáveis por fato algum; em outras palavras, responsabilidade moral é incompatível com o determinismo.

Após apresentar o Argumento Direto e motivá-lo, van Inwagen diz ver duas opções para o compatibilista: (i) encontrar um conjunto de proposições mais plausíveis que a validade da Regra B e que impliquem na compatibilidade entre responsabilidade e determinismo, ou (ii) apresentar um contra-exemplo à Regra B que possa ser avaliado independentemente do problema do determinismo. Como veremos nos dois próximos capítulos, os compatibilistas foram mais inventivos que isso; encontramos, além de contra-exemplos à Regra B, objeções à Regra A e argumentos de cunho dialético.

2. Críticas à Regra A¹

No capítulo anterior, apresentei a formulação do Argumento Direto proposta por Peter van Inwagen. No presente capítulo, apresentarei as críticas contra a Regra A, uma das regras dedutivas do argumento. Outros Argumentos Diretos Lógicos que encontramos na literatura, como o de Ted Warfield (1996) e o de Justin Capes (2016), também dependem da validade da Regra A. As críticas que apresentarei aqui, portanto, se estendem a essas formulações.

Este capítulo se divide em cinco seções. Na primeira, eu desenvolverei, seguindo Merluzzi & Maruchi (2019), a relação entre responsabilidade e veridadores. Na segunda, apresentarei um contra-exemplo à Regra A encontrado na literatura e como ele se relaciona com o que terá sido apresentado na primeira seção. Na terceira, discutirei as instâncias confirmatórias para a Regra A apresentadas por van Inwagen seguindo a discussão presente em Hermes (2014). Na quarta, discuto a proposta de Justin Capes (2016) de reformular o Argumento Direto de modo a não usar a Regra A. Na quinta seção, faço algumas considerações finais sobre os debates em torno da Regra A.

2.1 Responsabilidade e Veridadores

Uma questão que ficou em aberto no capítulo anterior é como responsabilidade por ações se relaciona com responsabilidade por proposições. Nesta seção, eu apresentarei uma proposta sobre essa relação. Ao longo do capítulo, veremos como essa proposta gera problemas para o Argumento Direto apresentado no capítulo anterior.

No debate sobre o *Argumento da Consequência*, a ideia de escolha sobre uma proposição é explicada a partir da noção da capacidade para falsear uma proposição.² Teríamos, por exemplo, algo como:

Escolha O agente S tem escolha pela proposição P se, e somente se, S é capaz de fazer uma ação A , tal que, se S fizesse A , P seria falso;³

¹Parte considerável deste capítulo, em especial a primeira e a segunda seção, foram retiradas de Merluzzi & Maruchi (2019).

²A noção é introduzida no debate por van Inwagen e é discutida, por exemplo, em Huemer (2000), Pruss (2013), Schnieder (2004) e Lampert & Merluzzi (2021a).

³Essa definição é uma versão simplificada daquela apresentada em Pruss (2013).

Poderíamos tentar transpor *Escolha* para o debate sobre o Argumento Direto. Em vez de falar sobre “ter escolha”, falaríamos sobre “ser responsável”:

R-Capacidade O agente *S* é responsável pela proposição *P* somente se *S* é capaz de fazer uma ação *A*, tal que, se *S* fizesse *A*, *P* seria falso;

Um leitor atento perceberá q *R-Capacidade* fornece apenas uma condição necessária, mas não suficiente, para ser responsável por uma proposição. Isso porque – segundo a maior parte da literatura – responsabilidade moral tem duas condições independentes. Uma é uma condição metafísica: o agente precisa ter um certo controle sobre sua ação para ser responsável; a outra uma condição epistêmica: o agente deve saber certas coisas sobre sua ação para ser responsável. A proposta de análise apresenta foca apenas na condição metafísica e, portanto, oferece apenas uma condição necessária.

Feito esse esclarecimento, voltemos a *R-Capacidade*. Dado a dialética do debate, esse princípio não pode ser usado a favor do Argumento Direto. Vale lembrar que o oponente do Argumento Direto é um compatibilista *de origem*; ele aceita, por exemplo, que Gabriel é responsável em *Chip na Cabeça* pelo fato de que Renato morreu. Gabriel, porém, não satisfaz *R-Capacidade* nesse cenário. Essa análise está tentando capturar a noção de ter alternativa, o que não é relevante no debate sobre o Argumento Direto.

Que análise poderíamos ter então? Parece que Gabriel é responsável pelo fato de que Renato morreu em *Chip na Cabeça* porque a ação de Gabriel (pela qual ele é responsável) fornece uma explicação para a verdade do fato em questão. Um modo de capturar essa ligação explicativa entre a ação do agente e a verdade de uma proposição é através da noção de *verificador*. Um verificador é a entidade que torna verdadeira uma proposição. Por exemplo, a proposição <A bola é vermelha> tem como verificador a bola instanciando vermelhidão. Pode-se pensar também que o verificador para uma proposição *P* é a entidade com que a verdade de *P* está ontologicamente comprometida (Asay, 2018). Para que <A bola é vermelha> seja verdadeira, deve existir a bola instanciando vermelhidão.

Usando a noção de verificador, podemos propor a seguinte análise parcial de *responsabilidade por uma proposição*:

R-Verificador *S* é, pelo menos parcialmente, moralmente responsável pelo fato de que *P* somente se o verificador de *P* é uma ação livre de *S* ou é causado por uma ação livre de *S*.

Em *R-Verificador*, há o termo “ação livre”, pelo qual quero dizer uma ação na qual o agente satisfaz a condição metafísica da responsabilidade moral. Essa análise é dialeticamente apropriada; ela é *prima facie* aceitável tanto pelo compatibilista de origem quanto pelo proponente do Argumento Direto. Diferente de *R-Capacidade*, ela consegue explicar porque Gabriel é responsável em *Chip na Cabeça*.

Apresentei nesta seção uma análise parcial de *ser responsável por uma proposição*. Ainda que a maior parte da literatura trate dessa noção de maneira apenas intuitiva, uma análise dela pode desempenhar um papel importante na discussão: se a melhor análise implicar que o Argumento Direto é sólido, temos uma boa defesa do argumento; se implicar que ele não é sólido, temos um bom contra-argumento. No restante deste capítulo, veremos como *R-Verificador* põe em xeque a validade da Regra A. No capítulo seguinte, veremos como a análise pode ser usada para defender a Regra B.

2.2 Um Contra-exemplo à Regra A

Nesta seção, apresentarei um contra-exemplo à Regra A que encontramos na literatura. Não apresentarei todos os contra-exemplos, nem me estenderei na discussão que eles geraram. O ponto central desta seção não é mostrar que a Regra A é inválida, mas antes motivar duas teses: (i) *R-Verificador* nos dá uma razão para rejeitar a Regra A e (ii), *pace* van Inwagen, a validade da Regra A está longe de ser indisputável.

O contra-exemplo em questão foi apresentado por Charles Hermes (2014). O cenário para o contra-exemplo é o seguinte:

Assassinato Mary, satisfazendo as condições para ser moralmente responsável por sua ação, matou Matt;

Nenhum experimento mental muito elaborado aqui. Como esse cenário é um contra-exemplo à Regra A? Segundo Hermes (2013; 2014), o seguinte princípio é verdadeiro:

(RD) Se um agente é responsável pelo único disjuncto verdadeiro de uma disjunção, ele é responsável pela disjunção;

Assumindo *RD*, temos o seguinte argumento: Mary é responsável pelo fato de que *Mary matou Matt*; o único disjuncto verdadeiro na disjunção *Mary matou Matt ou não é Mary matou Matt*; assim, dado *RD*, Mary é responsável pelo fato de que *Mary matou Matt ou não*. Essa disjunção, entretanto, é apenas uma instância do terceiro excluído e, logo, necessariamente verdadeiro. Se o argumento é sólido, um agente pode ser responsável por uma verdade necessária e, portanto, a Regra A é inválida.

Uma peça central nesse contra-exemplo é *RD*. Que motivação temos para aceitar esse princípio? Um motivo para crer em *RD* é que ele se segue de *R-Verificador*. Assuma que um agente é responsável por *P*. Dado *R-Verificador*, isso implica que uma ação livre do agente causa o verificador de *P* obter. Agora considere a disjunção $P \vee Q$ onde *Q* é uma proposição falsa arbitrária. O verificador dessa disjunção é o mesmo que o verificador de *P*. Dado que uma ação livre do agente causou o verificador de *P* obter, sua ação também causou o verificador de $P \vee Q$ obter. Ou seja, ele é responsável por $P \vee Q$ (dado *R-Verificador*). Concluimos, portanto, que *RD* é verdadeiro se assumimos *R-Verificador*.

Se *R-Verificador* é a melhor análise que temos para responsabilidade por proposições, então temos uma razão para crer que podemos ser responsáveis por verdades necessárias (e, consequentemente, que a Regra A é inválida.). O *modus ponens* de um, porém, é o *modus tollens* de outro; alguém pode considerar que responsabilidade por verdades necessárias é algo tão absurdo que, na realidade, esse exemplo nos dá é uma razão para rejeitar *R-Verificador*. Ainda que isso possa ser o caso, enquanto o defensor da Regra A não propor uma alternativa a *R-Verificador*, essa é a melhor análise que temos, o que coloca em dúvida a plausibilidade da Regra A.

A discussão sobre a Regra A é extensa. Michael Robinson (2016) apresenta um contra-argumento a *RD*, Pedro Merluzzi & Gabriel Maruchi (2019) respondem ao contra-argumento. Stephen Kearns (2011) apresenta vários contra-exemplos à Regra A, os quais são respondidos por P. Roger Turner & Justin Capes (2018). Fábio Lampert & Pedro Merluzzi (2021b) contra-argumentaram a Turner & Capes. Fora isso, a discussão sobre a capacidade para fazer o impossível, presente em Spencer (2017) e Nguyen (2020), também é relevante para esse debate.

Não me aprofundarei nessa discussão. O contra-exemplo apresentado é suficiente para mostrar o que eu pretendia nesta seção: (i) *R-Verificador* nos dá uma razão para rejeitar a Regra A e (ii), *pace* van Inwagen, a validade da Regra A está longe de ser indisputável.

2.3 E as Instâncias Confirmatórias?

Van Inwagen tinha apresentado duas motivações para a Regra A: sua aparente indisputabilidade e algumas instâncias confirmatórias. Na seção anterior, vimos um putativo contra-exemplo à Regra A, o que é suficiente para rejeitarmos sua aparente indisputabilidade. Nessa seção, eu discutirei brevemente as instâncias confirmatórias apresentadas por van Inwagen.

Para motivar a Regra A, van Inwagen apresenta alguns exemplos de fatos necessários pelos quais não somos responsáveis. Para que esses exemplos sirvam de motivação à Regra A, é preciso que a inferência de $\Box P$ para *NRP* seja intuitiva para os fatos em questão. Temos então três pretensas instâncias confirmatórias para a Regra A:

$$\frac{\Box(49 \times 18 = 882)}{NR(49 \times 18 = 882)}$$

$$\frac{\Box(\text{a aritmética é essencialmente incompleta})}{NR(\text{a aritmética é essencialmente incompleta})}$$

$$\frac{\Box(\text{o número atômico do ouro é } 79)}{NR(\text{o número atômico do ouro é } 79)}$$

Um defensor da Regra A pode apelar a esses casos para defender que eles nos dão motivação

para aceitar a regra e, já que o debate sobre os contra-exemplos ainda está em aberto, essas instâncias confirmatórias são suficientes para aceitar a Regra A.

Charles Hermes (2014) contra-argumenta ao uso dessas instâncias confirmatórias para motivar a Regra A. Segundo ele, temos uma explicação melhor para nossa não-responsabilidade por esses fatos do que eles serem necessários: eles não são fatos sobre agência.⁴ Para fortalecer seu argumento, Hermes propõe uma analogia. Considere a seguinte regra dedutiva:

Regra C $P, \neg\Box P \vDash NRP$

A Regra C é uma regra dedutiva que infere nossa não-responsabilidade de qualquer fato contingente. Imagine que tentássemos argumentar pela Regra C apresentando as seguintes instâncias confirmatórias:

$$\frac{\begin{array}{l} \text{Dinossauros existiram} \\ \neg\Box(\text{Dinossauros existiram}) \end{array}}{NR(\text{dinossauros existiram})}$$

$$\frac{\begin{array}{l} \text{A Terra gira em torno do Sol} \\ \neg\Box(\text{A Terra gira em torno do Sol}) \end{array}}{NR(\text{A Terra gira em torno do Sol})}$$

Essas inferências, afirma Hermes, não dão suporte à Regra C. Isso porque, no final das contas, a verdadeira explicação para nossa não-responsabilidade por esses fatos é que eles não tem a ver com nossas ações. Do mesmo modo, ele conclui, as instâncias confirmatórias para a Regra A apresentadas por van Inwagen não dão suporte a ela.

O argumento de Hermes, entretanto, se baseia numa compreensão errada de instâncias confirmatórias. As instâncias dão suporte à Regra A se ela é a melhor explicação para a *validade intuitiva* da inferência. Não é preciso que ela seja a melhor explicação para nossa *não-responsabilidade* pelos fatos em questão. Nesse ponto que encontramos a diferença relevante entre a Regra A e a Regra C: mesmo se nenhuma das regras apresentar a melhor explicação para a não-responsabilidade dos fatos em questão, as inferências usando a Regra C não são nem intuitivamente válidas.

Para motivar meu ponto, trago um exemplo apresentado por Ori Beck (2020). Considere a seguinte inferência:

$$\frac{\text{O céu é azul} \wedge \text{a grama é verde}}{\text{O céu é azul}}$$

⁴A objeção de Hermes é inspirado numa objeção de McKenna (2008) contra a Regra B, a qual será apresentada no próximo capítulo.

Essa inferência obviamente não apresenta a melhor explicação para sua conclusão. O céu é azul não porque é verdade que *o céu é azul e a grama é verde*, mas por causa de alguns fatos físicos sobre a interação entre a luz solar e nossa atmosfera. Ainda assim, a inferência ser intuitivamente válida nos dá uma razão para aceitar que exclusão da conjunção é uma regra inferencial válida. Do mesmo modo, ainda que a Regra A não seja uma boa explicação da conclusão das instâncias confirmatórias, ela é motivada por elas.

Ainda que o argumento de Hermes não tenha sucesso, temos motivos para rejeitar as instâncias confirmatórias apresentadas por van Inwagen. Mesmo se concedermos a van Inwagen que as instâncias confirmatórias são intuitivamente válidas, temos uma explicação alternativa para essa aparente validade. Minha proposta é que as instâncias confirmatórias da Regra A parecem válidas porque estamos aceitando implicitamente algo como *R-Capacidade*. Uma razão para aceitar isso vem do próprio van Inwagen (1983, 184); logo após apresentar as instâncias confirmatórias, ele afirma que a razão para nossa não-responsabilidade por fatos necessários é nossa falta de escolha (no sentido apresentado em *Escolha*) sobre eles.

Se eu estou certo sobre *R-Capacidade* (ou algo similar) ser a origem da aparente validade das inferências em questão, devemos rejeitar apelos a essas intuições no debate sobre o Argumento Direto. Dado que *R-Capacidade* é dialeticamente inapropriado nesse debate, intuições fundamentadas nele também são. Van Inwagen não pode, portanto, apelar a validade intuitiva dessas inferências, o que faz com que elas não forneçam nenhum suporte à Regra A.

Outro modo de formular meu argumento é o seguinte. É preciso dar uma explicação teórica para a validade intuitiva das inferências em questão. O incompatibilista oferece a validade da Regra A como explicação. Estabelecendo a validade dela desse modo, é apropriado que ele use a regra no Argumento Direto. O que eu estou propondo é uma explicação alternativa: as inferências parecem válidas porque estamos assumindo algo como *R-Capacidade*, que não pode ser usado no Argumento Direto. Enquanto o incompatibilista não apresentar uma razão pela qual a Regra A é uma explicação *melhor* para essas intuições, as instâncias não dão suporte à regra.

Nesta seção, eu discuti as instâncias confirmatórias da Regra A apresentadas por van Inwagen. Apresentei o contra-argumento a elas proposto por Hermes. Eu rejeitei esse contra-argumento e apresentei outro. Meu argumento é que, ainda que as inferências sejam intuitivamente válidas, essa intuição é explicada por princípios dialeticamente inapropriados para o debate sobre o Argumento Direto e, portanto, devemos rejeitar essas instâncias confirmatórias. Se esse argumento for sólido, temos no momento nenhuma motivação a favor da Regra A na literatura.

2.4 O Argumento Direto Lógico sem a Regra A

Nas duas últimas seções, eu apresentei razões para rejeitarmos as duas motivações para a Regra A apresentadas por van Inwagen, sua aparente indisputabilidade e algumas instâncias confir-

matórias. O que o proponente do Argumento Direto pode fazer em relação à Regra A? Uma opção é tentar salvar a regra mostrando que os contra-exemplos falham e dando novos argumentos a favor dela. Outra opção é reformular o Argumento Direto de modo a não usar a regra. Um modo de fazer isso é defendendo um Argumento Direto *Causal*. Argumentos Diretos Causais não dependem da Regra A. Com uma breve exceção no capítulo seguinte, eu não discuto Argumentos Diretos Causais nesta monografia. Outra opção que encontramos na literatura é simplesmente assumir como premissa aquilo que pretendíamos inferir usando a Regra A. Quem propôs essa solução foi Justin Capes (2016). Nesta seção, eu considerarei (e rejeitarei) essa proposta.

Capes, tendo em mente evitar alguns contra-exemplos à Regra B (que serão apresentados no próximo capítulo), propõe o próprio Argumento Direto Lógico. Seu argumento usa uma regra similar à Regra A e que está sujeita às mesmas objeções.⁵ Capes, porém, considera que seu argumento é plausível mesmo sem essa regra; ele sugere que podemos simplesmente assumir como premissa aquilo que pretendemos inferir com a regra.

Eu não discutirei o Argumento Direto de Capes, mas aplicarei sua sugestão de abandonar a Regra A ao Argumento Direto Original. Teríamos então o seguinte Argumento Direto:

Argumento Direto _{-A}		
P1 _{-A}	$NR(P_0 \supset (L \supset P))$	Nova Premissa
P2 _{-A}	NRP_0	Não-Responsabilisade pelo passado
P3 _{-A}	NRL	Não-Responsabilidade pelas leis
C1 _{-A}	$NR(L \supset P)$	De P1 _{-A} , P2 _{-A} por Regra B
C2 _{-A}	NRP	De C1 _{-A} , P3 _{-A} por Regra B

P1_{-A} é equivalente a C2 do *Argumento Direto Original*, mas, em vez de uma conclusão tirada da definição de determinismo pela Regra A, é uma premissa do argumento. Que razão temos para aceitar essa premissa? Capes diz o seguinte:

Que um estado de coisas é a consequência lógica de outro simplesmente não parece o tipo de fato pelo qual uma pessoa poderia ser parabenizada ou culpada. A conclusão disso tudo é que mesmo se a regra A* for inválida, eu acho que isso faria pouco ou nada para diminuir o apelo do novo argumento direto. (Capes, 2016, 1492)

A passagem em questão me parece ambígua entre dois argumentos. O primeiro argumento seria que $P_0 \supset (L \supset P)$ não é o tipo de fato que poderia justificar atitudes reativas e, portanto, não é possível ser moralmente responsável por esse fato. O segundo argumento seria que, aparentemente, ninguém é responsável por $P_0 \supset (L \supset P)$ e, portanto, podemos assumir P1_{-A} como uma premissa.

⁵A única diferença entre a regra usada por Capes e a Regra A é que a regra dele é restrita a um único agente arbitrário. Ou seja, ela conclui que um determinado agente não é responsável por um fato.

Considero que o segundo argumento é o mais forte, mas, antes de discuti-lo, falarei sobre o primeiro. Vejo dois problemas com esse argumento. Primeiro, ainda que falemos de responsabilidade moral no debate sobre o Argumento Direto, estamos exclusivamente interessados na condição metafísica da responsabilidade. Por causa disso, não nos interessa a plausibilidade de termos atitudes reativas sobre um dado fato, mas sim se podemos ter o controle sobre esse fato que justificaria atitudes reativas se as outras condições para responsabilidade obterem. Segundo, vale lembrar que, para concluirmos que responsabilidade é incompatível com determinismo, o Argumento Direto deve ser generalizado substituindo P por cada um dos fatos. Logo, uma das versões de $P_0 \supset (L \supset P)$ é, por exemplo, $P_0 \supset (L \supset \textit{centenas de milhares de pessoas morreram de COVID no Brasil em 2021})$. Quando consideramos essa versão da premissa, não fica tão claro se realmente não podemos ter atitudes reativas sobre esse fato. Dado essa razões, creio que apelar à nossa falta de atitudes reativas sobre o fato em questão é uma má defesa da premissa.

Passemos agora ao segundo argumento. Parece-me que a ideia central nesse argumento é que as objeções contra a Regra A são irrelevantes para o papel que a regra tem no Argumento Direto. Sendo assim, podemos simplesmente pular seu uso e assumir a conclusão. Em outras palavras, ainda que haja razões para crer que a Regra A é inválida, não há razões para crer que $P_0 \supset (L \supset P)$ é um dos fatos necessários pelos quais podemos ser responsáveis.

Minha resposta a esse argumento é que, se aceitarmos *R-Verificador*, assumir $P1_{-A}$ é uma petição de princípio contra o compatibilista. Para que $P1_{-A}$ seja verdadeiro caso aceitemos *R-Verificador*, é preciso que não haja um agente cuja ação livre tenha causado o verificador de $P_0 \supset (L \supset P)$ obter. Qual o verificador de $P_0 \supset (L \supset P)$? Sendo uma implicação material com antecedente verdadeiro, o que a torna verdadeira é a verdade do consequente, i.e., $L \supset P$. Essa implicação também tem o antecedente verdadeiro e, de maneira similar, o que a torna verdadeira é P . Logo, o verificador de $P_0 \supset (L \supset P)$ é o verificador de P . Se a verdade de $P1_{-A}$ depende de que não haja um agente cuja ação livre tenha causado o verificador de $P_0 \supset (L \supset P)$ obter, ela também depende de que não haja um agente cuja ação livre tenha causado o verificador de P obter. Que não haja um agente cuja ação livre tenha causado o verificador de P obter acarreta, dado *R-Verificador*, *NRP*. Logo, a verdade de $P1_{-A}$ depende de *NRP* ser verdadeiro. *NRP*, porém, é a conclusão do Argumento Direto. Se o proponente do *Argumento Direto*_{-A} assume a conclusão do argumento nas premissas, o argumento é uma petição de princípio contra o compatibilista. Portanto, *pace* Capes, uma das motivações contra a Regra A – *R-Verificador* – nos dá uma razão para rejeitar o *Argumento Direto*_{-A}.

Nesta seção, eu considereei a proposta de Justin Capes de formular o Argumento Direto sem usar a Regra A. Argumentei contra essa proposta. A parte mais importante da minha argumentação foi apontar que, assumindo *R-Verificador*, o proponente do *Argumento Direto*_{-A} comete petição de princípio.

2.5 Considerações Finais

Neste capítulo, eu discuti algumas das críticas à Regra A que encontramos na literatura. Primeiro, eu apresentei uma proposta de análise da noção de *responsabilidade por uma proposição*. Na segunda seção, eu apresentei um contra-exemplo à Regra A e mostrei como ele é motivado pela análise proposta. Na terceira seção, eu discuti as instâncias confirmatórias apresentadas por van Inwagen. Eu expus, e depois rejeitei, uma objeção de Hermes a elas. Por fim, levantei minha própria objeção a essas instâncias confirmatórias. Na quarta seção, eu considerei e rejeitei uma formulação do Argumento Direto sem a Regra A.

Dado o que foi exposto neste capítulo, o que o proponente do Argumento Direto deveria fazer quanto à Regra A? Vejo duas soluções. A mais promissora me parece ser abandonar o *Argumento Direto Original* e focar os esforços em ou (i) polir uma boa versão do Argumento Direto Causal ou (ii) propor um Argumento Direto Lógico que não use a Regra A. Discutirei um pouco essas possibilidades na conclusão da monografia. Caso o incompatibilista queira salvar o *Argumento Direto Original*, o melhor parece ser encontrar uma nova análise da noção de *responsabilidade por uma proposição*; uma análise melhor que *R-Verificador* e que motive tanto a Regra A quanto a Regra B.

3. Críticas à Regra B

No capítulo anterior, apresentei algumas críticas à Regra A. Neste capítulo, apresentarei críticas à Regra B. Segundo van Inwagen, a Regra B era o elo mais fraco no *Argumento Direto Original*. Encontramos, realmente, mais críticas a ela do que às outras partes do argumento. A Regra B é amplamente considerada inválida devido a um contra-exemplo que apresentarei na primeira seção. Encontramos, entretanto, um argumento a favor da Regra B, que apresentarei na segunda seção. Na terceira seção, discutirei a principal crítica à Regra B (e que se estende para princípios similares): o proponente do Argumento Direto não deu motivações suficientes para aceitar a Regra B e, portanto, o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo. Na última seção, farei algumas considerações finais.

3.1 Um contra-exemplo à Regra B

Nesta seção, eu apresentarei um contra-exemplo à Regra B proposto por Mark Ravizza e depois defendido por ele e John Fischer (Ravizza (1994) e Fischer & Ravizza (1998, Capítulo 6)). A maior parte da literatura aceitou o contra-exemplo, o que levou a uma série de propostas de outros princípios de transferência de não-responsabilidade. Encontramos substitutos À Regra B em Warfield (1996), McKenna (2001; 2008) e Capes (2016).

Antes de passar ao contra-exemplo, vale a pena apresentar o diagnóstico de Ravizza quanto à Regra B; sua hipótese de porque a regra parece plausível e onde ela falha. Segundo Ravizza, a plausibilidade inicial da Regra B está em sua aparente validade nos casos mais usuais:

A plausibilidade inicial da Regra (B) vem, eu acho, do fato de que nos casos mais comuns parece intuitivamente que não-responsabilidade pode ser “transferida” do modo que o princípio sugere. Ou seja, se uma pessoa não é responsável por uma coisa, e também não é responsável por essa coisa levar a outra, então a pessoa aparentemente não é responsável pela outra. E van Inwagen apela a esses casos comuns para fomentar suporte a (B). (Ravizza, 1994, 67)

Como a passagem mostra, para Ravizza, as instâncias confirmatórias apresentadas por van Inwagen para a Regra B realmente dão suporte à regra. Ravizza, porém, acha que essa aparente validade está baseada na ideia de *inevitabilidade*: “parece que parte do que causa nossas intuições nesses casos pode ser nada mais que a ideia de que um agente não é responsável por

algo inevitável” (Ravizza, 1994, 67). Se P é inevitável e é inevitável que P leve a Q , então Q também é inevitável. Se a inferência anterior é válida e inevitabilidade impede responsabilidade, a Regra B é válida. Ravizza aponta, entretanto, que a ideia de que inevitabilidade impede responsabilidade é posta em xeque pelos casos de Frankfurt, que são aceitos pelo adversário do Argumento Direto.

A estratégia de Ravizza, portanto, é gerar contra-exemplos à Regra B tendo como base casos de Frankfurt, cenários em que um agente é responsável por algo que era, para ele, inevitável. Ravizza apresenta mais de um contra-exemplo; apresento aquele mais discutido na literatura:

Erosão Imagine que Betty plante seus explosivos na cratera de uma geleira e detone a carga em t_1 , causando uma avalanche que destrói o forte inimigo em t_3 . Entretanto, sem que Betty ou seus superiores soubessem, a geleira estava gradualmente derretendo, deslizando e erodindo. Se Betty não tivesse colocado o dinamite na cratera, gelo e pedra teriam se soltado e destruído a base inimiga em t_3 .

Com esse cenário, Ravizza pretende apresentar um contra-exemplo à Regra B. Sendo E a proposição de que a geleira está erodindo e A a proposição de que uma avalanche destrói a base inimiga em t_3 , podemos obter o seguinte contra-exemplo à Regra B a partir do cenário anterior:

ER1 NRE

ER2 $NR(E \supset A)$

ER3 NRA

(De $ER1$ e $ER2$ pela Regra B)

ER4 $\neg NRA$

Dado que Betty não sabia que a geleira estava erodindo – e não há por que crer que essa seja uma ignorância culpável –, $ER1$ é verdadeiro. A motivação para aceitar $ER2$ é que ninguém é responsável pelo fato de que a erosão da geleira é suficiente para a destruição da base em t_3 . De $ER1$ e $ER2$, a Regra B nos permite inferir que ninguém é responsável pelo fato de que a base inimiga é destruída em t_3 ($ER3$). Isso, porém, contradiz nossa intuição, expressa em $ER4$, de que alguém – Betty – é responsável pela destruição da base inimiga. Da contradição que obtemos de $ER3$ e $ER4$, Ravizza conclui que a Regra B é inválida.

Na próxima seção, apresentarei uma defesa da Regra B e apontarei onde creio que o contra-exemplo de Ravizza falha. Antes disso porém, exporei, na subseção seguinte, a resposta de McKenna contra *Erosão*. Sua resposta é a mais bem estabelecida na literatura e é o ponto de partida para os Argumentos Diretos Causais. Ainda que eu não vá discuti-los a fundo, é interessante ver como o debate sobre o Argumento Direto Original levou aos Argumentos Diretos Causais.

3.1.1 A resposta de McKenna

Nesta subseção, apresentarei a resposta de Michael McKenna a *Erosão*. A resposta de McKenna não é uma defesa da Regra B; ele aceita que *Erosão* mostra que a Regra B é inválida. Em vez disso, sua resposta consiste em argumentar que o artifício em *Erosão* que permite o cenário funcionar como um contra-exemplo à Regra B não está presente na inferência relevante ao Argumento Direto; uma mera reformulação da Regra B permitiria evitar o contra-exemplo sem prejudicar o Argumento Direto.

Por que *Erosão* funciona como um contra-exemplo à Regra B? Segundo McKenna, o elemento chave em *Erosão* é a presença de duas cadeias causais distintas: (i) a cadeia causal (não-ocorrente) da erosão à destruição da base – que assegura a inevitabilidade da destruição – e (ii) a cadeia causal da ação de Betty até a destruição da base – que fundamenta a responsabilidade de Betty pela destruição –. McKenna afirma, porém, que, se o determinismo é verdadeiro, a cadeia causal do passado até nossa ação e da nossa deliberação até nossa ação são a mesma. Diferente de *Erosão*, em um mundo determinístico similar ao nosso, não há duas cadeias causais, uma fundamentando a verdade do determinismo e outra fundamentando a responsabilidade moral. *Erosão*, portanto, ainda que um contra-exemplo à Regra B, não ataca a transferência de não-responsabilidade central ao Argumento Direto.

McKenna propõe, então, reformular a Regra B de modo a restringi-la a casos em que há apenas uma cadeia causal. Não apresentarei as propostas de McKenna, pois nos levaria a discutir os Argumentos Diretos Causais. Vale a pena, entretanto, discutir uma das questões chaves da resposta de McKenna. Para que o argumento de McKenna seja cogente, é importante que sua reformulação da Regra B se aplique à inferência presente no Argumento Direto. Ou seja, já que a regra é restrita a casos com uma única cadeia causal, parece que o proponente dessa nova regra dedutiva está comprometido com haver apenas uma cadeia causal nos mundos determinísticos relevantes para o debate.

Esse comprometimento é inocente o bastante para a proposta de McKenna ser aceitável? A resposta a essa pergunta vai variar conforme as teses que se aceitar em outros debates de metafísica. Alguém, por exemplo, que acredite em níveis de realidade e causação mental irreduzível, pode achar que o preço é alto demais; McKenna estaria se comprometendo com uma posição implausível. Já alguém que aceite que composição é identidade, provavelmente não verá problema em aceitar que os mundos relevantes possuem apenas uma cadeia causal.

McKenna, entretanto, não precisa necessariamente se comprometer com haver apenas uma cadeia causal. Ele pode aceitar a sugestão de Fischer (2004, 195) de usar o Argumento Direto restrito a uma cadeia causal como um primeiro passo argumentativo pelo incompatibilismo. A defesa incompatibilista teria um segundo passo além do Argumento Direto, em que se mostraria que cadeias causais adicionais não possibilitariam responsabilidade moral nos mundos que o Argumento Direto se aplica. Um modo de argumentar por essa estratégia de duas etapas seria a seguinte: o Argumento Direto mostra que não-responsabilidade transfere pela cadeia causal

mais fundamental; a não-responsabilidade pela cadeia causal mais fundamental transfere para as outras cadeias causais; portanto, independente de quantas cadeias causais distintas há no mundo, não somos moralmente responsáveis.

O problema dessa argumentação com duas partes é que regra de transferência de não-responsabilidade o incompatibilista usaria na segunda parte. Ele não pode apelar a uma regra que transfere não-responsabilidade por cadeias causais, já que a relação entre as distintas cadeias causais não seria de causação. Ele também não pode apelar a uma regra que implicasse na validade da Regra B, já que ele aceita que *Erosão* é um contra-exemplo de sucesso. Parece que ele precisaria apelar a uma regra específica à relação em questão, algo como fundamentação.¹

Vemos então que a resposta de McKenna contra *Erosão* tem seus custos; ao restringir a transferência de não-responsabilidade a uma cadeia causal só, o incompatibilista precisa (i) ou defender que só há uma cadeia causal nos mundos relevantes, (ii) ou defender que a não-responsabilidade por uma das cadeias causais transfere para as outras. Qualquer via que seja tomada, será preciso se aventurar em outros debates de metafísica, como os debates sobre composicionalidade e fundamentalidade.

3.2 Uma defesa da Regra B

Na seção anterior, apresentei o contra-exemplo à Regra B proposto por Ravizza. Nesta seção, apresentarei uma defesa da Regra B e explicarei porque *Erosão* falha. Uma versão mais detalhada dessa defesa foi apresentada anteriormente em Merluzzi & Maruchi (2019).

Minha defesa tem como base *R-Verificador* e se assemelha ao argumento que apresentei em 2.4. Uma inferência usando a Regra B tem como ponto de partida *NRP* e $NR(P \supset Q)$. Assumindo *R-Verificador*, para que $NR(P \supset Q)$ seja verdadeiro, é preciso que não haja um agente cuja ação livre tenha causado o verificador de $P \supset Q$. Qual o verificador de $P \supset Q$? Sendo uma implicação material com antecedente verdadeiro – já que *NR* é um operador factivo e *NRP* é verdadeiro –, o que a torna verdadeira é a verdade do consequente, i.e., *Q*. Logo, o verificador de $P \supset Q$ é o verificador de *Q*. Se não há um agente cuja ação livre tenha causado o verificador de $P \supset Q$, então também não há um agente cuja ação livre tenha causado o verificador de *Q*. Que não haja um agente cuja ação livre tenha causado o verificador de *Q* acarreta, dado *R-Verificador*, NRQ . Assumindo *R-Verificador*, inferimos NRQ de *NRP* e $NR(P \supset Q)$; portanto, se *R-Verificador* é verdadeiro, a Regra B é válida.

Aceitando essa defesa da Regra B, *Erosão* falha como um contra-exemplo por causa de *ER2*. Já que Betty é responsável pela destruição da base, ela também é responsável por $E \supset A$; sendo assim, *ER2* é falso. Se *ER2* é falso, por que *Erosão* foi bem aceito como um contra-exemplo à Regra B? A hipótese que me parece mais plausível é que *ER2* é entendido como equivalente a

¹Um argumento que tenta transferir não-responsabilidade “debaixo pra cima” é o *Puppet Puzzle* apresentado por Andrew Bailey & Joshua Rasmussen (2020).

algo como:²

ER2* *E* é causalmente suficiente para *A*;

Porém, para que a inferência de *ER3* a partir de *ER1* e *ER2** seja válida, precisamos de uma regra diferente da Regra B e teríamos, portanto, um Argumento Direto diferente. *ER2** pode fazer parte de um contra-exemplo a uma regra de transferência de não-responsabilidade, mas não à Regra B. Parece que, em alguma medida, a literatura oscila entre uma interpretação lógica e uma interpretação causal da Regra B. Na próxima seção, eu explorarei essa questão e apontarei três possíveis soluções para ela. Essa discussão fará parte do debate sobre termos ou não uma motivação para aceitar Regra B (ou alguma outra regra de transferência de não-responsabilidade). O que mostrei nesta seção é que há uma motivação para a Regra B: *R-Verificador*.

3.3 Novamente, Instâncias Confirmatórias

Nesta seção, apresentarei a principal objeção contra a Regra B e outras regras de transferência de não-responsabilidade. A objeção foi apresentada por McKenna (2008) e foca na dialética do debate. Segundo McKenna, o compatibilista não precisa mostrar que a Regra B é inválida porque, diz ele, o proponente do Argumento Direto não estabeleceu que a Regra B é, ao menos, *prima facie* válida. Se isso for o caso, o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo e, portanto, o compatibilista não precisa se preocupar em responder ao argumento. O debate aqui gira em torno de instâncias confirmatórias que poderiam motivar a validade da Regra B. Após apresentar o argumento de McKenna, apresentarei minha própria defesa de que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo.

Primeiro, relembremos a principal motivação que van Inwagen dá a favor da Regra B. Ele apresenta duas instâncias confirmatórias para a Regra B, *Mordida de Cobra* e *Platão*. A intuição de que essas inferências são válidas serviria como uma razão *prima facie* para aceitar que a Regra B é válida. McKenna, porém, acha que os cenários que geram essas inferências têm uma característica que impede que elas deem suporte a Regra B:

Em *Mordida de Cobra*, Platão, [...], as cadeias de suficiência causal pelas quais qualquer transferência de não-responsabilidade é transmitida nunca “passam por” um agente funcionando normalmente que exercita suas capacidades deliberativas de modo desimpedido no processo de produção de uma ação (supostamente) livre pelo qual ele ou ela é moralmente responsável. (McKenna, 2008, 376)

²Uma evidência de que é assim que algumas pessoas entendem a *Regra B* está na passagem que citei do artigo de Ravizza na primeira seção deste capítulo. Nela, ao explicar a regra, Ravizza não fala de uma coisa implicar outra, mas sim de “essa coisa levar a outra”, o que claramente tem uma leitura causal.

Chamemos casos em que a não-responsabilidade não passa por um agente funcionando normalmente que exercita suas capacidades deliberativas de modo desimpedido no processo de produção de uma ação (supostamente) livre pelo qual ele ou ela é moralmente responsável de casos *não-agenciais*. *Mordida de Cobra e Platão* dão a entender casos não-agenciais. Segundo McKenna, as inferências geradas a partir de casos não-agenciais não podem ser usadas para motivar uma regra de transferência de não-responsabilidade. Para entendermos o argumento de McKenna, devemos primeiro ter em mente que as inferências da Regra B presentes no *Argumento Direto Original* não são de casos não-agenciais; elas transferem não-responsabilidade pela nossas ações, algumas das quais são supostamente livres. A ideia central do argumento de McKenna é que casos não-agenciais não podem ser usados para estabelecer uma regra de transferência de não-responsabilidade que será usada em casos como o do Argumento Direto. Antes de aplicar sua regra de transferência de não-responsabilidade no Argumento Direto, o proponente do argumento deve estabelecer que essa regra vale em casos onde a não-responsabilidade *passa* por um agente funcionando normalmente que exercita suas capacidades deliberativas de modo desimpedido no processo de produção de uma ação (supostamente) livre pelo qual ele ou ela é moralmente responsável.

O desafio que McKenna coloca ao incompatibilista, então, é encontrar esses casos que podem ser apropriadamente usados para motivar uma regra de transferência de não-responsabilidade. Enquanto o incompatibilista não apresentar esses casos, conclui McKenna, o Argumento Direto apresenta nem mesmo uma ameaça *prima facie* ao compatibilismo. McKenna acha que o tipo de caso relevante seria aquele que encontramos em Argumentos da Manipulação, como o Argumento dos Quatro Casos de Derek Pereboom (2001; 2014).

A objeção de McKenna recebeu dois tipos de resposta (em geral, de maneira complementar): (i) foi defendido que casos não-agenciais podem sim ser usados para estabelecer uma regra de transferência e (ii) foram apresentadas instâncias confirmatórias que supostamente satisfazem o *desideratum* que McKenna apresenta.

A literatura desse debate é extensa. Schnall & Widerker (2012) argumentaram, *pace* McKenna, em favor do Argumento Direto. Yael Loewenstein (2016) defendeu o argumento de McKenna contra a resposta deles. Finalmente, Ori Beck (2020) defendeu os argumentos de Schnall & Widerker contra os argumentos de Loewenstein. Fora isso, Justin Capes (2022) também apresentou uma defesa do Argumento Direto. Na subseção seguinte, eu argumentarei que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo de maneira independente dos argumentos de McKenna. Meu argumento, portanto, em grande medida será tangencial ao debate presente nessa literatura.

3.3.1 Causação e a Regra B

Nesta subseção, defenderei que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo. Meu argumento é diferente do de McKenna. O ponto de partida do meu argumento é

apontar que as considerações apresentadas a favor das instâncias confirmatórias da Regra B mostram que os proponentes delas oscilam (sem uma justificativa teórica) entre uma interpretação causal e uma interpretação não-causal da regra. Eu proporei três possíveis saídas para essa situação: (i) o proponente defende que as considerações causais motivam a Regra B (mesmo ela não sendo em termos de causação); (ii) ele apresenta uma defesa não-causal da Regra B; ou (iii) ele abandona a Regra B e usa as instâncias confirmatórias para motivar uma regra de transferência causal de não-responsabilidade. Argumentarei que, ainda que tomar qualquer uma dessas rotas aumente a clareza da discussão, todas falham em estabelecer o Argumento Direto como uma ameaça (pelo menos) *prima facie* ao compatibilismo.

Antes de apresentar meu argumento, apresento outras duas instâncias confirmatórias da Regra B, essas propostas por Schnall & Widerker (S&W daqui em diante). S&W pretendem satisfazer o *desideratum* de McKenna para uma instância confirmatória aceitável (ainda que eles argumentem contra o *desideratum*). Para os fins do meu argumento, todas essas instâncias confirmatórias são dialeticamente apropriadas. O cenário da primeira instância:

Repreensão Uma manhã, o oficial em comando de um sistema de defesas anti-mísseis do exército recebe um telefone de um de seus subordinados, Jones. Jones diz que está doente e não poderá ir trabalhar hoje. Sem que o oficial ou Jones soubessem, há um homem, Smith, na área que parece exatamente com Jones. Mais tarde naquele dia, o oficial está indo para outra base quando vê Smith, que ele acha ser Jones, se divertindo na praia. O oficial, depois de certa deliberação, decide repreender duramente Jones no dia seguinte e assim o faz. Jones se sente extremamente ofendido pela repreensão. Ele nega ter estado na praia, mas isso não muda a reação do oficial, que insiste que viu Jones “com os próprios olhos”.

Desse cenário, obtemos a seguinte instância confirmatória:

RP1 NR(O oficial acredita que Jones estava se divertindo na praia)

RP2 NR(O oficial acredita que Jones estava se divertindo na praia \supset O oficial repreende Jones)

RP3 NR(O oficial repreende Jones) (De RP1 e RP2 pela Regra B)

O segundo caso é o seguinte:

Tornado Jones vê um tornado se aproximando. Ele delibera o que fazer por alguns segundos e então decide, para salvar sua vida, entrar no seu carro e dirigir para longe, e assim o faz

De *Tornado*, S&W propõe a seguinte instância confirmatória:

TN1 NR(Jones vê um tornado se aproximando)

TN2 NR(Jones vê um tornado se aproximando \supset Jones dirige para longe para salvar sua vida)

TN3 *NR*(Jones dirige para longe para salvar sua vida) (De TN1 e TN2 pela Regra B)

Temos, então, quatro instâncias da Regra B que, segundo seus proponentes, são intuitivamente válidas. O que pode ser dito a favor da afirmação de que são intuitivamente válidas? Van Inwagen não apresenta um argumento a favor dela, mas S&W sim.

Para defender que as instâncias são intuitivamente válidas, S&W propõe o que eles chamam de ‘argumento reverso’. O formato de um argumento reverso é o seguinte: Imagine um agente que aceita as premissas de uma instância confirmatória, mas não sua conclusão. Ficaríamos intrigados com esse conjunto de crenças – no sentido de acharmos que a pessoa tem um conjunto *inconsistente* de crenças –. A explicação para essa nossa reação, conclui o argumento reverso, é que a instância confirmatória é intuitivamente válida. Por questões de simplicidade, apresentarei apenas o argumento reverso correspondente a Repreensão. Imagine uma pessoa, Alice, que aceita *RP1* e *RP2*, mas não *RP3*. Ou seja, Alice acredita que o oficial não é responsável por crer que Jones estava se divertindo na praia, não é responsável por *se o oficial acredita que Jones estava se divertindo na praia, então ele repreende Jones*, mas crê que o oficial é responsável por repreender Jones. S&W concluem seu argumento reverso do seguinte modo:

Nesse ponto, ficaríamos bem intrigados com a posição de Alice; pois parece que nesse cenário, se o oficial não é culpável por repreender Jones [então \neg *NR*(O oficial repreende Jones)], isso só pode ser ou porque ele é culpado por acreditar que Jones estava na praia [então \neg *NR*(O oficial acredita que Jones estava se divertindo na praia)] ou porque **ele é culpado por essa crença levá-lo a fazer a repreensão [logo \neg *NR*(O oficial acredita que Jones estava se divertindo na praia) \supset O oficial repreende Jones)]. Isto é, se alguém nega a conclusão desse argumento [da Regra B], seria natural assumir que essa pessoa nega uma das premissas – um forte indicativo de validade. (Schnall & Widerker, 2012, 30, grifo meu)**

Nessa passagem, S&W parecem entender *RP2* como equivalente ao seguinte:

RP2’ *NR*(A crença do oficial causa ele repreender Jones);

Não é necessário explicar porque *RP2* e *RP2’* não são equivalentes; o primeiro é sobre não-responsabilidade por uma implicação material, o segundo sobre não-responsabilidade por uma relação causal. Os próprios S&W apontam essa diferença:

O novo princípio (que [a Regra B] expressa na forma de uma regra de inferência) é que *não-responsabilidade transfere por um condicional material*

[...]

Note que esse não é o mesmo que o princípio que não-responsabilidade é preservada por *determinação causal* por fatos pelos quais o agente não é responsável.

(32)

Temos então um problema: eles apelam a uma intuição sobre não-responsabilidade por uma relação causal para motivar uma afirmação de não-responsabilidade por uma implicação material. Sem uma justificativa teórica para isso, esse passo é inapropriado (e acabamos sem uma motivação a favor da afirmação sobre a implicação material).

O problema, porém, é mais profundo; encontramos essa oscilação entre considerações causais e não-responsabilidade sobre implicações em outras defesas das instâncias confirmatórias. Por exemplo, ao discutir uma versão de *Mordida de Cobra* usando a versão de Capes da Regra B, Beck justifica sua contraparte de SB2 dizendo que John não é responsável pelo “fato que aquelas circunstâncias realmente levaram a sua morte” (Beck, 2020, 1813). Esse fato, porém, é sobre causação e, logo, distinto do fato relevante para SB2.

Se as intuições que os proponentes da Regra B apelam para motivar suas instâncias confirmatórias são geradas por considerações causais – não tendo sido apresentado uma razão para crer que elas justificam as não-responsabilidades pelas implicações materiais em questão –, o compatibilista tem uma boa razão para considerar essas intuições enganosas. Ou seja, o compatibilista teria uma *teoria de erro* para as intuições em questão: a Regra B parece intuitivamente válida, mas essa intuição não é gerada pela validade da regra, mas sim por certas considerações causais.³ Sendo assim, o compatibilista pode, de maneira dialeticamente apropriada, rejeitar as instâncias confirmatórias apresentadas e defender que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo.

Como o proponente do Argumento Direto pode resolver esse problema? Vejo três possíveis soluções: (i) defender que as considerações causais justificam as não-responsabilidades pelas implicações materiais em questão, (ii) defender a Regra B sem apelar a causação, (iii) reformular o Argumento Direto em termos de causação. Nas próximas três sub-subseções, defenderei que nenhuma dessas estratégias têm sucesso. Numa quarta subseção, concluirei minha argumentação.

De Causação à Regra B

Argumentei até agora que há um problema no debate sobre o Argumento Direto; não-responsabilidade sobre fatos causais é usado para motivar não-responsabilidade sobre implicações. Nesta sub-subseção, eu tentarei ligar esses dois tipos de não-responsabilidade. Ou seja, considerarei a estratégia de defender as instâncias confirmatórias da Regra B argumentando que a não-responsabilidade sobre os fatos causais em questão acarreta a não-responsabilidade sobre as implicações em questão. Defenderei que essa estratégia falha em estabelecer o Argumento Direto como uma ameaça (pelo menos) *prima facie* ao compatibilismo.

O proponente do Argumento Direto pode defender que os fatos causais aos quais apela para motivar a Regra B acarretam a implicação em questão. A partir disso, ele pode argumentar que

³Uma teoria de erro para uma intuição de que *P* é uma explicação de porque *P* parece verdadeiro, mas que não dá justificativa para crer que *P*.

não-responsabilidade transfere por esse acarretamento. Nesse caso, teríamos, por exemplo, a seguinte instância a partir de *Repreensão*:

RP1 NR(O oficial acredita que Jones estava se divertindo na praia);

RP1.5 \square (A crença do oficial causa ele repreender Jones \supset (O oficial acredita que Jones estava se divertindo na praia \supset O oficial repreende Jones));

RP2' NR(A crença do oficial causa ele repreender Jones);

RP2 NR(O oficial acredita que Jones estava se divertindo na praia \supset O oficial repreende Jones);
(De RP1.5 e RP2' pela Regra A e pela Regra B)

RP3 NR(O oficial repreende Jones); (De RP1 e RP2 por Regra B)

Essa formulação de *Repreensão* evita a objeção apresentada no começo dessa subseção. Essa formulação, porém, tem outro problema. Seu proponente não pode simplesmente assumir a inferência de RP2 a partir de RP1.5 e RP2'. No final das contas, essa inferência depende da Regra B; assumir sua validade seria cometer petição de princípio. Logo, uma motivação a favor dessa inferência precisa ser fornecida.⁴

O proponente dessa formulação poderia apelar a um argumento reverso para motivar a inferência em questão. O argumento reverso seria o seguinte. Suponha que Alice acredita que o oficial é responsável por *Se o oficial acredita que Jones estava se divertindo na praia, então o oficial repreende Jones* (i.e., ela nega RP2), mas ela também acredita que ele não é responsável por sua crença levá-lo a repreender Jones (i.e., ela aceita RP2'). Por fim, ela acredita que o fato de que sua crença o levou a repreender Jones acarreta que, *se ele acredita que Jones estava se divertindo na praia, então ele repreende Jones* (i.e., ela aceita RP1.5). Nesse ponto do argumento, seu proponente diz “Nós ficaríamos profundamente intrigados pelas crenças de Alice. Isso mostra que essa inferência é intuitivamente válida”,

É difícil avaliar o sucesso desse argumento. Minha resposta sincera a ele é que eu não me sinto profundamente intrigado pelas crenças de Alice. Não apenas isso, mas me parece que, dada a complexidade e technicalidade das crenças dela, ninguém poderia ter uma intuição forte sobre esse caso. Àqueles que se sentem profundamente intrigados pelas crenças dela, minha resposta será insatisfatória. Ainda assim, acredito que os compatibilistas ficarão satisfeitos em negar a intuitividade da inferência de RP2 a partir de RP1.5 e RP2'. Se a plausibilidade do Argumento Direto depende de uma intuição tão controversa, o argumento dificilmente é uma ameaça *prima facie* ao compatibilismo.

⁴Um segundo problema com a inferência é que ela depende da Regra A. Como vimos no capítulo anterior, também não foram apresentadas instâncias confirmatórias de sucesso a favor da regra.

Veridadores ao Resgate

Nem todos os argumentos a favor da Regra B presentes na literatura sobre a dialética do Argumento Direto dependem de considerações causais. Defendendo *Tornado* contra uma objeção de Loewenstein, Beck (2020) apresenta uma defesa da Regra B em termos de veridadores. Eu não apresentarei a objeção nem o contra-argumento. Isso porque o argumento de Beck nada mais é que uma versão menos desenvolvida da defesa da Regra B que apresentei na seção 3.2. Argumentarei que a defesa com veridadores da Regra B falha em estabelecer o Argumento Direto como uma ameaça (pelo menos) *prima facie* ao compatibilismo.

A estratégia do proponente do Argumento Direto seria a seguinte aqui. Primeiro, motivar *R-Verificador* como uma análise de responsabilidade por fatos. Em seguida, mostrar que de *R-Verificador* temos uma defesa da Regra B (como fiz na seção 3.2). Desse modo, seria possível evitar a discussão sobre instâncias confirmatórias; a motivação para aceitar a Regra B não vem de instâncias confirmatórias, mas sim de *R-Verificador*.

Creio que essa estratégia estabelece a Regra B como *prima facie* válida. O problema é que isso só é suficiente para estabelecer o Argumento Direto como uma ameaça *prima facie* ao compatibilismo se nossa defesa da Regra B não por em xeque alguma outra parte do argumento. Como vimos em 2.2, porém, *R-Verificador* nos dá uma motivação para recusar a Regra A. Sendo assim, ainda que o princípio possa ser usado para motivar a Regra B, ele não pode ser usado para motivar o Argumento Direto como um todo. Portanto, essa estratégia falha em estabelecer o Argumento Direto como uma ameaça (pelo menos) *prima facie* ao compatibilismo.

Por que não um Argumento Direto Causal?

Como argumentei nas duas últimas sub-subseções, os proponentes do Argumento Direto não tem um bom argumento a favor da Regra B. Parte da razão para isso é que seus argumentos a favor da regra estão em termos causais, mas a Regra B não. Dado esse cenário, a melhor solução pode ser reformular a Regra B – e consequentemente o Argumento Direto – em termos causais. Eu desenvolverei essa estratégia nesta sub-subseção e argumentarei que ela falha em estabelecer o Argumento Direto como uma ameaça (pelo menos) *prima facie* ao compatibilismo.

McKenna (2001; 2008) apresentou formulações causais do Argumento Direto com o intuito de evitar o contra-exemplo à Regra B apresentado em 3.1. Encontramos, portanto, Argumentos Diretos Causais na literatura. Eu não apresentarei essas formulações, já que elas merecem uma discussão mais pormenorizada e que foge ao interesse desta seção, a dialética do argumento. Apresentarei uma versão simplificada de Argumento Direto Causal, porém as críticas que apresentarei nesta sub-subseção valem para outras formulações.

O Argumento Direto Causal terá apenas uma regra dedutiva. Tendo ‘*a*’ e ‘*b*’ representando eventos e ‘*t*₁’ e ‘*t*₂’ intervalos de tempo, a regra dedutiva é a seguinte:

Transfer-NR $NR(a \text{ ocorre em } t_1), NR(a \text{ é parte da sequência atual de eventos que causa } b \text{ em } t_2) \vDash NR(b \text{ ocorre em } t_2)$

Agora, sendo p_0 o evento do universo sendo como era em um período do passado remoto, t_0 o intervalo de tempo em que p_0 ocorreu, e p um evento arbitrário ocorrendo em um tempo arbítrio t_1 , temos o Argumento Direto Causal :

P1* NR(p_0 ocorre em t_0)

P2* NR(p_0 é parte da sequência atual de eventos que causa p em t_1)

C1* NR(p ocorre em t_1) (De P1 e P2 por Transfer-NR)

Uma vantagem do Argumento Direto Causal sobre o Argumento Direto Original é que as considerações causais apresentadas para motivar as instâncias confirmatórias realmente as motivam. Por exemplo, temos a seguinte versão de *Repreensão*:

RP1* NR(O oficial acredita que Jones estava se divertindo na praia em t_1)

RP2* NR(O oficial acreditar que Jones estava se divertindo na praia é parte da sequência atual de eventos que causa o oficial repreender Jones)

RP3* NR(O oficial repreende Jones em t_2) De RP1* e RP2* por Transfer-NR

As justificativas a favor de *Repreensão* apresentadas por S&W (como o argumento reverso) nos dão razão para aceitar essa versão da instância (e, portanto, parecem motivar Transfer-NR). Resolvida a objeção que eu apresentei anteriormente – que considerações causais não podiam motivar a Regra B, poderíamos então continuar o debate; veríamos se as objeções de Lowenstein e os contra-argumentos de Beck são cogentes.

Como já tinha dito anteriormente, essa não é a rota que eu seguirei. Isso porque, mesmo se o compatibilista conceder que Transfer-NR é *prima facie* válida, o Argumento Direto Causal é nem mesmo *prima facie* uma ameaça ao compatibilismo.

Para que o Argumento Direto Causal seja uma ameaça ao compatibilismo, não precisamos apenas de motivação para sua validade, mas também para suas premissas. P1* não é um problema. A premissa é bem intuitiva; ela apenas afirma que ninguém é moralmente responsável pelo passado remoto ser como era. Não há, porém, nem mesmo uma motivação *prima facie* para P2*. A contraparte de P2* no *Argumento Direto Original* é C2, que é concluída a partir da Regra A. O Argumento Direto Causal, entretanto, não faz uso da Regra A. Isso tem a vantagem de evitar as discussões apresentadas no Capítulo 2, mas gera o problema de motivar a premissa P2*.

Um proponente do argumento pode dizer que P2* é equivalente a *ninguém é moralmente pela verdade do determinismo*, o que é intuitivamente verdadeiro e, logo, P2* é intuitivamente verdadeiro. Não é isso, porém, que a premissa afirma. Ela afirma que ninguém é responsável pela cadeia causal ligando o passado e o evento em questão. Se esse evento for uma ação – nosso interesse nesse debate no final das contas –, essa cadeia causal passa pela deliberação que gera

a decisão do agente. Sendo assim, o proponente do Argumento Direto Causal está nos pedindo para aceitar como premissa que, se o determinismo é verdadeiro, ninguém é responsável pela relação causal entre nossas deliberações e ações. Obviamente, o compatibilista não precisa fazer essa concessão. Portanto, não temos uma motivação para P2* e o Argumento Direto Causal é nem *prima facie* uma ameaça ao compatibilismo.

É curioso notar que o próprio McKenna levanta o ponto apresentado no último parágrafo:

Mas no fundo dessas aparentemente inocentes afirmações estão afirmações sobre não-responsabilidade segundo as quais não se é nem parcialmente responsável pelo fato que a sua deliberação leva à sua ação. Isso, o compatibilista não deveria conceder somente porque essa relação é determinística. (McKenna, 2008, 379)

O erro de McKenna, me parece, é que ele caracteriza esse problema como uma questão sobre a validade da Regra B (e similares) e não um problema com premissas como P2*. Se pensarmos no *Argumento Direto Original*, a objeção que McKenna apresenta nessa passagem não deveria ser dirigida contra a Regra B, mas sim contra a Regra A. É na inferência de C1 para C2 que o incompatibilista, de maneira velada, conclui que não somos responsáveis pela ligação entre nossas deliberações e nossas ações.

De que outro modo o proponente do Argumento Direto Causal poderia tentar motivar P2*? Ele poderia apelar a *R-Verificador*. O proponente do *Argumento Direto Original* não podia fazer isso pois *R-Verificador* lança dúvidas sobre a Regra A, mas o Argumento Direto Causal não depende dela. Uma defesa de P2* a partir de *R-Verificador* seria a seguinte. Nós não causamos o verificador de p_0 é parte da sequência atual de eventos que causa p em t_1 obter. Se isso é o caso, assumindo *R-Verificador*, nós não somos responsáveis por esse fato e, portanto, P2* é verdadeiro.

Qual, porém, é o verificador de um fato causal? A resposta a essa pergunta depende de qual teoria causal você aceita. Não havendo uma resposta genérica a essa questão, parece que o proponente do Argumento Direto Causal não pode apelar a verificadores para defender P2*; pelo menos não sem se comprometer com alguma teoria causal em específico. Parece, então, que para o debate sobre o Argumento Direto Causal avançar, devemos parar de discutir causalidade de maneira genérica e nos atentarmos às diferentes teorias sobre a natureza da causalidade. Discutirei mais isso na Conclusão.

Apresentei nesta sub-subseção uma estratégia para tentar estabelecer o Argumento Direto como uma ameaça (pelo menos) *prima facie* ao compatibilismo. A estratégia consiste em propor um Argumento Direto Causal. Defendi que, mesmo se esse argumento for *prima facie* válido, ele é nem *prima facie* uma ameaça ao compatibilismo. Isso porque não há motivação para uma de suas premissas.

Conclusão

Argumentei nesta subseção que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo. Meu argumento procedeu do seguinte modo. Primeiro, aponte que os proponentes do Argumento Direto parecem motivar suas instâncias confirmatórias a partir de considerações causais. Sendo isso o caso, temos uma teoria de erro para a intuitividade da validade dessas inferências: essas inferências *parecem* válidas por causa de certas considerações causais, as quais não nos dão um motivo para aceitar que a Regra B é válida. Apresentei então três estratégias pelas quais o proponente do Argumento Direto poderia tentar resolver esse problema: (i) defender que as considerações causais justificam as não-responsabilidades pelas implicações materiais em questão, (ii) defender a Regra B sem apelar a causalidade, (iii) reformular o Argumento Direto em termos de causalidade.

Argumentei que as três estratégias falham em estabelecer o Argumento Direto como uma ameaça (pelo menos) *prima facie* ao compatibilismo. A primeira estratégia depende de uma intuição duvidosa e bem técnica, o que nos dá um motivo para rejeitá-la. A segunda, ainda que estabeleça a validade *prima facie* da Regra B, falha em estabelecer o Argumento Direto como uma ameaça pois coloca em dúvida a validade da Regra A. A terceira – que consiste em apresentar um Argumento Direto Causal – falha por não apresentar alguma motivação para as premissas do argumento. Não havendo nenhuma estratégia de sucesso, podemos concluir que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo.

3.4 Considerações Finais

Neste capítulo, discuti a Regra B. Na primeira seção, apresentei o contra-exemplo à regra proposto por Ravizza, *Erosão*. É bem aceito que o contra-exemplo mostra que a Regra B é inválida. A principal resposta ao contra-exemplo é propor um novo princípio de transferência de não-responsabilidade para o Argumento Direto. Nesse ponto, discuti a ideia de McKenna de que *Erosão* funcionava como um contra-exemplo por ter duas cadeias causais distintas e, portanto, seria preciso apenas restringir o princípio de transferência de não-responsabilidade a casos com apenas uma cadeia causal. O problema apresentado contra essa resposta é que ela ou se compromete com a existência de apenas uma cadeia causal nos mundos relevantes para a discussão, ou o Argumento Direto precisa ser complementado com um segundo argumento, que mostraria que a não-responsabilidade por uma cadeia causal transfere para as outras cadeias causais.

Na segunda seção, apresentei uma defesa da Regra B. A argumentação tem como base *R-Verificador*; assumindo esse princípio, é fácil mostrar que a regra é válida. *Erosão* falha como um contra-exemplo porque uma das premissas extraídas do cenário é falsa dado *R-Verificador*.

Na terceira seção, considerei a principal crítica à Regra B: que as instâncias confirmatórias falham em motivar a regra e, portanto, o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo. Eu apresentei o modo pelo qual McKenna formula essa objeção. Segundo ele,

o proponente do Argumento Direto precisa apresentar instâncias confirmatórias nas quais a não-responsabilidade passa por um agente funcionando normalmente que exercita suas capacidades deliberativas de modo desimpedido no processo de produção de uma ação (supostamente) livre pelo qual ele ou ela é moralmente responsável – o que ainda não tinha sido feito. Eu não discuti as respostas à objeção de McKenna. Em vez disso, eu propus meu próprio argumento a favor da tese de que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo. Meu argumento consistiu em apontar que os proponentes do Argumento Direto motivam as instâncias confirmatórias a partir de considerações causais as quais, no final das contas, não dão suporte à Regra B. Considerei três estratégias que poderiam ser adotadas para resolver esse problema e defendi que todas falham. Conclui, portanto, que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo.

Conclusão

Nesta monografia, eu discuti o Argumento Direto. Mais especificamente, a formulação desse argumento apresentada por Peter van Inwagen. No primeiro capítulo, apresentei sua formulação e as motivações a favor do argumento. Em especial, foquei nas duas regras dedutivas para o operador modal de não-responsabilidade que são usadas no Argumento Direto Original, a Regra A e a Regra B.

Segundo van Inwagen, a Regra A lhe parecia indisputável. Apresentei no segundo capítulo, porém, algumas críticas à regra. Primeiro, motivei uma análise parcial de *responsável por uma proposição*. Em seguida, mostrei que essa análise, *R-Verificador*, corrobora um contra-exemplo à Regra A apresentado por Charles Hermes (2013). Eu não me aprofundi nessa discussão, mas usei o putativo contra-exemplo para mostrar que a principal motivação de van Inwagen a favor da Regra A, sua aparente indisputabilidade, não a motiva no final das contas – já que ela não é realmente indisputável. Passei, então, a segunda motivação de van Inwagen a favor da regra; algumas instâncias confirmatórias. Apresentei a objeção de Hermes (2014) a essas instâncias e rejeitei seu argumento. Argumentei, porém, que as instâncias não tem sucesso em motivar a Regra A. Isso porque a intuitividade de suas validades está baseada em um princípio como R-Capacidade, o qual é dialeticamente inapropriado no contexto do Argumento Direto. Por fim, considerei a sugestão de Justin Capes (2016) de um Argumento Direto Lógico sem a Regra A. Rejeitei a proposta argumentando que o proponente desse argumento estaria cometendo petição de princípio contra o compatibilista. A conclusão deste capítulo foi, portanto, que não há motivações a favor da Regra A.

No terceiro capítulo, discuti a Regra B. Comecei expondo o contra-exemplo à regra proposto por Mark Ravizza (1994), *Erosão*, e discuti a resposta de Michael McKenna (2001; 2008) ao contra-exemplo. McKenna aceita que o contra-exemplo obtém sucesso, mas crê que podemos salvar o Argumento Direto restringindo a regra de transferência de não-responsabilidade a casos com apenas uma cadeia causal. Eu apontei um problema com essa resposta: ou seu proponente se compromete com a existência de apenas uma cadeia causal nos mundos relevantes para a discussão, ou o Argumento Direto precisa ser complementado com um segundo argumento, que mostraria que a não-responsabilidade por uma cadeia causal transfere para as outras cadeias causais. Após essa discussão, apresentei uma defesa da Regra B a partir de *R-Verificador* e mostrei onde *Erosão* falha como um contra-exemplo se assumirmos o princípio. Por fim, discuti a principal objeção à Regra B (e aos Argumentos Diretos como um todo): O incompatibilista

falhou em motivar o bastante a Regra B, de modo que o Argumento Direto não é nem *prima facie* uma ameaça ao compatibilismo. Apresentei o modo como McKenna defende essa tese. Em vez de discutir a literatura gerada pelos argumentos de McKenna, eu propus minha própria defesa dessa tese. Mostrei que os proponentes das instâncias confirmatórias da Regra B apelam a considerações causais – que não podem ser usadas a favor da regra sem uma justificativa teórica – para motivar a intuitividade das inferências em questão. Se estou certo quanto a isso, o compatibilista tem uma teoria de erro para a aparente intuitividade dessas inferências, elas parecem válidas por causa dessas considerações causais, não por causa da validade da Regra B. Levantei três possíveis soluções a esse problema e argumentei que todos falham. Conclui, assim, que o Argumento Direto é nem *prima facie* uma ameaça ao compatibilismo.

Somando as conclusões obtidas nesses dois capítulos, podemos concluir que o Argumento Direto não é suficientemente motivado, falhando, assim, em ser uma ameaça ao compatibilismo. Em outras palavras, o compatibilista não precisa se preocupar em mostrar onde o argumento falha, já que ele nunca teve uma razão para achar que o argumento tem sucesso. Para que o debate possa prosseguir de maneira proveitosa, é preciso achar uma formulação do Argumento Direto que seja suficientemente bem motivada, de modo a colocar o compatibilismo de origem sob ataque. Vejo dois modos principais de se obter tal versão do argumento.

Primeiro, pode-se manter a ideia de ter um Argumento Direto Lógico, mas motivar as regras dedutivas do argumento a partir de uma análise mais formal do operador modal de não-responsabilidade. Por exemplo, assumindo *R-Verificador*, poderíamos tentar adicionar o operador *NR* numa lógica relevante com semânticas de veridadores, como a proposta por Mark Jago (2020), ou em lógicas deônticas com semânticas de veridadores, como a desenvolvida por Federico Faroldi (2019).

Segundo, pode-se abandonar a ideia de um Argumento Direto Lógico e focar nos Argumentos Diretos Causais. Eu apresentei, em 3.3.1, uma crítica aos Argumentos Diretos Causais; argumentei que eles são nem *prima facie* uma ameaça ao compatibilismo por não haver motivação para aceitar uma das premissas. Como eu disse nesta parte, creio que o jeito de evitar esse problema (e apresentar uma boa versão do Argumento Direto Causal) é ligando esse debate aos debates sobre a natureza da causação. Para que um Argumento Direto Causal tenha sucesso, é preciso que seu proponente aponte quais as propriedades da relação causal que dão suporte ao argumento.

Há duas características comuns em ambas as estratégias. Primeiro, ambas recusam *Determinismo* como a definição correta do determinismo. Isso porque, para a primeira, a relação de determinação – ainda que lógica – não pode ser mero acarretamento e, para a segunda, a relação é causal. Que deveríamos recusar *Determinismo* como a tese a ser debatida é algo defendido por Helen Steward (2021) e Carolina Sartorio (2015). Segundo, ambas as estratégias ligam o debate sobre o Argumento Direto com outros debates em metafísica, como discussões sobre veridadores e sobre causação. Ou seja, para sabermos se temos responsabilidade – e se minha irmã está certa em me culpar por não ter dado um presente de aniversário para ela este ano –,

vamos ter que estudar mais metafísica.

Bibliografia

- Andreoletti, G. and L. Vervoort (2022). Superdeterminism: A reappraisal. *Synthese* 200(5), 1–20.
- Asay, J. (2018). We don't need no explanation. *Philosophical Studies* 175(4), 903–921.
- Bailey, A. M. and J. Rasmussen (2020). A new puppet puzzle. *Philosophical Explorations* 23(3), 202–213.
- Beck, O. (2020). The direct argument is a prima facie threat to compatibilism. *Synthese* 199(1-2), 1791–1817.
- Beebe, H. and A. Mele (2002). Humean compatibilism. *Mind* 111(442), 201–223.
- Campbell, J. K. (2007). Free will and the necessity of the past. *Analysis* 67(2), 105–111.
- Campbell, J. K. (2010). Incompatibilism and fatalism: Reply to loss. *Analysis* 70(1), 71–76.
- Capes, J. (2022). Does the direct argument beg the question? *American Philosophical Quarterly* 59(1), 81–96.
- Capes, J. A. (2016). Incompatibilism and the transfer of non-responsibility. *Philosophical Studies* 173(6), 1477–1495.
- Earman, J. (1986). *A Primer on Determinism*. D. Reidel.
- Esfeld, M. (2019). Super-humeanism and free will. *Synthese*, 1–14.
- Faroldi, F. L. G. (2019). Deontic modals and hyperintensionality. *Logic Journal of the IGPL* 27(4), 387–410.
- Fischer, J. M. (2004). 9 the transfer of nonresponsibility. In J. K. Campbell (Ed.), *Freedom and Determinism*. Cambridge Ma: Bradford Book/Mit Press.
- Fischer, J. M. and M. Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.
- Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy* 66(23), 829–839.

- Goldstein, S. (2021). Bohmian Mechanics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). Metaphysics Research Lab, Stanford University.
- Hermes, C. (2013). A Counterexample to A. *Philosophia (United States)* 42(2), 387–389.
- Hermes, C. (2014). Truthmakers and the direct argument. *Philosophical Studies* 167(2), 401–418.
- Huemer, M. (2000). Van inwagen’s consequence argument. *Philosophical Review* 109(4), 525–544.
- Jago, M. (2020). Truthmaker semantics for relevant logic. *Journal of Philosophical Logic* 49(4), 681–702.
- Kearns, S. (2011). Responsibility for necessities. *Philosophical Studies* 155(2), 307–324.
- Lampert, F. and P. Merluzzi (2021a). Counterfactuals, counterfactuals, and free choice. *Philosophical Studies* 178(2), 445–469.
- Lampert, F. and P. Merluzzi (2021b). How (not) to construct worlds with responsibility. *Synthese* 199(3-4), 10389–10413.
- Lewis, D. K. (1981). Are we free to break the laws? *Theoria* 47(3), 113–21.
- Loewenstein, Y. (2016). Why the direct argument does not shift the burden of proof. *Journal of Philosophy* 113(4), 210–223.
- Mason, E. (2019). *Ways to Be Blameworthy: Rightness, Wrongness, and Responsibility*. Oxford University Press.
- McKenna, M. (2001). Source incompatibilism, ultimacy, and the transfer of non-responsibility. *American Philosophical Quarterly* 38(1), 37–51.
- McKenna, M. (2008). Saying good-bye to the direct argument the right way. *Philosophical Review* 117(3), 349–383.
- McKenna, M. (2012). *Conversation & Responsibility*. Oup Usa.
- Mele, A. R. (2021). Direct versus indirect: Control, moral responsibility, and free action. *Philosophy and Phenomenological Research* 102(3), 559–573.
- Merluzzi, P. and G. Maruchi (2019). Transferring non-responsibility. *ethic@ - An international Journal for Moral Philosophy* 18(3), 285–298.
- Nguyen, A. (2020). Unable to do the impossible. *Mind* 129(514), 585–602.
- Pereboom, D. (2001). *Living Without Free Will*. Cambridge University Press.

- Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford University Press.
- Pruss, A. R. (2013). Incompatibilism proved. *Canadian Journal of Philosophy* 43(4), 430–437.
- Ravizza, M. (1994). Semi-compatibilism and the transfer of non-responsibility. *Philosophical Studies* 75(1-2), 61–93.
- Rice, H. (2018). Fatalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 ed.). Metaphysics Research Lab, Stanford University.
- Robinson, M. (2016). Truthmakers, Moral Responsibility, and an Alleged Counterexample to Rule A. *Erkenntnis* 81(6), 1333–1339.
- Ross, W. D. (1930). *The Right and the Good. Some Problems in Ethics*. Clarendon Press.
- Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society* 13, 1–26.
- Sartorio, C. (2015). The problem of determinism and free will is not the problem of determinism and free will. In Alfred R. Mele (Ed.), *Surrounding Free Will: Philosophy, Psychology, Neuroscience*. Oxford University Press USA.
- Sartorio, C. (2017). Frankfurt-style examples. In G. Meghan, T. Kevin, and L. Neil (Eds.), *Routledge Companion to Free Will.*, pp. 179–190. Routledge.
- Schnall, I. M. and D. Widerker (2012). The direct argument and the burden of proof. *Analysis* 72(1), 25–36.
- Schnieder, B. S. (2004). Compatibilism and the notion of rendering something false. *Philosophical Studies* 117(3), 409–428.
- Shoemaker, D. (2015). *Responsibility From the Margins*. Oxford University Press.
- Shoemaker, D. (2020). Responsibility: The state of the question fault lines in the foundations. *Southern Journal of Philosophy* 58(2), 205–237.
- Spencer, J. (2017). Able to Do the Impossible. *Mind* 126(502), 464–497.
- Steward, H. (2021). What is determinism? why we should ditch the entailment definition. In M. Hausmann and J. Noller (Eds.), *Free Will: Historical and Analytic Perspectives*, pp. 17–43. Springer Verlag.
- Strawson, P. (1962). Freedom and resentment. In *Proceedings of the British Academy, Volume 48: 1962*, pp. 1–25.
- Strawson, P. F. (1974). *Freedom and Resentment and Other Essays*. London, England: Routledge.

- Tiberius, V. (2014). *Moral Psychology: A Contemporary Introduction*. Routledge.
- Timpe, K. (2016). Leeway vs. sourcehood conceptions of free will. In K. Timpe, M. Griffith, and N. Levy (Eds.), *Routledge Companion to Free Will*, pp. 213–224. Routledge.
- Tognazzini, N. A. and J. M. Fischer (2017). Incompatibilism and the fixity of the past. In J. Keller (Ed.), *Being, Freedom, and Method: Themes From the Philosophy of Peter van Inwagen*, pp. 140–148. Oxford: Oxford University Press.
- Turner, P. R. and J. Capes (2018). Rule a. *Pacific Philosophical Quarterly* 99(4), 580–595.
- van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press.
- Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oup Usa.
- Warfield, T. A. (1996). Determinism and moral responsibility are incompatible. *Philosophical Topics* 24(2), 215–26.