UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS SOCIAIS E HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

Marcelo Fischborn

APRIMORAR A RESPONSABILIDADE: DIREÇÕES PARA UMA INVESTIGAÇÃO INTERDISCIPLINAR

Santa Maria, RS
2018

Marcelo Fischborn

APRIMORAR A RESPONSABILIDADE: DIREÇÕES PARA UMA INVESTIGAÇÃO INTERDISCIPLINAR

Tese apresentada ao Programa de Pós-Graduação em Filosofia da Universidade Federal de Santa Maria (UFSM, RS) como requisito parcial para obtenção do título de **Doutor em Filosofia**.

Orientador: Prof. Dr. Frank Thomas Sautter

Santa Maria, RS
2018

Ficha catalográfica elaborada através do Programa de Geração Automática
da Biblioteca Central da UFSM, com os dados fornecidos pelo autor.

Fischborn, Marcelo
Aprimorar a responsabilidade: Direções para uma investigação interdisciplinar /
Marcelo Fischborn. - 2018.
117 p.; 30 cm

Orientador: Frank Thomas Sautter
Tese (doutorado) - Universidade Federal de Santa Maria, Centro de Ciências Sociais e Humanas, Programa de Pós-Graduação em Filosofia, RS, 2017

1. responsabilidade moral 2. livre-arbítrio 3. práticas de responsabilidade 4. punição 5. culpa. I. Sautter, Frank Thomas II. Título

Marcelo Fischborn

APRIMORAR A RESPONSABILIDADE: DIREÇÕES PARA UMA INVESTIGAÇÃO INTERDISCIPLINAR

Tese apresentada ao Programa de Pós-Graduação em Filosofia da Universidade Federal de Santa Maria (UFSM, RS) como requisito parcial para a obtenção do título de **Doutor em Filosofia**.

Aprovada em 23 de fevereiro de 2018:

_____
Frank Thomas Sautter, Dr. (UFSM)
(Presidente/Orientador)

_____
Beatriz Sorrentino Marques, Dra. (UFMT) – Videoconferência

_____
Gilberto Lourenço Gomes, Dr. (UENF) – Videoconferência

_____
Róbson Ramos dos Reis, Dr. (UFSM)

_____
Rogério Passos Severo, Dr. (UFRGS)

Santa Maria, RS
2018

AGRADECIMENTOS

Contei com o apoio de inúmeras pessoas, e de maneiras variadas, durante a escrita desta tese. Tenho certeza de que não conseguiria, a esta altura, agradecer a cada uma delas. Por essa razão, gostaria de reservar meus agradecimentos aqui aos que contribuíram de maneira especial com meu trabalho ao longo dos últimos quatro anos. Primeiramente, agradeço ao professor Frank Thomas Sautter por sua dedicação à orientação deste trabalho, pelas inúmeras conversas e discussões, e pela preocupação com meu futuro profissional. Agradeço ao professor Rogério Passos Severo pela orientação que recebi durante o mestrado (e que certamente trouxe para esta pesquisa) e pelo apoio acadêmico, profissional e pessoal que seguiram constantes desde então. Agradeço ao professor Gilberto Lourenço Gomes, pelas várias discussões e contribuições prontamente oferecidas ao longo da escrita e defesa desta tese. À professora Beatriz Sorrentino Marques e ao professor Róbson Ramos dos Reis, agradeço pelos comentários e questionamentos recebidos durante a defesa da tese. Agradeço ao professor Alfred Mele, pela orientação durante o período em que estive na Florida State University e a Lieke Asma pelos comentários a várias partes da tese. Agradeço ao professor Flavio Williges pelas discussões e pelo apoio de várias ordens, e aos professores Leonardo Ribeiro e Silvio Vasconcellos pelas contribuições na ocasião de meu exame de qualificação. Por contribuições variadas em momentos específicos da pesquisa, agradeço a Adina Roskies, Eddy Nahmias, Thiago Santin, Daniela Tocchetto, Thomas Nadelhoffer, Stephen Kearns, Gilson Olegario da Silva, Ricardo di Napoli, Ronai Rocha, Cristina Nunes, César dos Santos, Gabriel Bilhalva, Ana Pivetta e Daniel Albanio. E agradeço aos que contribuíram anonimamente: os participantes do estudo experimental e vários pareceristas. De modo especial, agradeço a Natália Rigue, não só pelo amor e companheirismo, mas também pelas inúmeras conversas e trocas sobre temas relacionados à tese, e a Davi Rigue Fischborn, por trazer alegria a meus dias e por ter me obrigado a pensar sobre o despertar da responsabilidade. Agradeço a meus pais, Clóvis Fischborn e Fátima Fischborn, pelo incentivo e apoio, e também a Bruno Fischborn, Luciano Fischborn, Amarildo Rigue e Neusa Rigue. Finalmente, agradeço à CAPES pela bolsa de doutorado, e à CAPES e à Comissão Fulbright Brasil pela bolsa para estágio nos Estados Unidos.

RESUMO

APRIMORAR A RESPONSABILIDADE: DIREÇÕES PARA UMA INVESTIGAÇÃO INTERDISCIPLINAR

AUTOR: Marcelo Fischborn

ORIENTADOR: Frank Thomas Sautter

As práticas de responsabilidade presentes em nossas vidas cotidianas envolvem, entre outras coisas, padrões sobre como se deve elogiar, culpar ou punir as pessoas por suas ações, bem como atos particulares que seguem esses padrões em maior ou menor medida. Uma questão clássica em filosofia pergunta se os seres humanos podem de fato ser moralmente responsáveis pelo que fazem. Esta tese defende que abordar essa questão clássica é insuficiente se quisermos que a investigação sobre a responsabilidade moral contribua com o objetivo de aprimorar as práticas cotidianas de responsabilidade. Como alternativa, ofereço direções para uma investigação interdisciplinar que defendo estar em melhores condições de promover o objetivo em questão. Essa proposta é defendida ao longo de cinco artigos e uma seção final de discussão. Os quatro primeiros artigos apontam limitações de propostas céticas, as quais negam a existência da responsabilidade moral. O primeiro artigo avalia um argumento cético baseado em resultados da neurociência que pretende mostrar que o livre-arbítrio não existe. Argumento que uma das premissas desse argumento—que diz que eventos no cérebro determinam escolhas—não é justificada pelos resultados disponíveis. O segundo artigo defende que, apesar dos resultados considerados não mostrarem que escolhas sejam determinadas por eventos no cérebro, outros estudos da neurociência poderiam em princípio fazê-lo. O terceiro artigo inicia a discussão de limitações que dizem respeito à implementabilidade de algumas modificações das práticas de responsabilidade sugeridas no interior de propostas céticas. Especificamente, descrevo desafios que tentativas de se reduzir a severidade da punição imposta no âmbito legal provavelmente enfrentariam devido a fatos psicológicos sobre a crença no livre-arbítrio e o desejo de punir. O quarto artigo apresenta resultados de um experimento original que buscou testar uma hipótese sobre o funcionamento da crença no livre-arbítrio e o desejo de punir, qual seja, a hipótese de que o desejo de punir afeta causalmente crenças sobre o livre-arbítrio. Os resultados não apoiaram essa hipótese. Finalmente, o quinto artigo desenvolve o que chamo de modelo de aprimoramento, uma proposta de estruturação de uma investigação interdisciplinar capaz de promover o aprimoramento das práticas cotidianas de responsabilidade. A seção final de discussão mostra como o modelo de aprimoramento supera algumas limitações da discussão recente sobre a existência da responsabilidade moral, a qual inclui não apenas a literatura cética considerada nos artigos anteriores, mas também contribuições que afirmam a existência da responsabilidade moral e do livre-arbítrio. A proposta central desta tese, portanto, diz que é possível reorganizar a investigação sobre a responsabilidade moral de uma maneira que lhe permita melhor promover o objetivo de aprimorar as práticas cotidianas de responsabilidade.

Palavras-chave: Responsabilidade Moral. Livre-arbítrio. Práticas de Responsabilidade. Punição. Culpa. Elogio.

ABSTRACT

ENHANCING RESPONSIBILITY: DIRECTIONS FOR AN INTERDISCIPLINARY INVESTIGATION

AUTHOR: Marcelo Fischborn

ADVISOR: Frank Thomas Sautter

Responsibility practices that are part of our daily lives involve, among other things, standards about how one should praise, blame, or punish people for their actions, as well as particular acts that follow those standards to a greater or lesser extent. A classical question in philosophy asks whether human beings can actually be morally responsible for what they do. This dissertation argues that addressing this classical question is insufficient if one wants the investigation of moral responsibility to serve the goal of improving ordinary responsibility practices. As an alternative, I offer directions for an interdisciplinary investigation that I take to be in a better position to promote that goal. My argument is developed in five articles and a discussion section. The first four articles describe limitations of skeptical views, which deny the existence of moral responsibility. The first article assesses a skeptical argument based on results from neuroscience that intends to show that there is no free will. I argue that a premise in the argument—which says that choices are determined by events in the brain—is not supported by the available results. The second article argues that, despite the fact that existent results do not show that choices are determined by brain events, further studies in neuroscience could in principle do that. The third article begins the discussion of limitations that concern the implementability of some of the changes in responsibility practices recommended in skeptical approaches. Specifically, I describe challenges that attempts to reduce the severity of legal punishment are likely to face due to psychological facts about belief in free will and desire to punish. The forth article presents results from an original experiment that sought to test a hypothesis about the workings of belief in free will and the desire to punish, namely the hypothesis that the desire to punish causally affects beliefs about free will. Results failed to support the hypothesis. Finally, the fifth article presents what I call the enhancement model, i.e., a proposal about how to structure an interdisciplinary investigation that can promote the enhancement of ordinary responsibility practices. The final discussion section shows how the enhancement model overcomes some of the limitations of recent discussions about the existence of moral responsibility, which includes not just the skeptical views considered in earlier articles, but also views that affirm the existence of moral responsibility and free will. The central claim of this dissertation, therefore, is that the investigation of moral responsibility can be rearranged so as to further the goal of improving ordinary responsibility practices.

Keywords: Moral Responsibility. Free Will. Responsibility Practices. Punishment. Blame. Praise.

SUMÁRIO

INTRODUÇÃO

Os seres humanos são moralmente responsáveis por suas ações? O interesse nesta questão, que tem alimentado a discussão filosófica há milênios, não é apenas teórico. A pergunta também revela um interesse no valor de um conjunto de práticas que adotamos comumente em nossas vidas cotidianas. Ao perguntar se os seres humanos são moralmente responsáveis, queremos saber se a maneira como colocamos a ideia de responsabilidade em prática —como quando elogiamos, culpamos ou desejamos a punição de alguém—é moralmente aceitável. Mais do que isso, a discussão sobre a responsabilidade moral traz ainda consigo um interesse no aprimoramento das práticas cotidianas. Pois, nos casos em que a investigação leva à conclusão de que essas práticas são em alguma medida problemáticas, surge naturalmente uma preocupação com sua modificação e aperfeiçoamento.

A tese central que unifica este trabalho envolve um componente negativo e outro positivo. O componente negativo diz que a discussão filosófica tradicional sobre a existência da responsabilidade moral é insuficiente para se promover o aprimoramento das práticas de responsabilidade tal como se apresentam em nossas vidas cotidianas. O componente positivo desenvolve, como alternativa, uma proposta de discussão interdisciplinar que considero mais bem equipada para realizar a tarefa em questão. Os elementos que permitirão defender essa proposta surgirão ao longo dos cinco artigos que compõem o trabalho. Por esses artigos terem sido compostos como contribuições independentes, o argumento que os articula só será explicitamente desenvolvido na seção de Discussão que lhes sucede. O restante desta introdução tem dois objetivos principais. O primeiro é apresentar alguns conceitos centrais que serão pressupostos no decorrer do trabalho e o modo como a discussão sobre a existência da responsabilidade moral tem se dado nas últimas décadas (seções 1 a 3). O segundo é delinear o conteúdo dos cinco artigos e como contribuem com a defesa da tese central proposta (seção 4)

1 Responsabilidade moral, atribuição de culpa e punição

O que queremos dizer quando dizemos que uma pessoa é responsável por certo acontecimento? Suponhamos que um torcedor tenha agredido um torcedor rival durante uma partida de futebol, causando-lhe certos ferimentos, e que depois de um julgamento o agressor seja considerado culpado pelo mal causado à vítima e condenado a pagar algumas cestas básicas. Esse

caso fictício certamente envolve práticas que podemos associar à responsabilidade moral. Mas em que consistem exatamente?

De acordo com um entendimento que remonta ao menos até Aristóteles, a responsabilidade tem a ver com a adequação de certas maneiras de responder a uma pessoa em virtude de algo que tenha feito.[1] Mais precisamente, dizer que um agente é responsável por uma certa ação, nessa concepção, significa dizer que é adequado responder ao agente de uma certa maneira por causa de sua ação (ver, por exemplo, Zimmerman, 2015; Eshleman, 2016). Mas de que respostas estamos falando e em que sentido são ditas adequadas?

Pode-se separar os tipos de respostas associados à responsabilidade moral em dois grupos, dependendo da valência moral das ações a que se dirigem. Se uma ação é moralmente negativa (errada ou condenável) os exemplos paradigmáticos de respostas incluem a atribuição de culpa, a condenação, a censura, o repúdio ou mesmo alguma forma de punição. Se a ação é positiva, por outro lado, a resposta pode ser o elogio, a gratidão, o louvor ou ainda uma premiação ou gratificação. Dizer que uma pessoa é moralmente responsável por uma ação, portanto, é dizer que é apropriado dirigir-lhe pelo menos uma das respostas de um dos grupos considerados, sendo que o valor da ação (positivo ou negativo) determina a qual grupo a resposta apropriada pertence. Noto que, ao fazer essa divisão, não quero sugerir que toda ação precise corresponder a um desses dois grupos. É possível que haja ações neutras ou mesmo ambíguas, incluindo, por exemplo, aspectos positivos e negativos. Não discutirei nesta tese se há alguma resposta característica da responsabilidade moral que seja adequada para ações desses tipos.

Mas o que significa dizer que alguma resposta é *adequada*? Suponhamos que seja adequado atribuir culpa ao agressor do caso previamente considerado. Em que poderia consistir essa adequação? Ao longo da história da filosofia, encontramos duas interpretações principais da noção de adequação, sendo uma delas de tipo consequencialista e a outra baseada na noção de merecimento (Eshleman, 2016, seção 1). De acordo com uma das interpretações, dizer que é apropriado culpar um agressor, por exemplo, significa dizer que podemos esperar *boas consequências* da atribuição de culpa. Assim, culpar o agressor poderia diminuir as chances de que repita o comportamento condenável no futuro, ou mesmo que outros imitem tal comportamento. Uma segunda maneira de interpretar a noção de adequação de uma resposta recorre à noção de *merecimento*. Nessa vertente, dizer que é apropriado atribuir culpa ao agressor equi-

1 Aristóteles apresenta essa concepção no Livro III da *Ética a Nicômaco* (Aristóteles, 1973). A antologia editada por Derk Pereboom (2009) é um bom ponto de partida para acompanhar a história da questão desde a Antiguidade até os dia atuais.

vale a dizer que merece ser culpado simplesmente em virtude daquilo que fez, em um sentido que permite desconsiderar por completo as consequências da responsabilização.[2]

Ao longo deste trabalho, meu foco será na concepção da responsabilidade moral como merecimento. A razão para fazê-lo é que esta posição tem sido central no debate filosófico das últimas décadas.[3] Quando falar no debate sobre a existência da responsabilidade moral, consequentemente, estarei falando sobre o debate que discute se os seres humanos são ou não responsáveis no sentido de merecerem, em função de suas ações, respostas características da responsabilidade moral. Vale a pena enfatizar que a presença ou ausência desse tipo de merecimento é relativa ao tipo específico de resposta que se estiver considerando (Zimmerman, 2015). É possível, por exemplo, que as condições necessárias para que um agressor mereça algum tipo de punição difiram das condições necessárias para que mereça ser culpado ou mesmo para que mereça um tipo diferente de punição. Por essa mesma razão, especificar as condições da responsabilidade moral depende de se entender exatamente em que consiste cada tipo de resposta a ser considerada.

Em que, então, consistem as respostas usualmente associadas à responsabilização? No caso da punição, gostaria de destacar pelo menos três traços definidores (ver, por exemplo, Brooks, 2012, pp. 1-2; Bedau & Kelly, 2015, section 2). Primeiramente, a punição é uma resposta a um ato *condenável* realizado pelo agente a quem se dirige—ou, no caso da punição legal, uma resposta a um ato que viola a lei. Em segundo lugar, a punição é uma resposta que acarreta algum tipo de *ônus* ou dano àquele que é punido. Formas recorrentes de punição incluem multas, restrição de liberdade ou a perda de algum direito. E, em terceiro lugar, a punição é imposta *intencionalmente* à pessoa a ser punida. Se, por exemplo, um juiz batesse seu carro acidentalmente contra o carro do agressor do caso anterior, não diríamos que esse acontecimento, apesar de oneroso para o agressor, contaria como uma punição por sua agressão.

---

2 Há também posições mistas, que combinam a consideração de merecimento e consequências. No caso da responsabilidade moral, esse tipo de posição é defendido, por exemplo, por Manuel Vargas (2013). No caso específico da punição, uma abordagem mista é desenvolvida por H.L.A. Hart (2008, ver também Brooks, 2012, parte II). A investigação que busca o aprimoramento das práticas de responsabilidade que desenvolvo no quinto artigo e na seção de Discussão também abre espaço para a consideração simultânea de merecimento e consequências.

3 Eshleman (2016, seção 2) diz que apesar de desenvolvimentos de concepções que adotam a interpretação consequencialista, "o trabalho sobre o conceito de responsabilidade moral nos últimos cinquenta anos tem se concentrado cada vez mais em: a) oferecer versões alternativas da concepção baseada no merecimento; e b) questionar se há apenas um conceito de responsabilidade moral". A tese de que a interpretação baseada no merecimento é a que melhor captura o interesse tradicional pela responsabilidade moral é defendida, por exemplo, por Pereboom (2014, p. 2; 2001) e Caruso e Morris (2017).

Essas três características parecem centrais para qualquer caracterização da punição, ainda que uma definição completa possa exigir elementos adicionais.

Outros tipos de resposta característicos das práticas de responsabilidade, como o elogio e a culpa, não parecem ser tão bem compreendidos. O caso da culpa, por exemplo, só recentemente começou a receber maior atenção. Coates e Tognazzini (2012) consideram algumas alternativas. Uma sugestão é que a atribuição de culpa consiste em um tipo de ação que expressa indignação e exige um pedido de desculpas—atos de repúdio poderiam envolver alterar o tom da voz em resposta às más ações do culpado (p. 198). Mas também pode-se pensar a atribuição de culpa como um juízo (julgar que o agente agiu mal), como um estado afetivo (raiva pelo agente ter feito o que fez), ou um estado conativo (um desejo de que o agente não tivesse feito o que fez).

Qualquer que seja a concepção adotada sobre punição, culpa ou elogio, o que cumpre destacar neste momento é que, na interpretação da responsabilidade investigada aqui, as condições para que alguém seja moralmente responsável refletirão, no fim das contas, as condições para que um agente mereça culpa, punição, elogio ou alguma outra resposta. Ao falar do ‘debate sobre a existência da responsabilidade moral’, portanto, estarei falando abreviadamente do debate sobre se os seres humanos merecem, sob algum conjunto de condições, elogio, culpa, punição ou algum outro tipo característico de reação. Também ressalto desde já que, como é comum na literatura sobre livre-arbítrio e responsabilidade em geral, estarei na maior parte do tempo focado no aspecto negativo da responsabilidade—i.e., na punição e no repúdio—e não nos aspectos positivos como gratidão e elogio. Isso se justifica em parte porque essas noções parecem ser as que mais pedem por justificação—repudiar, multar ou mesmo prender uma pessoa parece demandar uma justificação moral muito mais urgentemente do que elogiar ou premiar. Por essa razão, pode ser que nem tudo o que for dito aqui tendo a punição e o repúdio em mente possa ser transferido sem ressalvas para, por exemplo, premiações e elogios. Em outras palavras, fica em aberto a possibilidade de que exista algum tipo de assimetria no interior da responsabilidade moral (ver, por exemplo, Wolf, 1980).

2 Condições para a existência da responsabilidade moral

A seção anterior concentrou-se em estabelecer o que significam noções como ‘responsabilidade moral’, ‘punição’ e ‘culpa’. Trata-se de um trabalho sobre a definição desses termos (Zimmerman, 2015). Uma outra parte do trabalho a ser realizado em uma investigação sobre a

existência da responsabilidade moral é estabelecer as condições necessárias e suficientes para que, dado um certo entendimento da responsabilidade e suas respostas características, um agente seja de fato moralmente responsável por alguma ação. Esta seção exemplifica algumas das condições que foram consideradas nessa investigação.

Aristóteles (1973, Livro III) foi pioneiro, novamente, ao propor que a responsabilidade moral envolve dois tipos de condições. Segundo ele, não é apropriado responsabilizar um agente se, por um lado, tiver sido forçado a agir como agiu ou se, por outro, tiver agido por ignorância. Tal como são conceitualizadas contemporaneamente, essas considerações traduzem-se, respectivamente, nas condições de conhecimento e controle da responsabilidade moral (ver, por exemplo, Eshleman, 2016, seção 1). A condição de controle é, de longe, a que tem sido mais discutida na tradição filosófica.[4] Identifica-se comumente a noção de controle em questão com o livre-arbítrio, que é entendido como uma condição necessária para a responsabilidade moral. Em outras palavras, o livre-arbítrio é entendido precisamente como um tipo de controle sobre as próprias ações sem o qual um agente não pode ser moralmente responsável por elas. Alfred Mele, por exemplo, define o livre-arbítrio como uma capacidade para realizar ações livres "ao nível da responsabilidade moral", sendo uma ação livre nesse sentido aquela em que

> se todas as condições independentes da liberdade para a responsabilidade moral por uma ação particular fossem satisfeitas sem que isso bastasse para o agente ser moralmente responsável por ela, a adição da liberdade da ação a esse conjunto de condições implicaria que ele é moralmente responsável por ela. (Mele, 2006, p. 17)

A suposição de que o livre-arbítrio é uma condição necessária para a responsabilidade moral é praticamente universal no debate sobre a responsabilidade moral. Por essa razão, a investigação sobre a existência da responsabilidade moral concentrou-se, em muitos casos, em investigar se o livre-arbítrio existe.

Apesar deste acordo superficial, no entanto, há muito espaço para controvérsia. Uma das controvérsias mais perenes diz respeito a estabelecer se a falsidade da tese do determinismo é ou não uma condição necessária para a existência do livre-arbítrio (e, portanto, da responsabilidade moral). Uma maneira recorrente de apresentar a tese do determinismo é a seguinte:

4 Para um livro recente sobre a condição de conhecimento, ver Robichaud e Wieland (2017).

> Determinismo: Uma descrição completa do estado do universo em qualquer instante t, em conjunção com todas as leis da natureza que são verdadeiras nesse universo, implica todas as proposições verdadeiras nesse universo em instantes posteriores a t.

É frequentemente admitido no debate contemporâneo sobre o livre-arbítrio que a tese do determinismo implica que cada ação ou decisão que um ser humano realize não poderia ser diferente do que foi. Essa implicação depende da suposição de que o passado e as leis da natureza sejam fixos. Assim, a questão da compatibilidade pergunta se pode haver livre-arbítrio se a tese do determinismo for verdadeira. As respostas a essa questão se dividem em dois grandes grupos. Por um lado, o compatibilismo é a tese de que podemos ter livre-arbítrio ainda que a tese do determinismo seja verdadeira. Dennett (1984), Fischer (1994), Fischer e Ravizza (1998) e Nelkin (2011) estão entre os defensores contemporâneos do compatibilismo. O incompatibilismo, por outro lado, é a tese de que não podemos ter livre-arbítrio se a tese do determinismo for verdadeira.

Vale a pena tecer algumas notas sobre como as teses compatibilista e incompatibilista se relacionam com a verdade das teses do determinismo e de que o livre-arbítrio existe. Nem o compatibilismo e nem o incompatibilismo, por um lado, implicam isoladamente que o determinismo seja verdadeiro e tampouco que seja falso. Compatibilismo e incompatibilismo também não implicam que o livre-arbítrio exista ou que não exista.[5] Como questão contingente, entretanto, muitos compatibilistas contemporâneos não emitem juízo sobre a verdade da tese do determinismo e se comprometem com a existência do livre-arbítrio. No caso dos incompatibilistas, por outro lado, a única combinação de teses impossível é defender que temos livre-arbítrio e que a tese do determinismo é verdadeira. Alguns incompatibilistas negam que haja livre-arbítrio. Os incompatibilistas que afirmam a existência do livre-arbítrio (e que rejeitam, portanto, a tese do determinismo) são conhecidos como libertistas. Chisholm (1964), Kane (1996) e O'Connor (2002) estão entre os defensores recentes do libertismo.

Finalizo esta seção com alguns comentários sobre a atitude mais geral diante da tese do determinismo e da disputa entre compatibilistas e incompatibilistas que adoto ao longo desta tese. Primeiramente, noto que a tese do determinismo é altamente exigente. Ela envolve, por exemplo, noções como a descrição *completa* do estado do universo em algum momento e de suas leis, e afirma que *todo* evento é determinado. É duvidoso, para dizer o mínimo, que

5 Em outras palavras, pode-se dizer que a tese do determinismo e a tese de que há livre-arbítrio são ambas independentes tanto do compatibilismo quanto do incompatibilismo.

algo do tipo possa algum dia ser efetivamente alcançado. No debate sobre o livre-arbítrio, por vezes se considera que o determinismo é uma tese cuja verdade poderia ser decidida pela física (ver, por exemplo, Fischer e Ravizza, 1998, pp. 14-15). É disputável que as teorias físicas em voga sejam deterministas (Earman, 1986). E também é disputável que, ainda que o fossem, poderiam por si sós estabelecer o determinismo relevante para a literatura sobre o livre-arbítrio, já que seria necessário supor também alguma forma de redutibilidade de eventos mentais (como decisões) a eventos físicos. Sem esta última suposição, poderia haver indeterminação no âmbito dos eventos mentais ainda que eventos físicos operassem de modo determinista (ver Steward, 2008; Davidson, 1970; Fischborn, 2014; Lycan, 2015).

Uma das contribuições que desenvolvo nesta tese diz que em um domínio muito mais restrito—a saber, que fala apenas da relação entre eventos no cérebro e decisões—não encontramos suporte para a existência de relações deterministas (ver o primeiro artigo). Por essa razão, considero extremamente improvável que a tese do determinismo seja verdadeira. Essa consideração também influencia minha atitude em relação à questão sobre a (in)compatibilidade entre livre-arbítrio e determinismo. Ao longo da tese, não me comprometo com o compatibilismo e nem com o incompatibilismo. Porque considero a verdade da tese do determinismo implausível, considero também que a questão da compatibilidade é pouco urgente para o debate sobre a existência do livre-arbítrio e da responsabilidade moral. Além disso, se a tese do determinismo for falsa—e, mais precisamente, se escolhas humanas forem indeterminadas—mesmo os compatibilistas poderão ter de revisar suas teorias. Muitos dos compatibilistas que afirmam a existência do livre-arbítrio afirmam que a indeterminação buscada pelos libertistas não aumenta o controle que os agentes podem ter sobre suas ações. Ora, se a indeterminação for um fato sobre como o mundo é, então também os compatibilistas terão de dizer por que essa indeterminação não mostra que o livre-arbítrio não existe. Essa atitude mais geral sobre a tese do determinismo e a questão da compatibilidade não é propriamente defendida nesta tese, mas explicitá-la aqui poderá auxiliar na compreensão das preocupações que orientaram a sua escrita.

3 Ceticismo sobre a existência da responsabilidade moral

A posição mais comum entre os que se dedicam aos tópicos da responsabilidade moral e do livre-arbítrio é que os seres humanos são capazes de exercerem sua liberdade e dignos de responsabilização por pelo menos algumas das ações que realizam. Uma maneira de mapear o

debate sobre a existência da responsabilidade moral, portanto, é a partir do que têm dito aqueles que vão na contramão da posição majoritária. Nesta seção ilustrarei brevemente dois argumentos em favor da tese de que os seres humanos nunca são moralmente responsáveis pelo que fazem. Ambos os argumentos serão consideradas em maior detalhe ao longo do trabalho.

Uma primeira variante de ceticismo sobre a responsabilidade moral é derivada das suspeitas sobre a existência do livre-arbítrio motivadas por um conjunto de estudos publicados na década de 1980 pelo neurocientista Benjamin Libet e colaboradores (Libet et al., 1983; Libet, 1999). Sabia-se, à época, que movimentos espontâneos simples como flexionar o dedo são antecedidos por um padrão de atividade neural conhecido como potencial de prontidão. Libet buscou investigar a relação temporal entre o início do potencial de prontidão que antecede um movimento espontâneo e a decisão de se realizar esse movimento. A descoberta relativamente surpreendente de Libet foi que o potencial de prontidão antecedeu em cerca de 350 milissegundos o momento em que os participantes do estudo relatavam ter decidido realizar esses movimentos. Libet afirmou que seus resultados restringiam o tipo de controle que podemos ter sobre nossas ações e sugeriu que o controle consciente sobre as ações poderia ser exercido não no início da preparação da ação, mas sob a forma de um poder de veto que poderia permitir ou interromper uma ação cuja preparação já havia se iniciado. Embora o próprio Libet não tenha pensado que seus resultados ameaçassem a existência do livre-arbítrio, outros autores foram mais longe. Por exemplo, Haynes (2011) e Misirlisoy e Haggard (2014) argumentaram que os resultados de Libet—e de um estudo similar realizado mais recentemente por Soon et al. (2008)—ameaçam a existência do livre-arbítrio porque mostram que padrões de atividade neural anteriores ao momento da decisão são em alguma medida preditivos da decisão da qual o sujeito só terá consciência posteriormente. Esse argumento será considerado em maiores detalhes no decorrer desta tese. Neste momento, é suficiente destacar que se trata de um argumento cético que nega a existência do livre-arbítrio (e, portanto, também da responsabilidade moral) baseado em estudos sobre o funcionamento das capacidades humanas de decisão.

Derk Pereboom (2001, 2014) desenvolveu um outro tipo de argumento cético sobre o livre-arbítrio e a responsabilidade moral. Um primeiro passo do argumento é a tese de que o livre-arbítrio exige uma capacidade conhecida como ‘causalidade do agente’ (2014, capítulo 2). A noção de causalidade do agente envolve sutilezas metafísicas. Se um agente tivesse essa capacidade, suas ações e decisões seriam causadas pelo próprio agente *enquanto uma substância*, e não simplesmente por *eventos* ou estados que acontecessem *no* agente, como crenças

ou desejos. Como é comum entre aqueles que defendem que a responsabilidade moral exige a causalidade do agente, a alegação é que essa capacidade possibilitaria ao agente exercer um tipo mais robusto de controle sobre aquilo que faz. Como segunda premissa, Pereboom defende também que as teorias físicas atuais tornam implausível que sejamos agentes com o tipo de poder causal em questão (2014, capítulo 3). Finalmente, Pereboom rejeita que teorias compatibilistas do livre-arbítrio possam oferecer uma alternativa satisfatória para mostrar que o livre-arbítrio existe (capítulo 4) e, como consequência, conclui que é implausível que sejamos agentes com as capacidades para escolha e ação que seriam necessárias para que pudéssemos alguma vez merecer elogio, culpa ou punição por nossas ações. O ceticismo de Pereboom foi endossado recentemente por Gregg Caruso (2016), e ambos os autores têm se dedicado a explorar as consequências do ceticismo sobre o livre-arbítrio e a responsabilidade moral a respeito das práticas cotidianas de responsabilidade. No decorrer deste trabalho, não me ocuparei diretamente com o argumento cético de Pereboom e Caruso, mas com a discussão das implicações do argumento para o modo como deveríamos revisar as práticas cotidianas de responsabilidade. Minha atitude geral a esse respeito é de simpatia pelas questões práticas que os céticos colocam em discussão, mas de suspeita quanto ao ceticismo sobre a existência do livre-arbítrio e da responsabilidade moral que os motiva.

4 Visão geral da proposta

O restante deste trabalho compõe-se de cinco artigos e uma seção final de discussão. Como disse antes, a tese que unifica esses artigos tem dois componentes. O primeiro diz que a discussão especificamente sobre a existência da responsabilidade moral é insuficiente para promover o objetivo de aprimorar as práticas cotidianas de responsabilidade e o segundo oferece uma proposta de investigação alternativa. Nesta seção, descrevo o conteúdo dos cinco artigos e como contribuem com o argumento desenvolvido ao final.

Os dois primeiros artigos situam-se no interior do debate sobre o impacto dos estudos em neurociência mencionados na seção anterior para a discussão sobre a existência do livre-arbítrio. O artigo 1 (“Libet style experiments, neuroscience, and libertarian free will”) busca examinar se experimentos como os de Libet mostram, ou poderiam mostrar, que nossas escolhas são determinadas pelos potenciais de prontidão de uma maneira que ameace o libertismo. Como disse antes, alguns cientistas argumentam que os resultados disponíveis afetam essas concepções ao revelarem que padrões neurais são preditivos de escolhas específicas que serão

posteriormente realizadas por uma pessoa (Haynes, 2011; Misirlisoy & Haggard, 2014). Por outro lado, alguns filósofos argumentaram que resultados do tipo em questão não poderiam sequer em princípio ameaçar as concepções libertistas (Roskies, 2006; Nahmias, 2014). Argumento que ambas as posições são exageradas. Contrariando Nahmias e Roskies, defendo que a neurociência poderia em princípio mostrar que escolhas são determinadas por eventos neurais de uma maneira que feriria os pressupostos de pelo menos algumas teorias sobre a natureza do livre-arbítrio (a saber, algumas teorias libertistas). Apesar disso, contrariando agora Haynes, Haggard e Misirlisoy, argumento que os resultados disponíveis parecem estar longe de mostrar que tal tipo de determinação de fato ocorra.[6]

O artigo 2 (“Neuroscience and the possibility of locally determined choices: Reply to Adina Roskies and Eddy Nahmias”) continua essa discussão e busca responder às críticas que Roskies e Nahmias apresentaram em um comentário ao primeiro artigo (Roskies & Nahmias, 2017). Apesar de concordarem que experimentos como os de Libet não tenham mostrado que escolhas sejam determinadas no sentido relevante para o debate, eles insistem que a neurociência não poderia sequer em princípio mostrar que tal tipo de determinação ocorre. Em minha resposta, procuro mostrar que seus argumentos não fazem mais do que mostrar que a determinação relevante é muito improvável dado o conhecimento sobre a mente e o cérebro atualmente disponível—algo com que estou de acordo—mas que essa baixa probabilidade não acarreta a impossibilidade em princípio. Por essa razão, mantenho a tese de que, apesar de improvável que de fato venha a fazê-lo, a neurociência poderia em princípio mostrar que nossas escolhas são determinadas por certos padrões de atividade neural anteriores, em um sentido que falsificaria certas teorias libertistas.

Os artigos 3 e 4 buscam avaliar o que poderíamos chamar de “plausibilidade psicológica” de algumas implicações do ceticismo sobre a responsabilidade moral para as práticas de responsabilidade cotidianas. O artigo 3 (“How should free will skeptics pursue legal change?”) explora como certas implicações do ceticismo de Pereboom e Caruso para as práticas de punição aplicadas no âmbito legal poderiam vir a ser implementadas.[7] A principal con-

6 Uma questão relacionada, que não será abordada em detalhes nesta tese, é se os potenciais de prontidão, mesmo que não *determinem* a ocorrência de escolhas, poderiam ainda assim causar, de uma maneira não-determinista, escolhas ou ações. Essa questão continua um tópico de intenso debate (ver, por exemplo, Alexander et al., 2015; Trevena e Miller, 2010; Gomes, 2010; Schurger, Sitt & Dehaene, 2012; Marques, 2017).

7 É comum distinguir entre os âmbitos moral e legal, e, mais especificamente, entre a responsabilidade moral e a responsabilidade legal (ver, por exemplo, Duff, 2009). Em alguns momentos desta tese, focarei em questões que envolvem a relação entre os dois âmbitos. Faço isso por duas razões. Primeiramente, porque as abordagens céticas de Pereboom e Caruso que estarei considerando envolvem propostas

sequência desse tipo de ceticismo é que deveríamos buscar reduzir, tanto quanto possível, a severidade da punição dirigida àqueles que realizam crimes. Argumento, com base em resultados empíricos sobre o desejo de punir, a crença no livre-arbítrio e a criminalidade, que é implausível pensar que as pessoas em geral eventualmente reduzirão sua crença no livre-arbítrio e que, devido a isso, apoiarão a redução da severidade da punição que o ceticismo implica. Por essa razão, proponho que o cético sobre o livre-arbítrio e a responsabilidade moral deveria buscar uma via alternativa para a implementação do seu projeto. Segundo essa alternativa, a maneira mais viável de se buscar a redução da severidade da punição deveria buscar a redução do desejo de punir que surge naturalmente quando se é exposto ao comportamento imoral ou criminoso. Uma maneira de fazer isso seria justamente apresentar políticas efetivas de redução da criminalidade, políticas estas que os próprios Pereboom e Caruso também defendem, mas que não se seguem estritamente do ceticismo sobre o livre-arbítrio e a responsabilidade moral.

O artigo 4 (“Does desire to punish affect belief in free will? An experimental report”) apresenta os resultados de um experimento em psicologia social que buscou contribuir com o estudo das causas da crença no livre-arbítrio. Em um estudo anterior, Clark et al. (2014) mostraram que a exposição a uma ação imoral leva a um fortalecimento da crença no livre-arbítrio, e sugeriram que um desejo aumentado de punir explicaria esse resultado. Essa hipótese contraria a expectativa tradicional de que nosso desejo de punir deveria ser apenas influenciado por, e não influenciar, a crença no livre-arbítrio. O experimento relatado no quarto artigo buscou testar experimentalmente a hipótese aludida por Clark et al. (2014). Os resultados encontrados não confirmaram a hipótese. Participantes induzidos a desacreditar em possíveis efeitos positivos da punição acabaram por recomendar punições de duração menor para um criminoso fictício (relativamente a um grupo de controle), mas ainda assim mantiveram os mesmos níveis de crença no livre-arbítrio. Esse resultado é compatível com a hipótese de que o desejo de punir, ao menos quando reduzido por razões de tipo consequencialista, não afeta causalmente a crença no livre-arbítrio. O resultado sugere ainda que a modificação de alguns aspectos de nossas práticas de responsabilidade (especificamente, o nível de punição considerado apropriado em alguma situação) pode ser aprovada sem que a crença no livre-arbítrio seja reduzida.

---

sobre a punição no âmbito legal. E, em segundo lugar, porque o desejo de punições que faz parte de nossa moralidade pode ter um papel para a legitimação e estabilização do sistema legal (ver, por exemplo, de Keijser e Elffers, 2009).

No argumento desenvolvido na seção de Discussão, os quatro artigos iniciais ajudam a mostrar as limitações das posições céticas em promover o aprimoramento das práticas de responsabilidade cotidianas. Ao rejeitar um argumento contra a existência do livre-arbítrio baseado em resultados da neurociência, os dois primeiros artigos somam-se a uma literatura mais ampla que rejeita o ceticismo e afirma a existência do livre-arbítrio e da responsabilidade moral. Esse resultado ameaça o potencial das teorias céticas de oferecer recomendações visando o aprimoramento das práticas cotidianas. Pois, se o ceticismo que motiva essas recomendações é falso, então fica difícil defender que as modificações propostas de fato representariam um *avanço* relativamente às práticas de responsabilidade atuais. Os artigos 3 e 4, por outro lado, oferecem razões para se questionar o potencial do ceticismo de levar as pessoas em geral a endossar o tipo de alteração das práticas de responsabilidade que propõe, especificamente no que diz respeito à redução da severidade da punição. Na seção de Discussão, argumentarei que as limitações das posições céticas se estendem também a posições que afirmam a existência da responsabilidade moral e do livre-arbítrio.

Esses problemas abrem espaço para minha proposta de uma maneira alternativa de investigar a responsabilidade moral. Apresento essa proposta no artigo 5 (“Questions for a science of moral responsibility”) que tem como objetivo principal entender como a ciência pode contribuir com a investigação sobre a responsabilidade moral. Depois de examinar como resultados científicos têm sido introduzidos na discussão filosófica sobre a responsabilidade moral, o artigo desenvolve o que chamei de ‘modelo de aprimoramento’. O modelo de aprimoramento propõe uma maneira de integrar filosofia e ciência em uma investigação cujo fim último é promover o aprimoramento das práticas cotidianas de responsabilidade. Nessa investigação, a ciência tem duas contribuições principais a oferecer. Primeiramente, cabe à ciência descrever o funcionamento das práticas de responsabilidade em nossas vidas cotianas, ou seja, descrever as causas e efeitos de comportamentos como culpar, elogiar ou punir alguém. Em segundo lugar, a ciência pode descrever maneiras pelas quais se poderia efetivar modificações de eventuais aspectos problemáticos do funcionamento dessas práticas. Essa segunda tarefa da ciência, no entanto, depende da contribuição de uma investigação de tipo normativo que permita avaliar o funcionamento das práticas cotidianas de responsabilidade. Segundo o modelo de aprimoramento, a filosofia tem uma contribuição central a oferecer nessa investigação normativa.[8] Como se pode ver, o modelo de aprimoramento propõe direções gerais para uma in-

8 Essa sugestão recebe alguns detalhes adicionais na seção de Discussão.

vestigação interdisciplinar que envolve a interação não só de disciplinas científicas diferentes, mas também a interação entre a investigação de tipo normativo e a investigação de tipo mais puramente descritivo. E, se o argumento central desta tese estiver correto, essa investigação promete estar mais bem equipada para promover o aprimoramento das práticas cotidianas de responsabilidade do que a investigação filosófica tradicional focada exclusivamente na existência da responsabilidade moral.

ARTIGO 1: LIBET-STYLE EXPERIMENTS, NEUROSCIENCE, AND LIBERTARIAN FREE WILL*

Abstract: People have disagreed on the significance of Libet-style experiments for discussions about free will. In what specifically concerns free will in a libertarian sense, some argue that Libet-style experiments pose a threat to its existence by providing support to the claim that decisions are determined by unconscious brain events. Others disagree by claiming that determinism, in a sense that conflicts with libertarian free will, cannot be established by sciences other than fundamental physics. This paper rejects both positions. First, it is argued that neuroscience and psychology could in principle provide support for milder deterministic claims that would also conflict with libertarian free will. Second, it is argued that Libet-style experiments—due to some of their peculiar features, ones that need not be shared by neuroscience as a whole—currently do not (but possibly could) support such less demanding deterministic claims. The general result is that neuroscience and psychology could in principle undermine libertarian free will, but that Libet-style experiments have not done that so far.
Keywords: Benjamin Libet; determinism; free will; libertarianism; neuroscience

1 Introduction

Recent discussions about free will and cognitive science (especially neuroscience) were largely influenced by some intriguing and controversial experiments conducted by Benjamin Libet and others in the 1980s (see Libet, Gleason, Wright & Pearl, 1983; Libet, Wright, Feinstein & Pearl, 1982; and Libet, 1999). It was known at the time that a specific sort of neural activity called ‘readiness potential’ (RP) preceded voluntary movements (Kornhuber & Deecke 1965). Libet sought to investigate the temporal relation between RPs, movements, and the moment when subjects become conscious of wanting to move. He found that RPs start on average approximately 350 milliseconds before the subjects’ reported times of a conscious urge or wish to flex a finger, and approximately 500 milliseconds before actual movement (Libet et al., 1983; Libet 1999).[1] Libet concluded that the voluntary acts under examination

* The Version of Record of this manuscript has been published and is available in *Philosophical Psychology* 29.4 (2016): 494–502. <http://www.tandfonline.com/doi/full/10.1080/09515089.2016.1141399>.

1 This information refers only to what Libet calls ‘type II’ RP, i.e., RPs preceding movements for which subjects reported no previous planning of the moment to move. For other conditions, see Libet et al. (1982, 1983). It is worth noting that both the specific measurements and the implications for free will of Libet’s results are a matter of dispute. On the former, mentioned difficulties include the effects of instructions and training during the experiments, and subjects’ ability to accurately report the time of de-

are initiated unconsciously in the brain. More recently, Soon, Brass, Heize, and Heynes (2008) found neural activity that predicts which of two buttons a subject will push 7 seconds (or even 10 seconds) before the subject has consciously decided between the options. Although the accuracy of the prediction is less than roughly 60%, the authors conclude that conscious decisions are determined by unconscious neural activity.

There has been considerable disagreement about the significance of this kind of result for debates about the existence of free will. The aim of this paper is to assess these divergences with regard to a particular conception of free will, namely, libertarian free will. For the purposes of this paper, let us understand as 'libertarian' any conception that holds that free will is incompatible with determinism. ('Determinism' will be characterized in section 3.) Below, I start by framing current disputes on the impact of Libet-style experiments on libertarian free will (section 2), and then I argue for two theses. The first is that, contrary to what some have defended, neuroscience and psychology can, in principle, establish modest deterministic claims that might threaten libertarian free will (section 3). The second is that Libet-style experiments have not so far established that sort of claim, though they could in principle (section 4). Neuroscience and psychology could in principle undermine libertarian free will, but Libet-style experiments have not yet done that.

2 Disputes on the impact of Libet-style experiments on libertarian free will

Some people have interpreted results from Libet-style experiments as a straightforward case against free will. Haynes, the senior author in Soon et al. (2008), for example, describes the challenge as follows:

> our and Libet's findings do address one specific intuition regarding free will, that is the naïve folk-psychological intuition that at the time when we make a decision the outcome of this decision is free and not fully determined by brain activity. (Haynes, 2011, p. 92)

Similarly, Misirlisoy and Haggard describe a

> personal experience [that] provides a powerful impetus for the folk concept of free will. We consciously decide on a course of action and only then we do carry out the relevant actions to fulfill it. When presented with a choice of two options, we may think about them, and then we perform a conscious selection be-

cisions (see, e.g., Gomes, 1998; Banks & Isham, 2011; and Maoz, Mudrik, Rivlin, Ross, Mamelak et al., 2015). Questions related to the latter point include the representativeness and significance of finger flexions for free will, the precise nature of the mental phenomena investigated, and various others (see, e.g., Mele, 2006, 2009, the essays in Mele, 2015, in Sinnott-Armstrong & Nadel, 2011, in Part II of Pockett, Banks & Gallagher, 2006, and most of what is discussed below).

> tween them by exercising our will. In this sense, our will is experienced as free. (Misirlisoy & Haggard, 2014, p. 37)

And they add—partly on the basis of the results in Soon et al. (2008)—that neuroscience has "called this intuition into question, by showing that unconscious activity in the brain preceding our intention—activity that we are never aware of—predicts the emergence of that specific intention to act" (Misirlisoy & Haggard, 2014, p. 38).

The reasoning in these passages seems to be as follows. First, our intuitive conception of free will is said to require that our decisions are not determined by previous (allegedly unconscious) activity in the brain; in other words, a libertarian view of free will is considered a common intuition. But, second, Libet-style experiments are said to undermine this intuition. As a consequence, our intuitive, libertarian notion of free will is an illusion.

Such confidence in the implications of neuroscience for the free will debate has been challenged by others, notably in philosophy. Nahmias (2014b, p. 5) offers the following argument schema as a means of clarifying how Libet-style experiments and other results from cognitive science can have an impact on the debate:

1. Free will requires that X is not the case.
2. Science is showing that X is the case (for humans).
3. Thus, science is showing that humans lack free will.

He then analyzes a group of candidates for "X", the first of which is "determinism." He gets the following argument (see Nahmias, 2014b, p. 5):

D1. Free will requires that determinism is not the case.
D2. Science is showing that determinism is the case (for humans).
D3. Thus, science is showing that humans lack free will.

Premise D1 states a form of incompatibilism, and given premise D2, the argument as a whole is a form of hard determinism: free will requires determinism to be false, but since determinism is true, there is no free will.

Nahmias denies, first, that Libet-style experiments can support premise D2 because they would not be in a position to establish determinism such as it is understood by incompatibilists:

> In incompatibilist arguments, determinism is defined as the thesis that a complete description of a system (e.g., the universe) at one time and of all the laws that govern that system logically entails a complete description of that system at any future time. (2014b, p. 6)

Nahmias says that this sort of determinism "requires a closed system," and then objects that the brains and behaviors studied by cognitive scientists are not closed systems. He adds that results such as those in Soon et al. (2008) "do *not* show that, given prior events … certain decisions or behavior *necessarily* occur" (Nahmias, 2014b, p. 6).

Roskies (2006) offers a similar argument for the claim that neuroscience cannot tell whether the universe is, at a fundamental level, determinist. She argues that observed determinism or indeterminism at one level of description cannot be taken as evidence that another level is deterministic or indeterministic. For example, neuroscientists could come to the conclusion that brains are indeterministic. But, due to the possibility of deterministic chaos, she says, "apparent indeterminism in one level of description is entirely compatible with determinism at the fundamental physical level" (2006, pp. 420–421). In this way, Roskies accepts that "neuroscience can indicate … that, regardless of whether or not the universe is deterministic, the brain effectively is" (2006, p. 421), but insists that it is determinism at the fundamental physical level that is critical for the traditional debate about free will.

Before going ahead, I should mention that Nahmias and Roskies also doubt premise D1 in the argument above. Nahmias argues that cognitive scientists cannot simply assume that premise D1 accurately represents philosophers' and laypersons' views. According to him, most philosophers as well as most laypersons seem to be *compatibilists*. Regarding philosophers' beliefs, we have evidence from Bourget and Chalmers' (2013) online survey. And, regarding laypersons' beliefs, Nahmias mentions results in experimental philosophy by himself and colleagues (Nahmias, Coates, & Kvaran, 2007; Nahmias, Morris, Nadelhoffer, & Turner, 2006; see also note 5). And Roskies (2006, p. 422), partly drawing on the same experimental data, also doubts that neuroscience could have an impact on ordinary practices of responsibility, even if it could affect ordinary conceptions about free will.

In the following sections, I do not focus on the question whether compatibilism is conceptually stronger, nor on whether it represents common thought more accurately than incom-

patibilism. Instead, the focus is on whether Libet-style experiments (and neuroscience, more generally) are, or can be, a threat to free will *if* incompatibilism is correct, or, as we may put it, whether Libet-style experiments (and neuroscience) do, or could, undermine a libertarian conception of free will. This is precisely what is at issue: The scientists mentioned above claim that such experiments actually exclude libertarian free will; the philosophers mentioned claim that neuroscience could not do that in principle.

3 Neuroscience, determinism, and libertarian free will

Let us begin by assessing the claim that neuroscience cannot establish a sort of determinism that is incompatible with libertarian free will. It is true, as Nahmias says, that in discussions between compatibilists and incompatibilists, determinism is often characterized as a thesis concerning the workings of the universe as a whole. In that sense (let us label it 'universal determinism'), the thesis says, roughly, that the occurrence of *all* events in the universe—including, of course, human decisions and actions—can be deduced from a complete description of previous events and the laws of nature. For the purposes of this paper, I will ignore whether neuroscience can support determinism so defined. I want to ask instead if there are more modest forms of determinism that are both (a) capable of undermining libertarian free will, and (b) supportable, at least in principle, by neuroscience. I claim that there are, and in order to develop my argument I focus first on why incompatibilists take universal determinism to threaten free will.

In general, libertarians reject universal determinism because, for them, free will requires that we do have (at least sometimes) alternative possibilities for what we do *and choose*. Chisholm (1964), for example, claims that one acts freely only if one could have done otherwise. But he rejects a (compatibilist) conditional analysis of "could have done otherwise," that is, an interpretation in which "one could have done otherwise" means that "one would have done otherwise *if* one had chosen otherwise." Instead of such an analysis—which is consistent with the possibility that, given prior events and the laws of nature, she could not choose otherwise—Chisholm holds that "one could have done otherwise" requires "one could have *chosen* otherwise":

> Suppose, after all, that our murderer could not have *chosen*, or could not have *decided*, to do otherwise. Then the fact that he happens also to be a man such that, if he had chosen not to shoot he would not have shot, would make no difference. For if he could *not* have chosen *not* to shoot, then he could not have done anything other than just what it was that he did do. (1964, pp. 175-176)

In a similar way, Kane says that

> when we wonder about whether the *wills* of agents are free, it is not merely whether they could have done otherwise that concerns us […] What concerns us is whether they could have done otherwise *voluntarily* (or *willingly*)… (2009, p. 275)

In order to be able to do otherwise voluntarily, as Kane says, one must be able to choose otherwise. We have again the requirement of alternative choice possibilities. However, Kane does not think it generalizes to every action. For him, libertarian free will requires alternative possibilities only for *some* actions, those which he labels "self-forming actions" (SFAs). In this way, an action results from free will if it is either an SFA or formed on the basis of previous SFAs (Kane, 2009, p. 272). It is because universal determinism entails that (given what happened in the past and the laws of nature) we *never* have alternative possibilities that libertarians regard it as incompatible with free will. For if everything (including actions and decisions) is determined by past events and the laws of nature, then no one can ever choose otherwise.

But now it should become clear that even less demanding forms of determinism can conflict with libertarian free will. As a first possibility, we might have claims that *particular sorts* of events are determined—I will refer to these as "statements of local determination." Consider the following schema for generating statements of this sort:

LD. For any event *x*, if an *x* that is *P* occurs, then another event, *y*, that is *Q*, will occur.[2]

LD says that whenever there is an event of sort P, this fact entails that there will be a second event of sort Q, that is, events of sort P *determine* the occurrence of events of sort Q. We can imagine a similar law that would prevent an individual from choosing otherwise given the occurrence of some previous event whose occurrence was not within the individual's control:

LD1. For any event *x*, and any subject *s*, if an *x* that is a pattern of neural activity of type *B* occurs in *s*'s brain, then *s* will decide to push a given button.

2 This is a modified and simplified version of an analysis of causal laws developed by Davidson (1967, p. 158).

Here, whenever a specific pattern of neural activity happens in a subject's brain, a specific decision results, namely, a decision to push a given button. It should be clear that we could generate a potentially infinite number of statements of local determination like LD1.

Statements like LD1, if true, can have an impact on the sort of libertarian free will that we have been examining. Consider Chisholm's case. If an action is to be free in his libertarian sense, then the agent has to be able to do and choose otherwise. By this criterion, and given LD1, if a pattern of neural activity of type B occurs in a subject's brain, then, in this particular situation, this subject would be unable to choose otherwise.[3] Consequently, an action resulting from such a decision would not be free in Chisholm's sense. Additionally, the more decisions happened to be determined according to that sort of law, the less would be the space for choices and actions that are free in his libertarian sense.

The impact of statements of local determination on Kane's account is more subtle. As we have seen, he only requires SFAs to be such that the agent could have done and chosen them otherwise. On his account, if a pattern of neural activity of type B occurs in a subject's brain and determines a particular behavior according to LD1, this does not entail that the action is not free in a libertarian sense, but merely that it is not an SFA. A free action or choice *can* be deterministically caused on Kanes's account, provided that the causal chain originated in a past SFA (see Kane, 2009, pp. 271-272). Thus, the truth of LD1 would not directly shrink the number of actions resulting from libertarian free will, but only the number of SFAs. But this still allows that the discovery of more and more laws similar to LD1 could decrease our confidence in the existence of SFAs, or at least challenge those who believe in their existence to provide some evidence. For if many choices could be shown to occur deterministically, then it would be natural to ask whether libertarians can support their claim that there is a special class of decisions that are not so determined.

In addition to statements of local determination, we might have a thesis about the deterministic nature of choices in general that would conflict with libertarian free will in a more radical way. Such a thesis is far more demanding than individual statements of local determi-

3 Strictly speaking, the subject would be unable not to choose to push the given button. The *logical* possibility (whatever its empirical plausibility) remains that the subject could make simultaneously other, unrelated decisions. What is usually taken to be relevant in the free will debate, however, is the possibility of *not* choosing in a particular way. For example, it could be that the murderer in Chisholm's example could choose *both* to shoot and to shoot with a black (rather than, say, a gray) gun. But this additional choice would not make the shooting free on his account. The relevant possibility for free will would still be that of *not* choosing to shoot.

nation, but still far less demanding than universal determinism. We can express it in the following way:

DNC. For any subject *s*, any choice *x*, and any course of action *X*, if *s* chooses to do *X*, then there is a previous event *y* of a type *Y* in *s*'s brain, such that whenever an event of type *Y* occurs in someone's brain, then this subject will choose for the course of action *X*.

DNC basically says that every choice occurs according to some statement of local determination. For assume that DNC is true, and that a given subject decided to push a given right button. Then, according to DNC, there would be a previous event of a type (say, of type P) in the subject's brain that is such that any subject in whose brain an event of type P occurred would also decide to push a given right button. But this is to say that there is a statement of local determination about choices of this kind. DNC thus generalizes the idea of statements of local determination by saying that all decisions are determined according to one such statement.

Now, what would be the impact of DNC on Chisholm's and Kane's accounts? Once DNC entails that every choice is determined according to statements of local determination, it follows that if DNC is true, then there are no decisions that are free in the libertarian senses of both Chisholm and Kane. As we have seen, in Chisholm's account a decision and the corresponding action are free only if the decision is not determined. And even though determined actions and decisions can be free on Kane's account—provided that they originated in a previous SFA that was not determined—DNC entails that there are no SFAs. Therefore, DNC would completely undermine the existence of libertarian free will even in Kane's sense.

The result from the discussion so far is that, contrary to the suggestions by Nahmias and Roskies, deterministic statements less demanding than universal determinism can also threaten libertarian free will. And it seems clear that sciences other than fundamental physics, such as neuroscience and psychology, could in principle support those deterministic statements. In the case of statements of local determination, LD1 itself suggests this, since I have deliberately designed it to resemble the results reported by Soon et al. (2008). As for DNC, neuroscience and psychology should also be able to support it, since it is just a generalization about statements of local determination and choices. Despite these possibilities, what has been said is not meant to suggest that it would be easy to discover whether specific brain areas and patterns of neural activity in fact determine specific kinds of choices. Despite great progress

in the study of neural and behavioral aspects involved in making decisions, much remains to be discovered on these matters (see Balleine, 2007; Dayan 2012; Glimcher, 2005, 2013; Gold & Shadlen 2007; Murray, O'Doherty & Schoenbaum 2007; Symmonds & Dolan 2012). The point here is just that we have no reason to think that neuroscience and psychology could not, in principle, find evidence supporting the relevant deterministic statements about decisions. The next section concentrates on the question whether Libet-style experiments have already, as a matter of fact, established some statement of local determination.

4 Libet-style experiments and statements of local determination

The question now is whether results from Libet-style experiments support some deterministic claim that potentially threatens libertarian free will. I will argue that they do not. The argument is based on the claim that results currently available are insufficient to establish even such weaker deterministic statements as LD1. If this is correct, we are even further away from establishing the stronger DNC.

In order to interpret Libet's original results in the light of LD1, we can propose something like this:

LDL. For any event *x*, and any subject *s*, if an *x* that is an RP-II occurs in *s*'s brain, then *s* will decide to flex his/her finger "now" and move his/her finger.

If LDL is true, we can say that readiness potentials of type II determine a peculiar sort of choice, namely, choices to "move now" that are accompanied by actual movement. However, the results fall short of definitely establishing the truth of LDL. Libet measured the time lapse between voluntary movement and RP onset by averaging the EEG signal recorded from 1.4 seconds before finger movements (Libet et al. 1982, p. 324; see also Haynes, 2011, p. 86; Pockett & Purdy, 2011, pp. 35-37; and the commentary following Libet, 1985). Only data within this time interval was actually stored and analyzed. That means that, due to its very design, Libet's original experiment could not find an RP-II that is *not* followed by a decision to "flex now," and by actual movement. But this is critical for assessing the truth of LDL. The only way to falsify it is by finding an RP-II that is not followed by a decision to "flex now." Therefore, Libet's results support in fact the claim that some RPs of type II are followed by decisions to "flex now," rather than the stronger LDL. In other words, Libet's results leave it

open whether RP-II determines decisions to “flex now,” or if it is just something that precedes the sort of action investigated, but that could also precede other sorts of actions and states.[4]

Consider now the experiments by Soon et al. (2008). Here subjects were asked to choose between a left and a right button, press it immediately after deciding for one of them, and then report the time of the decision. During this process their brain activity was scanned with fMRI. Using advanced decoding techniques, the authors were able to show that the spatial pattern of activation is some brain regions (e.g. BA10 in frontopolar cortex) contained predictive information about which button the subject would choose and actually press. This information was available in the brain at about 7-10 seconds before the time subjects reported to have consciously decided, and it predicted the result with nearly 60 % accuracy (see Soon et al., 2008, p. 544, figure 2). In more precise terms, the authors were able to identify some patterns of neural activity whose occurrence indicated that a particular decision would follow with a probability of approximately 60 %—when the chance probability is 50 %.

We could also try to infer something similar to LD1 here. We would get a statement of local determination whose antecedent specifies some pattern of neural activity, and whose consequent specifies a particular choice accompanied by behavior (pressing a right or a left button). As in Libet’s case, the study excludes from the start the possibility of identifying those same patterns of neural activity in situations that are not followed by decisions and movements of the types under investigation. But here the possibility of inferring a deterministic statement is even smaller (indeed null). Since the accuracy is of just 60 %, it follows that in approximately 40 % of cases those patterns of neural activity were followed by a different decision than the one to be expected. For the sake of argument, name “XYR” a pattern of neural activity whose occurrence raises the probability of a decision to push the right button to 60 %. Given that the right-button and left-button options are mutually exclusive, it follows that we should expect XYR neural activity to be followed by decisions to press the *left* button in approximately 40 % of cases. That means that in some occasions XYR neural activity is *not* followed by decisions to push a right button. Therefore, a statement of local determination with the occurrence of XYR as its antecedent and the occurrence of a decision to press the right button as its consequent would be false. Of course, it remains an open question whether

4 Pockett and Purdy (2011, pp. 36–37) say that “Waveforms that look like RPs have been known for decades to occur before a variety of expected events that are not movements.” This suggests that RPs in fact are not uniquely related to decisions to “flex now”. It should also be mentioned that Libet’s experiments on ‘veto’ conditions—when subjects were instructed to prepare to move at a prearranged time and, shortly before, block that preparation—indicated that a great initial portion of an RP of *type I* may not be followed by actual movement (see Libet, 1985, pp. 537-538, especially Figure 2).

future studies could improve accuracy and reveal whether we are facing deterministic processes still poorly known, or processes that are intrinsically stochastic (see Haynes, 2011, p. 93). Either way—and this is the important point here—we are far from having established a deterministic claim that could conflict with libertarian conceptions of free will.

The previous arguments suggest that Libet-style experiments have not so far provided results that could undermine libertarian free will, although neuroscience and psychology more generally could in principle do that. Could Libet-style experiments themselves some day affect libertarian free will? A first step in answering this question is to flesh out what possible result from a Libet-style experiment would lend support to a statement of local determination. The key difficulty, as we have seen, is to establish that some sort of neural activity occurs *exclusively* in situations that are followed by a particular sort of decision—as contrasted, for example, with establishing that a particular sort of decision is always preceded by some sort of neural activity. There are technical difficulties here. In the case of type-II RPs, one needs a reference point on the basis of which EEG recordings from many trials can be averaged. This makes it difficult, practically, to investigate if RPs that are candidates for determinants of specific decisions can appear without the expected decisions and movements (see Libet, 1985, p. 538; Gomes, 1999, p. 64). But practical difficulty does not mean impossibility. One possibility would be to add some form of intervention to Libet-style experiments that induced RPs whose effects could then be analyzed. Additionally, one could have a comparison between intervention and control conditions—a methodology widely used in attempts to infer causal connections. In the case of Soon et al. (2008), a first and crucial limitation is the low accuracy of predictions: we do not have at present a plausible candidate for neural determinant of a particular sort of decision.

What are the prospects for future investigations? Haynes himself (2011, p. 94) has suggested developing a “decision prediction” machine for predicting choices in individual trials in real time. This would allow conducting some relevant experiments. For example, if one could predict the decisions subjects are going to make in real time, “one could ask them to change their mind and take the *opposite* option” (p. 94). It could then be assessed whether some candidate for neural determinant of a particular sort of decision is always followed by the expected decision.

Results currently available, moreover, suggest that this sort of experiment could be implemented in the near future. For example, some studies have achieved higher accuracies in

the prediction of choices from neural events, even on a single-trial basis and even in real time. Maoz, Ross, Ye, Mamelak and Koch (2012) were able to determine in real time from intracranial recordings which hand subjects would raise half a second before a "go" signal. They achieved accuracies above 68 %. Similarly, Salvaris and Haggard (2014) used EEG signals to predict in real time whether subjects would follow or not a given instruction. They achieved approximately 75 % near the "go" signal. Curiously, this is significantly less than the accuracy achieved in conditions in which subjects were asked just to follow a given instruction (approximately 82%; see Salvaris & Haggard, 2014, p. 8, figure 7). The authors themselves interpreted this as suggesting that free actions are less predictable because agents "could have done otherwise" (Salvaris & Haggard, 2014, p. 10). At any rate, despite the increase in accuracy, both Maoz et al. (2012) and Salvaris and Haggard (2014) failed to measure the *time* of subjects' decisions (for difficulties involved in this task, see Banks & Isham, 2011 and Maoz, Mudrik, Rivlin, Ross, Mamelak et al., 2015). This is a shortcoming for the present purposes because we cannot know if the information decoded was predictive of a *forthcoming* choice. In contrast, Fried, Mukamel and Kreiman (2011) have monitored the time of a conscious decision between left and right hand in a similar experiment, but accuracy remained below 70 % before the time of decision (see figure S7E, supplemental information). This again supports the idea that neuroscience *could* provide evidence for deterministic statements that would conflict with libertarian free will, although that evidence has not been provided so far.

5 Conclusion

There has been divergence about the significance of Libet-style experiments for discussions about free will. In what concerns specifically libertarian free will, it turns out that parties have drawn exaggerated conclusions. Contrary to what one side has defended (e.g. Nahmias, 2014b; Roskies, 2006; Sinnott-Armstrong, 2011), experiments in neuroscience and psychology could, in principle, support deterministic statements that undermine libertarian free will. But, contrary to what those in the opposite side have insisted (e.g. Misirlisoy & Haggard 2014; Haynes, 2011), results so far obtained fall short of actually supporting even those weaker statements of local determination. Assumptions involving libertarian free will are often in place in discussions about free will and neuroscience. First, because libertarianism is a more demanding view, both metaphysically and empirically, some have assumed that if science leaves space for free will at all, then it must be for some weaker, compatibilist sort of

free will (see, e.g., Koch, 2012, p. 111; Schlosser 2012). This, together with a second assumption that libertarianism is the correct view (or that it better represents laypersons' views), has also led some to conclude that neuroscience shows that free will (in itself, or in the way it is commonly understood) is an illusion (e.g., Haynes, 2011; Misirlisoy & Haggard, 2014, p. 37; Harris, 2012, p. 16).[5] If the present results are correct, data from Libet-style experiments lend support to none of those assumptions, although they (as well as other studies in neuroscience and psychology) could in principle do that.[6]

References

Andow, J., & Cova, F. (in press). Why compatibilist intuitions are not mistaken: A reply to Feltz and Millan. *Philosophical Psychology*.

Balleine, B. (2007). The neural basis of choice and decision making. *Journal of Neuroscience*, 27, 8159–8160.

Banks, W., & Isham, E. (2011). Do we really know what we are doing? Implications of reported time of decision for theories of volition. In W. Sinnott-Armstrong & L. Nadel (Eds.), *Conscious will and responsibility: A tribute to Benjamin Libet* (pp. 47–60). Oxford: Oxford University Press.

Bourget, D., & Chalmers, D. (2013). What do philosophers believe? *Philosophical Studies*, 170, 465–500.

Chisholm, R. (1964). Human freedom and the self. In D. Pereboom (Ed.), *Free will* (pp. 172–184). Indianopolis, IN: Hackett.

Davidson, D. (2001). Causal relations. In *Essays on actions and events* (2nd ed.; pp. 149–162). Oxford: Oxford University Press.

Dayan, P. (2012). Models of value and choice. In R. Dolan & T. Sharot (Eds.), *Neuroscience of preference and choice: Cognitive and neural mechanisms* (pp. 33–52). Amsterdam: Academic Press.

Deery, O., Davis, D., & Carey, J. (2015). The free-will intuitions scale and the question of natural compatibilism. *Philosophical Psychology*, 28, 776–801.

Feltz, A., & Cova, F. (2014). Moral responsibility and free will: A meta-analysis. *Consciousness and Cognition*, 30, 234–246.

5 As I have noted earlier, controversies remain in the philosophical debate on compatibilism versus incompatibilism, as well as in the experimental research on laypersons' beliefs about free will. On the latter, see, for example, Nahmias, Morris, Nadelhoffer & Turner (2006), Nahmias, Coates & Kvaran (2007), Nichols & Knobe (2007), Rose & Nichols (2013), Deery, Davis & Carey (2014), Feltz & Cova (2014), Nahmias (2014a), and Andow & Cova (forthcoming).

6 For comments, I thank Gilberto Gomes, Alfred Mele, Frank Sautter, Rogério Severo, and four anonymous referees. Thanks also to Eddy Nahmias and Adina Roskies for their comments at the 2015 Minds Online Conference, and to Ricardo di Napoli, Gilson Olegario, Danilo Dantas, Lucas Roisenberg, Márlon Teixeira, Cristina Nunes, Lucas Dalsotto, and Rafael Cortes for discussion. This work has been financially supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [grant number BEX 4789/15-6], and by Fulbright Brasil.

Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron*, 69, 548–562.

Glimcher, P. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology*, 56, 25–56.

Glimcher, P. (2013). Value-based decision making. In P. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed.). (pp. 373–392). Amsterdam: Academic Press.

Gold, J., & Shadlen, M. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535–574.

Gomes, G. (1998). The timing of conscious experience: A critical review and reinterpretation of Libet's research. *Consciousness and Cognition*, 7, 559–595.

Gomes, G. (1999). Volition and the readiness potential. *Journal of Consciousness Studies*, 6, 59–76.

Harris, S. (2012). *Free will*. New York, NY: Free Press.

Haynes, J. (2011). Beyond Libet: Long-term prediction of free choices from neuroimaging signals. In W. Sinnott-Armstrong & L. Nadel (Eds.), *Conscious will and responsibility: A tribute to Benjamin Libet* (pp. 85–96). Oxford: Oxford University Press.

Kane, R. (2009). Rethinking free will: New foundations for an ancient problem. In D. Pereboom (Ed.), *Free will* (pp. 268–288). Indianopolis, IN: Hackett.

Koch, C. (2012). *Consciousness: Confessions of a romantic reductionist*. Cambridge, MA: MIT Press.

Kornhuber, H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger's Archiv*, 284(1), 1–17.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8, 529–566.

Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies*, 6, 47–57.

Libet, B., Gleason, C., Wright, E., & Pearl, D. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, 106, 623–642.

Libet, B., Wright, E., Feinstein, B., & Pearl, D. (1982). Readiness potentials preceding unrestricted 'spontaneous' vs. pre-planned voluntary acts. *Electroencephalography and Clinical Neurophysiology*, 54, 322–335.

Maoz, U., Mudrik, L., Rivlin, R., Ross, I., Mamelak, A., & Yaffe, G. (2015). On reporting the onset of the intention to move. In A. Mele (Ed.), *Surrounding free will: Philosophy, psychology, neuroscience* (pp. 184–202). Oxford: Oxford University Press.

Maoz, U., Ross, I., Ye, S., Mamelak, A., & Koch, C. (2012). Predicting action content on-line and in real time before action onset—An intracranial human study. *Advances in Neural Information Processing Systems*, 25, 881–889.

Mele, A. (2006). Free will: Theories, analysis, and data. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior?* (pp. 187–206). Cambridge, MA: MIT Press.

Mele, A. (2009). *Effective intentions: The power of conscious will*. New York, NY: Oxford University Press.

Mele, A. (Ed.). (2015). *Surrounding free will: Philosophy, psychology, neuroscience*. Oxford: Oxford University Press.

Misirlisoy, E., & Haggard, P. (2014). A neuroscientific account of the human will. In W. Sinnott-Armstrong (Ed.), *Moral psychology (Vol. 4): Free will and moral responsibility* (pp. 37–42). Cambridge, MA: MIT Press.

Murray, E., O'Doherty, J., & Schoenbaum, G. (2007). What we know and do not know about the functions of the orbitofrontal cortex after 20 years of cross-species studies. *Journal of Neuroscience*, 27, 8166–8169.

Nahmias, E. (2014a). Explaining away incompatibilist intuitions. *Philosophy and Phenomenological Research*, 88, 434–467.

Nahmias, E. (2014b). Is free will an illusion? Confronting challenges from the modern mind sciences. In W. Sinnott-Armstrong (Ed.), *Moral psychology (Vol. 4): Free will and moral responsibility*, (pp. 1–25). Cambridge, MA: MIT Press.

Nahmias, E., Coates, J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, 31, 214–232.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73, 28–53.

Nichols, S. & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, 41, 663–685.

Pockett, S., Banks, W., & Gallagher, S. (Eds.). (2006). *Does consciousness cause behavior?*. Cambridge, MA: MIT Press.

Pockett, S., & Purdy, S. (2011). Are voluntary movements initiated preconsciously? The relationships between readiness potentials, urges, and decisions. In W. Sinnott-Armstrong & L. Nadel (Eds.), *Conscious will and responsibility: A tribute to Benjamin Libet* (pp. 34–46). Oxford: Oxford University Press.

Rose, D., & Nichols, S. (2013). The lesson of bypassing. *Review of Philosophy and Psychology*, 4, 599–619.

Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Science*, 10, 419–423.

Salvaris, M., & Haggard, P. (2014). Decoding intention at sensorimotor timescales. *PLoS ONE*, 9, e85100.

Schlosser, M. (2012). Free will and the unconscious precursors of choice. *Philosophical Psychology*, 25, 365–384.

Sinnott-Armstrong, W. (2011). Lessons from Libet. In W. Sinnott-Armstrong & L. Nadel (Eds.), *Conscious will and responsibility: A tribute to Benjamin Libet* (pp. 235–246). Oxford: Oxford University Press.

Sinnott-Armstrong, W., & Nadel, L. (Eds.). (2011). *Conscious will and responsibility*. Oxford: Oxford University Press.

Soon, C., Brass, M., Heize, H., & Heynes, J. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 11, 543–545.

Symmonds, M., & Dolan, R. (2012). The neurobiology of preferences. In R. Dolan & T. Sharot (Eds.), *Neuroscience of preference and choice: Cognitive and neural mechanisms* (pp. 3–31). Amsterdam: Academic Press.

ARTIGO 2: NEUROSCIENCE AND THE POSSIBILITY OF LOCALLY DETERMINED CHOICES: REPLY TO ADINA ROSKIES AND EDDY NAHMIAS*

Abstract: In a previous paper, I argued that neuroscience and psychology could in principle undermine libertarian free will by providing support for a subset of what I called “statements of local determination.” I also argued that Libet-style experiments have not so far supported statements of that sort. In a commentary to the paper, Adina Roskies and Eddy Nahmias accept the claim about Libet-style experiments, but reject the claim about the possibilities of neuroscience. Here, I explain why I still disagree with their conclusion, despite being sympathetic to a lot of what they say in support of it.
Keywords: free will; neuroscience; determinism; incompatibilism; choice

In a previous paper (Fischborn, 2016), I argued that neuroscience and psychology could in principle undermine libertarian free will by providing support for a subset of what I called ‘statements of local determination’. I also argued that Libet-style experiments have not so far supported statements of that sort. In a commentary to the paper, Adina Roskies and Eddy Nahmias (2016) accept the claim about Libet-style experiments, but reject the claim about the possibilities of neuroscience. They argue that “neuroscience cannot establish the relevant kind of determinism” and that “in principle, neuroscience will not be able undermine libertarian free will” (2016, p. 2). In this short reply, I explain why I remain unpersuaded by their argument.

As I defined the notion, statements of local determination are statements of the following sort:

LD1. For any event *x*, and any subject *s*, if an *x* that is a pattern of neural activity of type *B* occurs in *s*’s brain, then *s* will decide to push a given button. (2016, p. 497)

I argued that LD1 entails that a choice resulting from the occurrence of a pattern of neural activity of type B is not free—if one is assuming an incompatibilist conception of free will such

* The Version of Record of this manuscript has been published and is available in *Philosophical Psychology*, 30.1-2 (2017): 198–201. <www.tandfonline.com/doi/full/10.1080/09515089.2016.1266319>

as Roderick Chisholm's (1964)—and that it is not an (undetermined) self-forming action—if one is assuming an incompatibilist conception of free will such as Robert Kane's (1996). And I described how the truth of a sufficiently large number of statements of local determination similar to LD1 would reduce the plausibility of the claim that choices that are free in an incompatibilist sense exist. I argued that "sciences other than fundamental physics, such as neuroscience and psychology, could in principle support those deterministic statements." (2016, p. 498).

Roskies and Nahmias reject what this last claim implies about neuroscience, and their argument relies on the possibilities that (1) mental states are realized by the brain in multiple ways (multiple-realizability), that (2) mental states are constituted by states beyond the brain (extended or embodied cognition), and that (3) neural networks are complex and chaotic (complexity/chaos). I would like to distinguish certain sorts of claims about what a specific science can do in order to show why I remain unpersuaded by their conclusion despite being sympathetic to a lot of what they say in support of it.

Regarding a hypothesis like LD1, we can distinguish the following claims:

(a) Neuroscience has supported LD1.
(b) Neuroscience can support LD1 if X is true (for some X).
(c) Neuroscience can support LD1 if LD1 is true.

Roskies, Nahmias, and I agree that (a) is false. We also agree that (b) is false if we substitute 'X' for the thesis that minds extend beyond the brain, or the thesis that mental states can be realized in multiple ways. (I also agree that (b) may be false if we substitute 'X' for the claim that neural networks are chaotic in an *in*deterministic way.) But I think (c) is true—and I am not sure Roskies and Nahmias need to disagree with this. As I understand it, (c) is similar to the claim that physics could show, for example, that time travel is possible. It can be true that physics could in principle show that time travel is possible even if time travel is actually physically impossible. Saying that physics can in principle show that time travel is possible (if time travel is actually possible) says nothing about the truth or plausibility of the possibility of time travel; it only says that physics is the right science for an investigation on the possibility of time travel. Moreover, (c) is not trivially true because, for example, the claim that "Sociology can support LD1 if LD1 is true" seems clearly false. Accordingly, the claim that neuro-

science can in principle support LD1 is not meant to suggest anything about the truth or likelihood of LD1. By asserting (c), I am just assigning to neuroscience the task of assessing the truth or falsity of LD1.

I take the previous distinctions to be perfectly consistent with my original claim that neuroscience could in principle support a group of relevant statements of local determination. As I originally said, the very fact that LD1 purposely resembles the way Soon, Brass, Heize, and Heynes (2008) report their results suggests that statements like LD1 are a topic for neuroscience. And I added that the claim that neuroscience could in principle support statements like LD1 "is not meant to suggest that it would be easy to discover whether specific brain areas and patterns of neural activity in fact determine specific kinds of choices" (2016, p. 498). Roskies and Nahmias say that "local determination is unlikely to be established by neuroscience in any form that should trouble compatibilists or libertarians" (2016, p. 1). I agree with that and, as I explained above, I take this agreement to be consistent with the claim that neuroscience could in principle support local determination. The mere fact that local determination of the relevant sort is unlikely is not a threat to my point.

A second objection Roskies and Nahmias raise is that "local determination would not raise a challenge to free will without assuming universal determinism" (p. 5). In my original paper, I described how statements of local determination could impact libertarian theories of free will such as Chisholm's and Kane's. Roskies and Nahmias note that such theorists endorse incompatibilism "only because they first develop or accept incompatibilist arguments using universal determinism" (p. 4). Therefore, they ask, "Why should we accept the incompatibilism that motivates these libertarians unless we are considering the sort of universal determinism that neuroscience cannot establish[?]" (p. 4). As Roskies and Nahmias properly note, I did not provide any argument for incompatibilism—and neither did I intend to do so. My aim was to examine whether the conditions for the *existence* of free will posited by some theories are ever satisfied, and in doing so I disregarded (at least for the moment) the question of why those theories arrived at conclusions about the *incompatibility* of free will and determinism. Thus, my claim that neuroscience could in principle undermine free will should be read conditionally: neuroscience could in principle undermine free will *if* free will requires that at least some choices are not determined by previous neural activity. For the theories I considered, part of the reason why such undetermined choices are required is that only some

forms of indetermination leave unconditional room for alternative choices and actions—a point that has been influential in different ways in the history of incompatibilism.

Roskies and Nahmias also point out that in theories of free will such as Kane's free will can exist even if most of our choices are determined, provided that a few of them are not. On this basis, they argue that evidence for the claim that some choices are determined by previous neural activity would not give us reasons to doubt that free will (in the sense described by such theory) exists. In order to show that no choice is free in the sense considered, Roskies and Nahmais say, one would need "reasons to think that indeterministic events do not affect brain activity, and thus to think that global determinism is true" (p. 7). In the paper, I described DNC, a version of a more global form of determinism (still weaker than universal determinism) that I thought would undermine the existence of free will in Kane's sense:

DNC. For any subject *s*, any choice *x*, and any course of action *X*, if s chooses to do *X*, then there is a previous event *y* of a type *Y* in *s*'s brain, such that whenever an event of type *Y* occurs in someone's brain, then this subject will choose for the course of action *X*. (2016, p. 498)

If we had reasons to think that DNC is true, I think we would have reasons to think that no choice is free in the sense postulated by Kane. Nahmias and Roskies certainly doubt that neuroscience will ever show that DNC is true; me too. But, no matter how unlikely DNC sounds given what we currently know about brains and decisions, I still believe that, if DNC is true, then neuroscience could in principle show that it is, and, therefore, that neuroscience can in principle undermine the existence of free will in certain incompatibilist senses.

References

Chisholm, R. (1964). "Human freedom and the self". In: Pereboom, D. (Ed.), *Free will*, pp. 172-184. Indianapolis: Hackett.

Fischborn, M. (2016). "Libet-style experiments, neuroscience, and libertarian free will", *Philosophical Psychology* 29: 494-502.

Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.

Roskies, A. and Nahmias, E. (2016). ""Local determination", even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn", *Philosophical Psychology* : 1-9.

Soon, C. S.; Brass, M.; Heinze, H.-J. and Haynes, J.-D. (2008). “Unconscious determinants of free decisions in the human brain”, *Nature Neuroscience* 11: 543-545.

ARTIGO 3: HOW SHOULD FREE WILL SKEPTICS PURSUE LEGAL CHANGE?*

Abstract: Free will skepticism is the view that people never truly deserve to be praised, blamed, or punished for what they do. One challenge free will skeptics face is to explain how criminality could be dealt with given their skepticism. This paper critically examines the prospects of implementing legal changes concerning crime and punishment derived from the free will skeptical views developed by Derk Pereboom and Gregg Caruso. One central aspect of the changes their views require is a concern for reducing the severity of current forms of punishment. The paper considers two strategies for pursuing such a reduction. By taking into account evidence from the psychology of belief in free will and desire to punish, it is argued that a strategy aiming at a reduction of people's natural desire to punish criminals can be successful if capable of providing alternatives to current forms of punishment satisfying three properties: they must be *less harmful* than current forms of punishment, *more effective* in preventing crime, and *incompatible* with current forms of punishment.

Keywords: free will skepticism; punishment; criminal behavior; public opinion; criminal law

1 Introduction

Theorists who doubt that people have free will often show a concern for the practical implications of their views. The concern derives from the fact that free will is usually thought to be a necessary condition for things such as moral and legal responsibility, desert, and punishment. Derk Pereboom (2001, 2014) illustrates this pattern well. While he defends the view that people are never free in a sense that would make them truly deserve to be praised, blamed, or punished for their actions—a view he calls 'free will skepticism'—he also puts forth a fairly elaborate account of how crime could be dealt with given such skepticism.[1] His account includes two main parts. One is the rejection of retributivistic justifications of punishment that appeal to the claim that criminals should be punished because they deserve it. The other is a proposal of independently acceptable ways to prevent crime that avoid inflicting harm as

* This paper has been published in *Neuroethics* 11.1 (2018): 47–54. It is available via Springer at <http://dx.doi.org/10.1007/s12152-017-9333-8>.

1 Although this paper focuses on Pereboom's project, I will also consider the way Caruso (2016) further develops it. Other free will skeptic proposals include the ones by Corrado (2013), Greene and Cohen (2004) and Vilhauer (2013).

much as possible. In synthesis, Pereboom's project begins with free will skepticism and goes all the way down to the claim that current punishment practices (and criminal law, more generally) should be modified.

The question I want to answer in this paper is the following: Were theorists to become convinced that no one has the free will that would be required for making punishment ever deserved, how should the relevant legal changes be pursued? In asking this question, I intend to contribute with the investigation of the practical viability of a free will skeptic project, but this is not meant to imply that I am endorsing skepticism. Indeed, free will skepticism is far from consensual among theorists, and I will not intervene in this debate here.[2] But as I will indicate in the end, this investigation has some lessons that can be relevant beyond the scope of free will skepticism.

Below, I begin by describing Pereboom's free will skeptic project in more detail, and the way Gregg Caruso (2016) has extended Pereboom's project (section 2). I then present and criticize what I call a cognitive strategy for pursuing relevant legal changes (section 3). Finally, I present a more promising, non-cognitive strategy for pursuing legal changes (section 4), and conclude by indicating how non-skeptics may also benefit from the previous reflection (section 5).

2 Free will skepticism and punishment

Free will skepticism is the view that no one has the kind of free will that is required for an agent to be morally responsible in what Pereboom calls the 'basic desert sense'. In this sense, responsibility practices are essentially backward-looking. If an agent is responsible in the basic desert sense, then she "would deserve to be blamed or praised just because she has performed the action, given an understanding of its moral status, and not, for example, merely by virtue of consequentialist or contractualist considerations" (Pereboom, 2014, p. 2; see also his 2001). Pereboom's argument for the claim that we do not have the kind of free will that is required for responsibility in the basic desert sense is based on the following premises. First, he rejects event-causal libertarian theories, and argues that free will requires agent-causation (2014, ch. 2). Second, he argues that the claim that we are agents with such causal powers is implausible given current physical theories (ch. 3). Finally, Pereboom also rejects

2 Among contemporary authors, opposition to free will skepticism comes mainly from proponents of theories that affirm the existence of free will, including libertarians (Kane, 1996), compatibilists (Fischer, 1994; Morse, 2013), and views that combine elements of both compatibilism and libertarianism in different ways (Mele, 1995; Mele, 2006; Vargas, 2013).

compatibilism as a viable alternative on the basis of manipulation arguments (ch. 4). It follows that it is implausible that we have the kind of free will that is required for responsibility in the basic desert sense.

For the purposes of this paper, the most relevant feature of Pereboom's free will skepticism is that it entails that people never deserve, in the basic desert sense, to be punished. If that is true, then no justification of punishment that assumes people deserve it (such as retributivism) can be successful. This is significant because retributivism is one of the leading alternatives when it comes to the justification of punishment, and also because current criminal justice systems seem to have a significant retributivistic character. Bedau and Kelly (2015), for example, argue that the very notion of punishment is 'inherently retributive'. As a consequence, describing acceptable alternative ways of dealing with criminal behavior is a pressing task for a free will skeptic.

An often considered alternative to retributivism is consequentialism. Purely consequentialist justifications of punishment avoid claims about what an agent deserves, and instead focus on its possible positive consequences, such as deterrence or incapacitation. This alternative has been embraced by some skeptics about the existence of free will (e.g. Greene and Cohen, 2014). On their view, increasing disbelief in free will would eventually motivate a transition from retributivistically-oriented forms of punishment to more consequentialist ones. But Pereboom rejects this move. Although he grants that consequentialism is consistent with free will skepticism, he argues that it faces serious ethical objections, including an inability to offer a principled rejection of patently problematic possibilities—such as punishing an innocent or punishing someone much more harshly than it would seem fair—if they happen to be conductive to the best consequences overall (see Pereboom, 2014, pp. 163-165). That is why Pereboom's proposal involves neither retributivism nor pure consequentialism.

As a viable alternative, Pereboom seeks to justify measures whose focus is on crime prevention and not on the response to crime that has already occurred. Here is a summary of his view:

> If the free will skeptic is right, criminal punishment for retributive reasons is ruled out. […] But a theory of crime prevention that would be acceptable whether or not the skeptic is right can be developed by analogy with our rationale for quarantining carriers of dangerous diseases. The core idea is that the right to harm in self-defense and defense of others justifies incapacitating the criminally dangerous with the minimum harm required for adequate protection. […] The free will skeptic would also endorse measures for reducing crime that aim at altering social conditions, such as improving education, increasing oppor-

> tunities for fulfilling employment, and enhancing care for mentally ill. (Pereboom, 2014, pp. 173-174; see also 2001, ch. 6)

As we can infer from this passage, Pereboom's proposal has two main components. One component consists in the rejection of punishment for retributive reasons.[3] This entails that, if current criminal justice systems do rely on retributive considerations to some extent, they need to be revised. A second component focuses on crime prevention and has two aspects. One aspect is the justification of measures to prevent imminent crime through the incapacitation of likely criminals with the minimum harm needed (e.g. detention) based on the right to harm in self-defense. According to Pereboom, there are various requirements for a proper implementation of this aspect of the preventive component. For example, the right to liberty, the concern that people will be used merely as means, and the possibility of misuse by the state should be seriously considered (2014, p. 170). Also, preventive detention would require the availability of very accurate, reliable, and non-invasive methods for detecting the likelihood of criminal behavior, as well as a concern for the wellbeing of, and an effort to rehabilitate detained individuals (pp. 170-171). The other aspect of the preventive component proposes measures that aim at preventing people from becoming likely criminals in the first place, and include the improvement of social and health conditions.

Pereboom's proposal has been defended and extended by Gregg Caruso (2016), and by Pereboom and Caruso (forthcoming) in collaboration. Caruso (2016) justifies the free will skeptic approach to criminality within the framework of public health ethics. He calls the resulting model a 'public health-quarantine model'. One of the innovations in this model is the claim that free will skeptics not just *can* adopt measures for crime prevention—as Pereboom originally suggested—but should *prioritize* doing so (2016, p. 31). As he says:

> [Pereboom's] quarantine analogy is narrowly focused on justifying the incapacitation of dangerous criminals. […] The public health-quarantine model justifies the incapacitation of dangerous criminals but the primary focus should always be on preventing crime from occurring in the first place by addressing the systemic causes of crime. Prevention is always preferable to incapacitation. (2016, pp. 35-36)

Caruso's model, therefore, includes at least two new elements: it goes beyond Pereboom's quarantine analogy when it comes to the justification of responses to crime; and it emphasizes prevention over incapacitation as the preferred form of response.

---

3 Sometimes, Pereboom seems to reject punishment completely, and not just punishment on retributive grounds (see 2014, p. 165).

For the purposes of this paper, it matters to identify a group of legal changes that is common to both Pereboom's and Caruso's proposals. The group of changes I have in mind includes only those that properly depend on, or are required by, their free will skepticism. Free will skepticism, for example, entails that an individual who has committed a crime does not deserve any punishment in response. It also entails that, at the very least, we should be concerned about the fact that most societies currently assign punishment in response to crime. But it does not follow from the thesis that we lack the kind of free will that would make punishment ever deserved that we should adopt any of the preventive measures Pereboom and Caruso describe. Pereboom himself describes some of those measures as "acceptable whether *or not* the skeptic is right" (emphasis added), which suggests that even non-skeptics might be willing to support them. And even though Caruso says his model requires active steps to prevent crime, such requirement derives not from his skepticism but from the public health ethics he also adopts. What is left as an essentially free will skeptic demand for legal change is a concern for reducing as much as possible the amount of harm to be inflicted upon those who have committed crimes. It might be suggested that free will skepticism requires the actual reduction of such harm, but Pereboom's claim (which Caruso also accepts) that "the minimum harm required for adequate protection" should be used seems consistent with the possibility that current forms of punishment did already reach that minimum. Consider, for example, incarceration as a punishment. A free will skeptic should propose at least a reduction in incarceration duration (reduction understood as consistent with elimination) *if* it can be shown that shorter incarceration sentences are enough for adequate protection.[4] Therefore, the best candidate for the sort of legal change free will skepticism requires is a concern for reducing as much as possible the amount of harm to be inflicted upon those who have committed a crime, leaving open the possibility of completely eliminating such harm.

Assume, now, that a convincing case can be made for the claim that a reduction in the harm involved in current forms of punishment is consistent with keeping or increasing the level of protection against damage or violence that individuals within a certain society enjoy. That would mean that a reduction in the *severity* of current punishment practices is consistent with keeping or increasing the level of protection within that society. The main question for this paper can now be reframed as follows: Given such an assumption, how should a free will skeptic pursue the relevant legal changes, i.e., changes aiming at the reduction of punishment

4 Of course, on the same condition, the free will skeptic should also propose the reduction of other harms often associated with imprisonment (e.g., prison violence, and prisoner mistreatment).

severity within a certain society? In the remainder of this paper, I consider two answers to this question.

3 Pursuing legal change: A cognitive strategy

How should a free will skeptic pursue legal changes aiming at the reduction of punishment severity within a given society? In this section, I begin by motivating an important constraint to be taken into account when assessing the prospects of different strategies for pursuing legal change. Then I describe a first obvious strategy available to the free will skeptic, and explain why I think it is unpromising. A more promising alternative will be presented in the next section.

A *desiderata* for any strategy for legal change is that it can make the changes proposed acceptable for a substantial part of a given society's members. The reason for this constraint on strategies for legal change is that I am interested in assessing the prospects of such strategies in *democratic* societies. In democracies, relevant changes in the law usually need to be sanctioned by a majority of legislators. Also, the wider population itself has to elect those representatives in the first place. Therefore, it is hard to see how new laws could be sanctioned without the wider population itself endorsing to some extent the views of those candidates who defend the legal changes under consideration. In addition, even if changes could come to be implemented without the agreement of most people, a huge discrepancy between what the population believes and desires and what the legal system provides might pose a threat to the system's legitimacy (see, e.g., de Keijser & Elffers, 2009, for an analysis of this problem in the context of The Netherlands). It follows that, other things being equal, free will skeptics interested in implementing the legal changes recommended by their views should favor a strategy that has a better chance of gaining popular adherence.

A first possible strategy for pursuing legal change can be easily derived from Pereboom's and Caruso's free will skeptic projects. Just as free will skeptics themselves propose legal changes because they believe no one has free will, legal change could be pursued by trying to convince the larger population that no one has free will. I will call this a *cognitive strategy* because it has belief in free will as a target in its pursuit of legal change. This does not seem a promising alternative, though.

First, people's current beliefs about free will, desert, and punishment are in clear opposition to free will skepticism. Studies on people's beliefs about free will indicate that they are

strong (Nadelhoffer et al., 2014, p. 38) and hard to manipulate (Schooler et al., 2015). Interestingly, many of the attempts to decrease belief in free will involve real or fictitious quotes from scientists saying that free will does not exist or is an illusion (see Vohs & Schooler, 2008, p. 50; Schooler et al., 2015, pp. 75-77). And even when experimental manipulation is statistically significant, belief in free will remains considerably strong. For example, Monroe, Brady and Malle (2016, study 1) successfully decreased belief in free will from 5.03 (SD = 1.17; 1 = 'strongly disagree', 7 = 'strongly agree') to 4.77 (SD = 1.25). If we take into account the fact that statements used to assess belief in free will included such strong claims as 'People always have free will' (see Nadelhoffer et al., 2014, p. 34 for the scale used in the study), it is fair to conclude that even participants who had their beliefs successfully decreased were far from becoming free will skeptics. Studies also confirm that people take punishment to be appropriate in a variety of contexts. For another item in Nadelhoffer et al.'s scale (2014, p. 38)—saying that "People who harm others deserve to be punished even if punishing them will not produce any positive benefits to either the offender or society—e.g., rehabilitation, deterring other would-be offenders, etc."—responses averaged 5.37 (SD = 1.46; 1 = 'strongly disagree', 7 = 'strongly agree'; n = 330; sample from the United States population). These results indicate that most people in this population are likely to disagree with the skeptic's denial of free will and its consequences for punishment. Moreover, there is some evidence that patterns in beliefs about free will and attitudes toward punishment are similar across different cultures. Sarkissian et al. (2010) found similarities in belies about free will across Western and Eastern countries. And Santin et al. (manuscript) found evidence of the transcultural validity of Nadelhoffer et al.'s scale in a Brazilian sample. Therefore, it is at least a worth considering hypothesis that belief in free will and support for punishment are similarly strong and robust across different cultures.[5]

Second, belief in free will and desire to punish are not independent. Rather, belief in the existence of free will and in the appropriateness of punishing criminals seem to be part of a natural and widespread strategy people adopt with the aim of repelling undesirable behavior. In a series of studies by Clark et al. (2014), people showed stronger belief in free will after exposure to crime and immoral behavior. This effect was found to be mediated by a stronger desire to punish the authors of those actions. These results suggest that confidence in the effec-

5 It is also a worth considering possibility that the relation between belief in free will and desire to punish indicated in the next paragraph is similar across cultures. It goes without saying that, were these aspects of belief in free will and attitudes toward punishment to be peculiar to some cultures, the overall conclusions of this paper should be relativized accordingly.

tivity of alternative ways of reducing crime might be a prerequisite for reducing belief in free will and then getting enough support for changes in more traditional forms of punishment. Therefore, it might be that a free will skeptic approach to crime can only be implemented if it can ensure the efficacy of alternative ways of reducing crime.

Third, there is some chance that less severe punishment may contribute to *increase* criminality, which might end up reinforcing belief in free will. It is accepted, for example, that punishment has preventive effects on crime. After critically reviewing studies on the preventive effects of punishment, Suhling and Greve (2009, p. 420) agree that "despite many methodological problems and often inconsistent results, a crime-preventive effect of the existence of the criminal justice system cannot be denied". It has also been pointed out that deterrence is somewhat correlated with the severity of punishment—although the connection is weaker than that between deterrence and punishment *certainty* (see von Hirsch et al., 1999; Friesen, 2012; but see also Doob &Webster, 2003). These effects of punishment are also present in laypersons' views. In another item in Nadelhoffer et al.'s study—saying that "People who perform harmful actions ought to be punished so that other potential offenders are deterred from committing similar harmful actions"—the responses averaged 5.78 (SD = 1.2), indicating again a substantial agreement with the statement. Thus, insofar as free will skepticism requires an attempt to reduce punishment severity, the risk of increasing crime might actually reinforce belief in the existence of free will and opposition to a free will skeptic project.

On the three aspects considered, therefore, a cognitive strategy for legal change having belief in free will as a main target looks unpromising. Belief in free will is currently strong, to begin with, and the way belief in free will relates to other beliefs and desires makes unlikely that it will get weaker without (at the very least) a significant reduction in crime. Free will skeptics, therefore, need a better strategy for pursuing relevant legal changes.

4 A non-cognitive strategy

One of the challenges for the cognitive strategy just described is that crime can trigger a desire to punish which has been shown to reinforce belief in free will. This sets the stage for considering an alternative strategy focusing on the desire to punish. A widespread reduction in people's desire to punish might, simultaneously, favor two central goals that free will skeptics may have. First, a reduced desire to punish would make support for a reduction in punishment severity easier to achieve. Second, a reduced desire to punish might contribute to a reduction

in the strength of people's belief in free will. Because the main target of this strategy for legal change is a *desire*, I call it a non-cognitive strategy.[6]

Is there any plausible way of reducing the desire to punish available for the free will skeptic? Studies on people's preferences concerning different sorts of policies on crime provide an interesting starting point. McCorkle (1993) found that people tend to favor both punitive and rehabilitative policies when their combination is possible. In another study, Baker et al. (2015) found that rehabilitation policies are preferred to punitive ones when people are forced to choose for only one of them. These results suggest some further constraints on how punishment severity can be successfully reduced. First, in accordance with the free will skeptic project, they must be *less harmful* than current practices. Second, they must be *more effective* in reducing crime; otherwise crime itself could reinforce a desire for punishment. And third, they must be *incompatible* with current forms of punishment because otherwise people might continue to support current forms of punishment *alongside* other measures. Should we become convinced of the existence of alternatives satisfying these three properties, I think it is a reasonable prediction that our natural desire to punish criminals would become weaker. Of course this is a prediction that needs further empirical investigation, but it seems our best guess given the evidence available so far.

Earlier in the paper I have distinguished between those legal changes that are entailed by free will skepticism (reduction in punishment severity) and those that are not (preventive measures). The relevance of this distinction can now be better appreciated and qualified. Only the legal changes derived from free will skepticism are incompatible with current punitive responses to crime, on the assumption that less harmful practices can ensure an adequate level of social protection. Hence only measures involving a reduction in punishment severity would be considered as a replacement for current practices. Pereboom's and Caruso's preventive measures, on their turn, are strictly consistent with non-skeptical views, and thus could be more easily supported by the public, although not instead of current practices.

As previously mentioned, the empirical studies on belief in free will and desire to punish suggest that a reduction in crime may be a necessary condition for reducing the desire to punish criminals. This suggests some qualification regarding the points made in the previous paragraph. For if measures for crime prevention are successfully implemented, one can also

6 It should not be assumed that the non-cognitive strategy needs to exclude the aim of convincing people that they lack free will. In this sense, the cognitive and non-cognitive strategies may be taken as two components that may or may not be combined in a real attempt to implement legal recommendations derived from free will skepticism.

expect an overall reduction in desire to punish and belief in free will. Success in prevention, therefore—even if not conceptually required by free will skepticism itself—may be contingently central for achieving some of the free will skeptic's goals. In other words, Pereboom and Caruso are right to emphasize preventive measures, although maybe for reasons not fully understood so far. On the other hand, those legal changes that free will skepticism strictly requires concern what the responses to actual crime should be. For this reason, even if crime prevention can reduce desire to punish on a general level, it does not follow that the specific desire to punish that arises after a particular crime is considered will also decrease. The question of whether free will skeptics should pursue the non-cognitive strategy, therefore, can only be answered by investigating the existence of alternatives involving a reduction in punishment severity that can be simultaneously successful in the prevention of crime.

Are there any good reasons to believe that such alternatives can be found? Physical punishment of children, although not a part of the criminal justice system itself, provides an interesting case for reflection. Research indicates that the practice of physically punishing children is widespread across the globe (see, e.g., Kish & Newcombe, 2015; Global Initiative, 2016). It has been suggested that parents' beliefs about its necessity for proper education and unharmfulness are among the possible causes of parental use of physical punishment (Kish & Newcombe, 2015). And yet, several studies have shown that physical punishment of children is associated with undesirable outcomes such as aggressiveness, antisocial behavior and psychological problems, among many others (Gershoff, 2002; Afifi et al., 2006; Durrant & Ensom, 2012). For these reasons, there have been initiatives to end up with the practice of physically punishing children, which have led many countries to prohibit its use. By 1990, four countries prohibited all corporal punishment of children. The number increased to eleven by 2000, to thirty-four by 2011, and is currently forty-seven (see Global Initiative, 2015; Durrant & Ensom, 2012, p. 1373).

Without overlooking important differences, the case of the physical punishment of children provides a model for thinking about alternatives to current forms of legal punishment. First, there is a convincing case for reducing the severity or harm involved in punishing children—actually, there is a convincing case for *abolishing* the practice of harming children in the process of educating them. Second, the evidence suggests that a non-punitive education is more effective in preventing certain undesirable attitudes, including aggressiveness. And third, the alternatives proposed are incompatible with preserving old punitive practices be-

cause physical punishment itself is identified as a causal factor for those negative outcomes. Therefore, the case for abolishing the physical punishment of children satisfies all of the three properties previously identified as necessary for a successful reduction in the severity of current forms of legal punishment. And if this alternative works in the case of children, it is an open possibility that something analogous could work for adults and make the non-cognitive strategy a promising one.

A challenge for the implementation of legal changes of the sort free will skepticism requires can now be more precisely stated. As in the case of children, what needs to be shown is that alternative ways of responding to criminals can be more effective in crime prevention *in virtue of being less harmful*. In other words, support for reduction in punishment severity can be expected if more severe punishment actually contributes to higher recidivism rates. A worth considering possibility is that some policies for crime prevention might turn out to work better in the absence of certain harmful aspects associated with punishment. For example, recidivism rates have been shown to be higher in overcrowded prisons (Farrington & Nuttall, 1980; Haney, 2006; Haney, 2015) and lower when education programs are available (Kim & Clark, 2013; Sellers, 2015). But these possibilities still fall short of showing that, say, shorter *sentences* would lead to lower recidivism rates. And, finally, an additional difficulty free will skeptics need to address is the unresolved issue about the relation between punishment severity and crime rates (von Hirsch et al., 1999; Friesen, 2012; Doob &Webster, 2003). These are all empirical questions that free will skeptics willing to use the non-cognitive strategy in an attempt to implement the legal changes required by their skepticism may need to address.

5 Concluding remarks

Free will skeptics such as Pereboom and Caruso argue that, even if humans do not have free will in the basic desert sense, we still have sufficient resources to deal with criminal behavior. This paper did not dispute the truth of this sufficiency claim if understood on purely conceptual or normative grounds. But I did put forth arguments that describe practical challenges that an attempt to implement this sort of proposal is likely to face. By considering how a central tenet derived from free will skepticism—a concern for reducing punishment severity as much as possible—could come to be implemented, I described two strategies for pursuing the relevant legal changes, as well as their respective requirements and challenges. If my arguments

are on the right track, the best strategy for changing the law in ways that reduce punishment severity is by seeking a reduction of people's natural desire to punish criminals. But this can only be achieved by finding alternatives to current punishment practices that cannot be implemented alongside current practices and that are more effective in preventing crime. If these conditions can be met, the prospects are high that a public support for the implementation of the relevant legal changes can be achieved.

As a final thought, I would like to emphasize that the concerns of this paper can be relevant beyond the scope of free will skepticism. As a first example, free will agnosticism, the view according to which no one *knows* whether humans have free will, has been said to have implications for punishment that are similar to those of Pereboom's account (Kearns, 2015, p. 249, n. 8). Second, practices of punishment currently accepted in some societies are sometimes claimed to be excessive. For example, some theorists consider capital punishment excessive on the argument that target criminals' history often include factors that reduce culpability (Steiker, 2011, p. 444-446). A third possibility would be to argue that punishment severity should be reduced in some cases *even if deserved*. One way to defend this is by distinguishing the conditions for an amount of punishment to be deserved from the conditions for an amount of punishment to be mandatory (see, Hart, 2008, p. 236; Steiker, 2011, p. 442; Zimmerman, 2015, p. 55). It is open for someone who makes this distinction to say that matters beyond desert—such as analyses of the effects or cost-effectiviness of current forms of punishment (see, e.g., Kleiman, 2009; Clear and Frost, 2014)—favor reducing punishment severity even if criminals deserve what is currently prescribed. The challenges and possibilities for the implementation of legal changes of the sort discussed in this paper can, therefore, be relevant far beyond the scope of free will skepticism. In fact, they concern any proposal that has a reduction in punishment severity as a consequence.[7]

References

Afifi, T. O.; Brownridge, D. A.; Cox, B. J. and Sareen, J. (2006). "Physical punishment, child abuse and psychiatric disorders", *Child Abuse and Neglect* 30: 1093-1103.

7 This work has been financially supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [BEX 4789/15-6] and by Fulbright Brasil. For comments on earlier versions of this paper, I would like to thank Lieke Asma, Gilberto Gomes, Stephen Kearns, Alfred Mele, Leonardo Ribeiro, Frank Sautter, Rogério Severo, Silvio Vasconcellos, and Flavio Williges. I also thank the audiences of the Writing Group at Florida State University, the 2016 UF/FSU Graduate Philosophy Conference, and the 3rd Workshop on Naturalism, and an anonymous referee for this journal.

Baker, T.; Metcalfe, C. F.; Berenblum, T.; Aviv, G. and Gertz, M. (2015). “Examining public preferences for the allocation of resources to rehabilitative versus punitive crime policies”, *Criminal Justice Policy Review* 26: 448-462.

Bedau, H. A. and Kelly, E. (2015). “Punishment”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*.

Caruso, G. D. (2016). “Free will skepticism and criminal behavior: A public health-quarantine model”, *Southwest Philosophy Review* 32: 25-48.

Clark, C. J.; Luguri, J. B.; Ditto, P. H.; Knobe, J.; Shariff, A. F. and Baumeister, R. F. (2014). “Free to punish: A motivated account of free will belief”, *Journal of Personality and Social Psychology* 106: 501-513.

Clear, T. R. and Frost, N. A. (2014). *The punishment imperative: The rise and failure of mass incarceration in America*. New York: New York University Press.

Corrado, M. L. (2013). “Why do we resist hard incompatibilism? Thoughts on freedom and punishment”. In: Nadelhoffer, T. A. (Ed.), *The future of punishment*, pp. 109-104. Oxford: Oxford University Press.

Doob, A. N. and Webster, C. M. (2003). “Sentence severity and crime: Accepting the null hypothesis”, *Crime and Justice* 30: 143-195.

Durrant, J. and Ensom, R. (2012). “Physical punishment of children: lessons from 20 years of research”, *Canadian Medical Association Journal* 184: 1373-1377.

Farrington, D. P. and Nuttall, C. P. (1980). “Prison size, overcrowding, prison violence, and recidivism”, *Journal of Criminal Justice* 8: 221-231.

Fischer, J. M. (1994). *The metaphysics of free will*. Oxford: Blackwell Publishers.

Friesen, L. (2012). “Certainty of punishment versus severity of punishment: An experimental investigation”, *Southern Economic Journal* 79: 399-421.

Gershoff, E. T. (2002). “Corporal punishment by parents and associated child behaviors and experiences: a meta-analytic and theoretical review”, *Psychological Bulletin* 128: 539-579.

Greene, J. and Cohen, J. (2004). “For the law, neuroscience changes nothing and everything”, *Philosophical Transactions of the Royal Society of London* 359: 1775-1785.

Haney, C. (2006). “The wages of prison overcrowding: Harmful psychological consequences and dysfunctional correctional reactions”, *Washington University Journal of Law & Policy* 22: 265-293.

Haney, C. (2015). “Prison overcrowding”. In: Cutler, B. L. & Zapf, P. A. (Ed.), *APA handbook of forensic psychology (Criminal investigation, adjudication, and sentencing outcomes)*, pp. 415-436. Washington, DC: American Psychological Association.

Hart, H. L. A. (2008). *Punishment and responsibility: Essays in the philosophy of law (Second Edition)*. Oxford: Oxford University Press.

von Hirsch, A.; Bottoms, A. E.; Burney, E. and Wikstrom, P.-O. (1999). *Criminal deterrence and sentence severity: An analysis of recent research*. Oxford: Hart Publishing.

Initiative (2015). “Global initiative to end all corporal punishment of children”, http://www.endcorporalpunishment.org/.

Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.

Kearns, S. (2015). “Free will agnosticism”, *Noûs* 49: 235-252.

de Keijser, J. W. and Elffers, H. (2009). “Punitive public attitudes: A threat to the legitimacy of the criminal justice system?”. In: Oswald, M. E.; Bieneck, S. & Hupfeld-Heinemann, J. (Ed.), *Social psychology of punishment of crime*, pp. 55-74. Chichester: Wiley-Blackwell.

Kim, R. H. and Clark, D. (2013). “The effect of prison-based college education programs on recidivism: Propensity Score Matching approach”, *Journal of Criminal Justice* 41: 196-204.

Kish, A. M. and Newcombe, P. A. (2015). ““Smacking never hurt me!” Identifying myths surrounding the use of corporal punishment”, *Personality and Individual Differences* 87: 121-129.

Kleiman, M. A. R. (2009). *When brute force fails: How to have less crime and less punishment*. Princeton: Princeton University Press.

McCorkle, R. C. (1993). “Research note: Punish and rehabilitate? Public attitudes toward six common crimes”, *Crime & Delinquency* 39: 240-252.

Mele, A. (2006). *Free will and luck*. Oxford University Press: Oxford University Press.

Mele, A. R. (1995). *Autonomous agents: From self-control to autonomy*. Oxford: Oxford University Press.

Monroe, A. E.; Brady, G. and Malle, B. F. (2016). “This isn’t the free will worth looking for: General free will beliefs do not influence moral judgments; agent-specific choice ascriptions do”, *Social Psychological and Personality Science* : 1-9.

Morse, S. J. (2013). “Compatibilist criminal law”. In: Nadelhoffer, T. A. (Ed.), *The future of punishment*, pp. 107-131. Oxford: Oxford University Press.

Nadelhoffer, T.; Shepard, J.; Nahmias, E.; Sripada, C. and Ross, L. T. (2014). “The free will inventory: Measuring beliefs about agency and responsibility”, *Consciousness and Cognition* 25: 27-41.

Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. New York: Oxford University Press.

Pereboom, D. and Caruso, G. D. (forthcoming). “Hard-incompatibilist existentialism:Neuroscience, punishment, and meaning in life”. In: Caruso, G. D. & Flanagan, O. (Ed.), *Neuroexistentialism: Meaning, morals, and purpose in the age of neuroscience*. New York: Oxford University Press.

Santin, T. R.; Vilanova, F.; Costa, Â. B.; Tocchetto, D. G.; Nadelhoffer, T. and Koller, S. H. (manuscript). “Evidências de validade do Inventário do Livre-Arbítrio (ILA) para a população brasileira”.

Sarkissian, H.; Chatterjee, A.; de Brigard, F.; Knobe, J.; Nichols, S. and Sirker, S. (2010). “Is belief in free will a cultural universal?”, *Mind & Language* 25: 346-358.

Schooler, J.; Nadelhoffer, T.; Nahmias, E. and Vohs, K. D. (2015). “Measuring and manipulating beliefs and behaviors associated with free will: The good, the bad, and the ugly”. In: Mele, A. R. (Ed.), *Surrounding free will: Philosophy, Psychology, Neuroscience*, pp. 72-94. New York: Oxford University Press.

Sellers, M. P. (2015). “Online learning and recidivism rates”, *International Journal of Leadership in Education* .

Steiker, C. (2011). “The death penalty and deontology”. In: Deigh, J. & Dolinko, D. (Ed.), *The Oxford handbook of the philosophy of the criminal law*, pp. 441-466. New York: Oxford University Press.

Suhling, S. and Greve, W. (2009). “The consequences of legal punishment”. In: Oswald, M. E.; Bieneck, S. & Hupfeld-Heinemann, J. (Ed.), *Social psychology of punishment of crime*, pp. 405-426. Chichester: Wiley-Blackwell.

Vargas, M. (2013). *Building better beings: A theory of moral responsibility*. Oxford: Oxford University Press.

Vilhauer, B. (2013). “Persons, punishment, and free will skepticism”, *Philosophical Studies* 162: 143-163.

Vohs, K. D. and Schooler, J. W. (2008). “The value of believing in free will: Encouraging a belief in determinism Increases cheating”, *Psychological Science* 19: 49-54.

Zimmerman, M. (2015). “Varieties of moral responsibility”. In: Clarke, R.; McKenna, M. & Smith, A. M. (Ed.), *The nature of moral responsibility: New essays*, pp. 45-64. New York: Oxford University Press.

ARTIGO 4: DOES DESIRE TO PUNISH AFFECT BELIEF IN FREE WILL? AN EXPERIMENTAL REPORT

Abstract: An experiment was conducted to test the hypothesis that a person's belief in the existence of free will can be causally influenced by how much the person desires to punish a criminal. Participants (N = 180) were randomly assigned to one of three conditions. Two conditions manipulated information about how effective punishing a fictitious criminal would be in preventing future criminal behavior relatively to an alternative program aiming at his social rehabilitation. A control condition included no such information. Results indicate that the manipulations led to significant reductions in the desire to punish, although no significant difference was found in both general and specific beliefs about free will.
Keywords: free will; punishment; blame; moral responsibility

1 Introduction

Philosophical tradition conceives of free will as a necessary condition for moral responsibility (O'Connor, 2013). Responsibility practices and attitudes, such as blaming, praising or even punishing an agent for an action can only be deserved if the action was done freely. Philosophers have investigated all sorts of questions about free will, including its nature, whether it can exist if determinism is true, and whether human beings have it. More recently, scientists from a variety of fields have also become interested in the subject, and a variety of empirical questions and data have since then helped to inform the debate. Two main lines of research can be identified in this empirical trend. One focuses on what may be called the cognitive science of decision-making and action, which investigates, for example, how humans actually reason and act (see, e.g., Libet et al., 1983; Libet, 1999; Gomes, 1999; Soon et al., 2008; Mele, 2009; Sinnott-Armstrong & Nadel, 2011; Sinnott-Armstrong, 2014). A second line of investigation is the psychology of belief in free will and responsibility practices, whose aim is to understand what laypersons think about free will and how their thoughts in this domain affect responsibility practices (see, e.g., Nahmias et al., 2005; Nahmias, Coates & Kvaran, 2007; Nichols & Knobe, 2007; Roskies & Nichols, 2008; Malle, Guglielmo & Monroe, 2014; Andow & Cova, 2015; Bear & Knobe, 2015; Feltz, 2015; Monroe, Brady & Malle, 2016).

In a recent contribution to the psychology of belief in free will, Clark, Luguri, Ditto, Knobe, Shariff and Baumeister (2014) investigated some of the causes of belief in free will.

They found that people report stronger belief in free will after considering an immoral action than after considering a neutral one. One of the proposed explanations for the result was that "a heightened desire to punish accounts for the heightened levels of both specific free will attributions and general free will belief" (p. 504). This explanation was supported by a mediation analysis, although it could not be tested experimentally (2014, p. 506). Clark et al. call the idea that belief in free will is reinforced by a desire to hold others morally responsible (and to punish the authors of wrongful actions) a motivational account of belief in free will.

The present study was designed to test experimentally the hypothesis that desire to punish causally influences general and specific beliefs about free will. Past research has found a connection between support for punishment and belief in punishment's efficacy in preventing crime. Thomas and Cage (1974) reported a correlation between perceptions of the effectiveness of punishment and the severity of sentencing for a variety of offenses (see also Miller & Vidmar, 1981). More recently, people tended to agree with a survey item stating that "People who perform harmful actions ought to be punished so that other potential offenders are deterred from committing similar harmful actions" (Nadelhoffer et al., 2014, p. 38, table 1). These findings, together with the motivational account of belief in free will, license the prediction that manipulating participants' confidence in the effectiveness of punishment would cause changes in both their desire to punish a criminal and in their beliefs about free will. The experiment reported below sought to test this prediction.

2 The present study

The present experiment was designed to test whether, and how, beliefs about the effects of punishment affect the desire to punish and beliefs about free will. The experiment included three conditions, all of which began with the description of a robbery. Condition A (Less Effective Punishment) then stated that the author of the robbery would have a high chance of repeating the criminal behavior if subjected to traditional punishment, but a very low chance of doing so if subjected exclusively to an alternative treatment program. Condition B (Similar Efficacy) described a situation in which either punishment or a treatment program would have a similar weak/moderate effect in preventing recidivism. Condition C (Punishment Only) was a control condition providing no information about the effects of punishment nor about alternative programs. The main hypotheses before running the experiment were that the desire to punish would be weaker in A than in C, and that, as a consequence, belief in free will would

also be weaker in A than in C. Condition B was designed to check whether either punishment or treatment would be preferred when both are thought to have similar effects on recidivism.

2.1 Method

Participants

Taking into account results from previous studies, the sample size for the present study was fixed at 60 valid responses per condition (N = 180). Desired responses were reached by inviting participants in Brazilian university groups on Facebook. Participants were balanced on sex (51% male) and from 15 different states (27% from Rio Grande do Sul). Most of them were young adults (58% were 21–30 years-old) who had attended undergraduate-level education (70%) in several fields (22% from Exact and Earth Sciences).

Procedure

The study was conducted online.[1] Participants were told that the study was about the relation between philosophical beliefs and social attitudes. After giving informed consent, participants were randomly assigned to one of the three conditions. In all conditions, participants read about a robbery committed by a 32 years-old man of initials M.C.D. M.C.D was said to have used a gun to threaten a person who was preparing to leave a market's parking area in order to steal the victim's motorcycle. Conditions differed in the following aspect designed to manipulate participants' desire to punish M.C.D.:

Condition A: participants were told that, according to a group of experts, M.C.D. satisfied the conditions for participation in a social reintegration program that would make him very unlikely to repeat the crime (10%), but that the program would not be efficacious if accompanied by punishment (80% of chance of repeating the crime). Participants were asked whether M.C.D. should receive the treatment program.

Condition B: participants were told that, according to a group of experts, the social reintegration program and traditional punishment would have similar effects on recidivism (40% chance of repeating the crime either participating in the program or being punished). Participants were also asked whether M.C.D. should receive the treatment program.

Condition C: no alternative to punishment was mentioned.

Henceforth, all participants indicated the amount of punishment (in years of imprisonment) M.C.D should receive (0 to 15 years; intervals of one year); this was used as the measure of

1 See *Apêndice A* for the materials used in the study.

desire to punish. In a 7-point Likert-scale (1 = totally disagree; 7 = totally agree), participants indicated whether M.C.D.'s action was free, and whether he was blameworthy and responsible for the action.[2] Participants also indicated their general beliefs about free will in a Brazilian version of the free will subscale of the Free Will Inventory (Nadelhoffer et al., 2014; Santin et al., manuscript). Finally, participants answered some demographic questions and were informed, upon conclusion, of the fictitious nature of the material they previously read.

2.2 Results

An analysis of variance revealed significant variation in the level of punishment considered appropriate across conditions, $F(2, 177) = 16.28$, $p < 0.001$, $\eta 2 = 0.155$. Post-hoc TukeyHSD comparisons indicated that punishment differed across all conditions ($ps < 0.05$). Punishment in condition A ($M = 3.25$; $SD = 4.89$) differed from condition B ($M = 5.63$, $SD = 4.84$), and both differed from the control condition ($M = 8.12$, $SD = 4.26$)—see Figure 1.

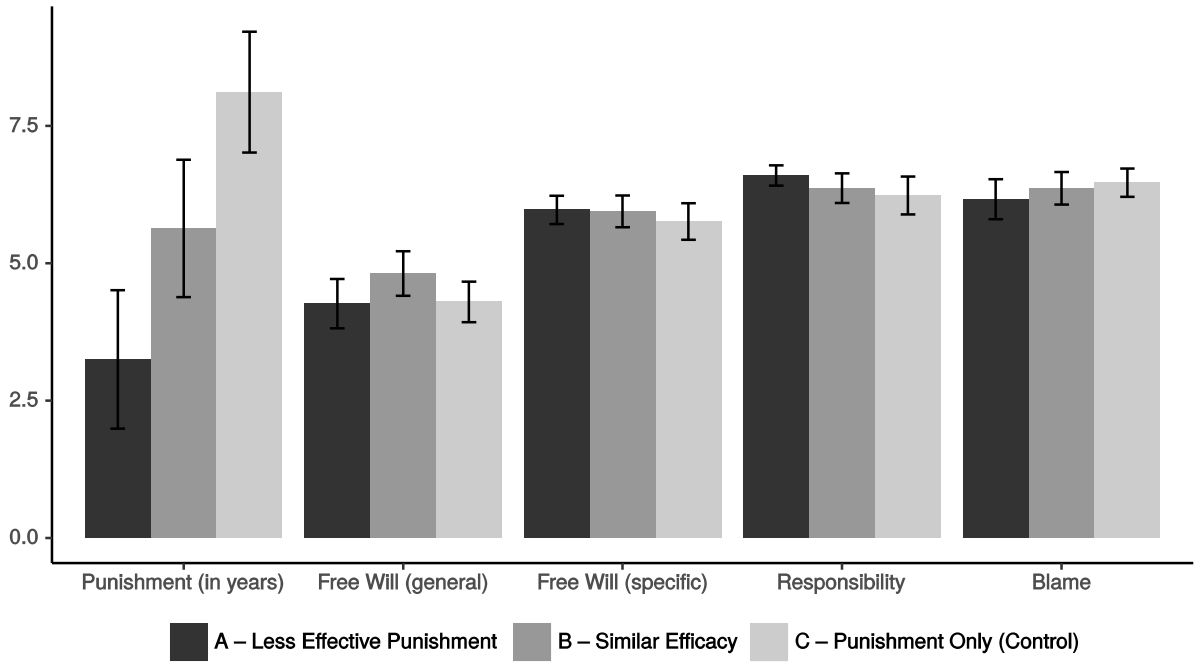

Figure 1 – Mean scores of dependent variables by condition. Error bars represent 95% confidence intervals.

2 Following Clark et al. (2014, p. 504), three statements assessed participants' beliefs about M.C.D.'s free will during the action: "M.C.D. exercised his free will during the robbery", "M.C.D. could have decided not to rob", and "M.C.D. decided to rob freely". The means for these three variables were used in subsequent analyses.

On the other hand, no significant difference was observed in the level of general belief in free will in each condition; $F(2, 177) = 2.277$, $p = 0.106$, $\eta 2 = 0.025$. Differences were also absent in specific belief in free will—$F(2, 177) = 0.61$, $p = 0.544$, $\eta 2 = 0.007$—and in specific attributions of responsibility—$F(2, 177) = 1.838$, $p = 0.162$, $\eta 2 = 0.020$—and blame—$F(2, 177) = 0.979$, $p = 0.378$, $\eta 2 = 0.011$; see also Figure 1.

2.3 Discussion

As predicted, manipulations of belief in the efficacy of punishment relative to a treatment program led to significant changes in the desire to punish the fictitious criminal. Contrary to predictions, however, significant changes were detected neither in general or specific beliefs about free will nor in ascriptions of responsibility and blame. These results show important connections to previous research.

Desire to punish and belief in punishment's efficacy. The present results extend previous research on the relation between support for punishment and belief in the efficacy of punishment for crime prevention. While previous research revealed correlation and association (Thomas & Cage, 1974; Miller & Vidmar, 1981; Nadelhoffer et al., 2014), the present results provide experimental evidence of a causal relation. Belief in punishment's efficacy not just correlates with desire to punish criminals, but also causally affects it.

Effect of desire to punish on belief in free will. Clark et al. (2014) found an effect of exposure to immoral behavior on beliefs about free will. A mediation analysis supported their hypothesis that participants' desire to punish mediated this effect, but they were unable to test the hypothesis experimentally (2014, p. 506). In the present experiment, a significant difference in the desire to punish did not lead to changes in either general or specific beliefs about free will. These results, therefore, failed to support experimentally Clark et al.'s hypothesis, although they leave untouched the general effect of increased belief in free will after exposure to immoral behavior. In sum, in the present study disbelief in the efficacy of punishment (as compared to an alternative treatment program) resulted in weaker desire to punish that was not followed by weaker belief in the existence of free will.

Policy and law. Discussions about the implications of different views about free will for punishment and law often take skepticism about the existence of free will as a reason for proposing changes in punishment practices. Greene and Cohen (2004) proposed that disbelief in free will would motivate support for more consequentialist and less retributive forms of punishment. The present results are compatible with the suggestion that some of these legal

changes could be supported even by people who firmly believe in free will. Similarly, Pereboom (2001, 2014) proposed that free will skepticism requires punishment practices to be modified in such a way that the level of harm criminals receive is not greater than the minimum required for social protection. On the assumption that current punishment practices sometimes inflict more than that minimum, it follows that free will skepticism requires a reduction in the severity of current punishment practices. The results described above suggest that public support for these sorts of changes can also be achieved in the absence of free will skepticism. Insofar as less severe measures that are at least as effective as punishment in reducing future criminality are available, one can expect them to receive substantial support from the public even if belief in free will remains stable.

Limitations. One of the results of the present study was a failure to reject the null hypothesis that the desire to punish does not causally affect general and specific beliefs about free will. This failure might be due to some limitation of the study itself rather than to the truth of the null hypothesis. First, there is a possibility that the effect failed to be replicated because of some cultural difference between Brazil and the United States (where Clark et al. 2014 found contrary evidence). Although this possibility cannot be excluded at this point, it should at least be noted that the present study employed a validated translation of a scale originally developed in the United States. In addition, it would be a relevant finding if there were cultural differences in this domain because other studies found cross-cultural similarities in beliefs about free will (Sarkissian et al., 2010). Another possibility that cannot be excluded is that a larger sample size would be required to detect the effect of interest. Although the present sample size (60 participants per condition) was larger than Clark et al.'s (48 participants per condition in their second study), other studies relied on even larger sample sizes to find significant variation in belief in free will—e.g. 133 per cell in Monroe, Brady and Malle (2017, p. 2). This is another possibility that cannot be excluded at this moment. A further limitation has to do with the fact that, in the present study, the desire to punish was reduced with the use of information about the consequences of punishment. The next section discusses this issue in the context of a broader literature on the relation between belief in free will and desire to punish.

3 Conflicting past results

The results just presented conflict with some of Clark et al.'s (2014) findings about the effects of desire to punish on general and specific beliefs about free will. This situation is not new in the broader literature on the relation between belief in free will and desire to punish. Conflicting results have been reported in correlational as well as experimental studies on the relation between belief in free will and desire to punish. In this section I review these conflicting results and discuss their implications for the interpretation of the present results.

Past studies have variably found general belief in free will to be positively associated with desire to punish, not associated, and even negatively associated. Krueger et al. (2014) categorized crimes as having either high or low affect and found an association between belief in free will and punishment severity only for low affect crimes. Viney, Waldman and Barchilon (1982) also found a negative correlation in some cases, but did not find it when the death penalty was considered (which can be interpreted as high affect). In contrast, Shariff et al. (2014, study 2) found an association between punishment severity and belief in free will for a crime described as "beating a man to death" (which can be interpreted as involving high affect). And, for low affect cases, Krueger et al. (2014) found an association between belief in free will and punishment severity, although Monroe, Brady and Malle (2017) did not for a crime described as one worker (in Amazon Mechanical Turk) stealing US$ 0.80 from another. Finally, for studies considering crimes of diverse or unspecified affect, Stroessner and Green (1990) found a correlation between belief in free will and punishment severity while Viney, Parker-Martin and Dotten (1988) did not. Table 1 summarizes these conflicting results.

Affect	No association	Positive association	Negative association
High	Viney et al. 1982 Krueger et al. 2014	Shariff et al. 2014	
Low	Monroe et al. 2016	Krueger et al. 2014	Viney et al. 1982
Unspecified	Viney et al. 1988	Strossener et al. 1990	

Table 1 – Summary of conflicting results in previous literature on the relation between general belief in free will and punishment recommendations.

One can infer from Table 1 that the uncontrolled influence of affect cannot be the whole story behind this remarkable amount of conflicting results. Even if affect can be in-

voked to explain differences within Viney, Waldman and Barchilon's (1982) and Krueger et al.'s (2014) studies, theirs and other's results on the relation between belief in free will and desire to punish also conflict when affect can be assumed to be similar. So what else might explain these divergences?

Shariff et al. (2014) and Viney, Parker-Martin and Dotten (1988) explore an additional factor that may help to make progress on these issues. Shariff et al. specifically looked at the effects of belief in free will on *retributive* punishment. In order to do that, participants were invited to recommend punishment for a target said to have already been successfully rehabilitated, in such a way that any concern for the effects of punishment on the criminal should be neutralized. Under such circumstances, there was an effect of belief in free will on punishment severity, as well as a correlation between them. When Monroe, Brady and Malle (2017) found no such effect, for example, they were looking at punishment for no specific reason.[3] Relatedly, Viney, Parker-Martin and Dotten (1988) found that a single individual can adopt either a more consequentialist or a more retributivist approach to punishment depending on the nature of the crime considered. Finally, McFatter (1978) found that the rationale behind punishment recommendations affects the relation between the seriousness of a crime and the punishment recommended.

These further results suggest that it is relevant to consider the rationale behind punishment recommendations in order to assess the relation between belief in free will and desire to punish. This consideration helps in the interpretation of the results from the present study. On the question of whether belief in free will is causally *affected* by desire to punish, the present study failed to find an effect of desire to punish on beliefs about free will. But in light of the studies just considered, it would be more appropriate to say that the present study failed to find an effect of desire to punish *for consequentialist reasons* on general and specific beliefs about free will, since the rationale behind the reduction in punishment severity was of a consequentialist kind. And although this is a failure to support experimentally Clark et al.'s (2014) hypothesis that desire to punish for *unspecified* reasons causally affects belief in free will, it would be consistent with the present results if desire to punish for retributive reasons

3 Also, there is reason to think that the concept of punishment underlying the study was not retributive at all. After saying that "Worker 1 had stolen US$ 0.80" from a second worker, "[p]articipants decided how much to punish Worker 1 using a slider bar incremented by US$ 0.10 between US$ 0.00 and US$ 0.80." (Monroe, Brady & Malle, 2017, p. 3). But it is hard to see how taking away stolen money (or part of it) from a stealer could possibly qualify as punishment, much less punishment for retributive reasons. Under normal circumstances, punishment is what a stealer receives after the stolen object is returned (when that is possible).

turned out to causally affect beliefs about free will. Also, the present results do not contradict the hypothesis that general belief in free will causally affects the desire to punish for retributive reasons (Shariff et al., 2014), even though others failed to find evidence for an effect on punishment for unspecified reasons (Monroe, Brady & Malle, 2017). It is a question for future investigation to determine more conclusively whether the relation between desire to punish for retributive reasons and beliefs about free will really involves bidirectional causality. Anyway, the main suggestion sketched above is that such an investigation will benefit from controlling for the rationale behind punishment recommendations.

4 Conclusion

The present experiment, although successful in reducing people's desire to punish a criminal, failed to support the hypothesis that reduced desire to punish results in reduced belief in free will. Contrary to a suggestion by Clark et al. (2014), neither general nor specific beliefs about free will changed after a significant reduction in the desire to punish. This null result, however, should be interpreted with caution. Some past studies indicate that the connection between desire to punish and beliefs about free will varies depending on the rationale behind punishment recommendations. Because most past studies failed to monitor the rationale behind punishment recommendations, uncontrolled variations in this regard may account for at least part of the seemingly inconsistent findings. If this suggestion is correct, future studies will benefit from taking note of the reasons that guide punishment recommendations in each particular case. Finally, the results reported here support the claim that beliefs about the efficacy of punishment (relative to the alternatives available) causally influence the desire to punish, at least on a general level. This information is relevant when it comes to assessing the public acceptability of proposed legal changes (especially in the criminal law) and policies for the prevention of violence and crime. Crucially, because belief in free will is so strong and widespread, it is relevant to know that many people are likely to support, under the appropriate conditions, policies that involve less severe punishment even if their beliefs in free will remain strong.

References

Andow, J. and Cova, F. (2016). "Why compatibilist intuitions are not mistaken: A reply to Feltz and Millan", *Philosophical Psychology* 29: 550-566.

Bear, A. and Knobe, J. (2016). “What do people find incompatible with causal determinism?”, *Cognitive Science* 40: 2025-2049.

Clark, C. J.; Luguri, J. B.; Ditto, P. H.; Knobe, J.; Shariff, A. F. and Baumeister, R. F. (2014). “Free to punish: A motivated account of free will belief”, *Journal of Personality and Social Psychology* 106: 501-513.

Feltz, A. (2015). “Experimental philosophy of actual and counterfactual free will intuitions”, *Consciousness and Cognition* 36: 113-130.

Gomes, G. (1999). “Volition and the readiness potential”, *Journal of Consciousness Studies* 6: 59-76.

Greene, J. and Cohen, J. (2004). “For the law, neuroscience changes nothing and everything”, *Philosophical Transactions of the Royal Society of London* 359: 1775-1785.

Krueger, F.; Hoffman, M.; Walter, H. and Grafman, J. (2014). “An fMRI investigation of the effects of belief in free will on third-party punishment”, *SCAN* 9: 1143-1149.

Libet, B. (1999). “Do we have free will?”, *Journal of Consciousness Studies* 6: 47-57.

Libet, B.; Gleason, C. A.; Wright, E. W. and Pearl, D. K. (1983). “Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act”, *Brain* 106: 623-642.

Malle, B. F.; Guglielmo, S. and Monroe, A. E. (2014). “A theory of blame”, *Psychological Inquiry* 25: 147-186.

McFatter, R. M. (1978). “Sentencing Strategies and Justice: Effects of Punishment Philosophy on Sentencing Decisions”, *Journal of Personality and Social Psychology* 36: 1490-1500.

Mele, A. (2009). *Effective intentions: The power of conscious will*. Oxford: Oxford University Press.

Miller, D. T. and Vidmar, N. (1981). “The social psychology of punishment reactions”. In: Lerner, M. J. & Lerner, S. C. (Ed.), *The justice motive in social behavior*, pp. 145-172. : Springer.

Monroe, A. E.; Brady, G. and Malle, B. F. (2017). “This isn’t the free will worth looking for: General free will beliefs do not influence moral judgments; agent-specific choice ascriptions do”, *Social Psychological and Personality Science* 8: 191-199.

Nadelhoffer, T.; Shepard, J.; Nahmias, E.; Sripada, C. and Ross, L. T. (2014). “The free will inventory: Measuring beliefs about agency and responsibility”, *Consciousness and Cognition* 25: 27-41.

Nahmias, E.; Coates, D. J. and Kvaran, T. (2007). “Free will, moral responsibility, and mechanism: Experiments on folk intuitions”, *Midwest Studies in Philosophy* 31: 214-242.

Nahmias, E.; Morris, S.; Nadelhoffer, T. and Turner, J. (2005). “Surveying freedom: Folk Intuitions about free will and moral responsibility”, *Philosophical Psychology* 18: 561-584.

Nichols, S. and Knobe, J. (2007). “Moral responsibility and determinism: The cognitive science of folk intuitions”, *Noûs* 41: 663-685.

O'Connor, T. (2013). “Free will”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, .

Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. New York: Oxford University Press.

Roskies, A. L. and Nichols, S. (2008). “Bringing moral responsibility down to earth”, *The Journal of Philosophy* 105: 371-388.

Santin, T. R.; Vilanova, F.; Costa, Â. B.; Tocchetto, D. G.; Nadelhoffer, T. and Koller, S. H. (manuscript). “Evidências de validade do Inventário do Livre-Arbítrio (ILA) para a população brasileira”.

Sarkissian, H.; Chatterjee, A.; de Brigard, F.; Knobe, J.; Nichols, S. and Sirker, S. (2010). “Is belief in free will a cultural universal?”, *Mind & Language* 25: 346-358.

Shariff, A. F.; Greene, J. D.; Karremans, J. C.; Luguri, J. B.; Clark, C. J.; Schooler, J. W.; Baumeister, R. F. and Vohs, K. D. (2014). “Free will and punishment: A mechanistic view of human nature reduces retribution”, *Psychological Science* 25.8: 1-8.

Soon, C. S.; Brass, M.; Heinze, H.-J. and Haynes, J.-D. (2008). “Unconscious determinants of free decisions in the human brain”, *Nature Neuroscience* 11: 543-545.

Stroessner, S. J. and Green, C. W. (1990). “Effects of belief in free will or determinism on attitudes toward punishment and locus of control”, *The Journal of Social Psychology* 130: 789-799.

Thomas, C. W. and Cage, R. J. (1974). “Correlates of public attitudes toward legal sanctions”, *Paper presented at the meeting of the Western Sociological and Anthropological Association, Banff, Canada*: 1-31.

Viney, W.; Parker-Martin, P. and Dotten, S. D. H. (1988). “Beliefs in free will and determinism and lack of relation to punishment rationale and magnitude”, *The Journal of General Psychology* 115: 15-23.

Viney, W.; Waldman, D. A. and Barchilon, J. (1982). “Attitudes toward punishment in relation to beliefs in free will and determinism”, *Human Relations* 35: 939-950.

Sinnott-Armstrong, W. & Nadel, L. (Ed.) (2011). *Conscious will and responsibility: A tribute to Benjamin Libet*. Oxford: Oxford University Press.

Sinnott-Armstrong, W. (Ed.) (2014). *Moral psychology: Free will and moral responsibility*. Cambridge, MA: MIT Press.

ARTIGO 5: QUESTIONS FOR A SCIENCE OF MORAL RESPONSIBILITY*

Abstract: In the last few decades, the literature on moral responsibility has been increasingly populated by scientific studies. Studies in neuroscience and psychology, in particular, have been claimed to be relevant for discussions about moral responsibility in a number of ways. And at the same time, there is not yet a systematic understanding of the sort of questions a science of moral responsibility is supposed to answer. This paper is an attempt to move toward of such an understanding. I discuss three models for framing scientific questions relevant to an investigation of moral responsibility. The favored model—the Enhancement model—proposes that a science of moral responsibility has two descriptive tasks. First, science can describe the causes and effects of the many sorts of responses that constitute the human practices of moral responsibility, such as praise, blame, and punishment. And, second, science can describe how modifications aiming at the improvement of such practices can be achieved. Relatively to the other models to be considered, the Enhancement model is broader in scope and less tied to the traditional philosophical agenda on moral responsibility.
Keywords: moral responsibility; free will; science; psychology; neuroscience

1 Introduction

Among the questions that science can in principle answer, which ones matter for an investigation of moral responsibility? Despite the increase in scientific studies related to moral responsibility and free will, little effort has been made to systematically define the questions for such an investigation. This paper discusses three general models that can be used to frame answers to the question posed, one of which is favored in the end. And while these answers are mainly a matter for philosophical debate, I hope the reflection can help those interested in the scientific study of moral responsibility think more explicitly about the questions they are—or could be—trying to answer.

The paper is structured as follows. Section 2 focuses on what I call the Minimal model, which gives science a relatively small set of questions about moral responsibility. Section 3 discusses a model that arises in the context of the experimental philosophy movement,

* This paper has been published in the *Review of Philosophy and Psychology* and is available at Springer via <http://dx.doi.org/10.1007/s13164-017-0360-5>.

the Folk intuitions model. Section 4 presents my favored model, namely, the Enhancement model. Relatively to the previous models, the Enhancement model innovates by integrating a broader set of scientific questions about moral responsibility, and by giving science not just purely descriptive questions (e.g., what does causally affect responsibility ascriptions?) but also questions that pertain to a more revisionist enterprise (e.g., how can morally objectionable aspects of responsibility ascriptions be modified?).

2 The Minimal model

Among the models for framing the scientific questions that matter for an investigation of moral responsibility to be considered here, the Minimal model is (although tacitly) the most widely assumed in contemporary philosophical discussion. According to this model, science is supposed to contribute with the assessment of the empirical conditions postulated by different views about the nature of moral responsibility. By telling whether such conditions are ever satisfied, science helps to tell whether there are any morally responsible human beings, i.e., beings that satisfy conditions that are necessary (or sufficient) for moral responsibility. I call this model 'minimal' because it denies science any further role in the investigation of moral responsibility. In particular, it denies that science can help to determine what the conditions of moral responsibility are.

Robert Kane's (1996) work on free will and moral responsibility is a clear example of the Minimal model at work. Kane is an incompatibilist about free will and determinism, and as such he assumes that the falsity of determinism is a necessary condition for the existence of free and morally responsible agents.[1] Kane defends incompatibilism on the basis of traditional philosophical arguments, but he turns for science when the question is whether determinism is actually false:

> There are empirical aspects of the free will issue that mere philosophical speculation cannot co-opt. If free will of a nondeterminist kind should exist in nature, then the atoms must somewhere 'swerve' to make room for it, and they must swerve in places where it matters—in the brain, for example. (Kane, 1996, p. 17)

The investigation about the existence of such undetermined events, according to Kane, turns on questions about the physical world in general—and our brains, in particular. Kane's project, therefore, assumes a conception about the role of science in the investigation of moral responsibility that is in accordance with the Minimal model as previously defined.

1 Determinism can be understood as the thesis that the complete state of the world at a given time and the natural laws that are true in this world fix the state of the world at any subsequent time (Hoefer, 2016).

Compatibilist theories—those that take free will and moral responsibility to be possible even if determinism is true—may also have empirically assessable commitments. John Martin Fischer and Mark Ravizza (1998), for example, argue that responsiveness to reasons (a capacity to use reasons to guide one's behavior) is a necessary condition of moral responsibility. Like in Kane's work, the condition is justified using philosophical arguments. But the claim that human beings are responsive to reasons as required is itself amenable to scientific scrutiny. For example, some other philosophers have examined whether the situationist literature—empirical results showing that minor aspects of the environment can influence our behavior in usually unnoticed ways—undermines reasons responsiveness (see, e.g., Nelkin, 2005; Schlosser, 2013; Vargas, 2013; Shepherd, 2015). This is another instance of the Minimal model at work.

Further examples arise in discussions about the impact of the so-called Libet-style experiments on a number of proposed conditions for free will and moral responsibility. Libet et al. (1983) found a pattern of neural activity called 'readiness potential' (RP) to precede simple spontaneous movements such as flexing a finger or a wrist, and also the moment subjects reported to have consciously felt the urge to execute those movements. Recent studies in this paradigm were able to use patterns of neural activity to predict which button (right or left) a subject would press approximately 7 seconds before actual movement, although with a relatively low accuracy (Soon et al., 2008; Haynes, 2011). It has been widely discussed whether and how these results do or could impact the satisfaction of a number of proposed conditions for the existence of free will and moral responsibility. For example, do the results show that decisions are deterministically caused by neural events (Roskies, 2006, 2011, 2014; Balaguer, 2009; Haynes, 2011; Misirlisoy & Haggard, 2014; Nahmias, 2014; Fischborn, 2016, 2017; Roskies & Nahmias, 2017)? Do they show that decisions, intentions, or conscious mental states in general do not play the role they are supposed to do in the generation of actions (Nahmias, 2002; Wegner, 2002; Pockett, Banks & Gallagher, 2006; Gomes, 2007; Mele, 2009; Baumeister, Masicampo & Vohs, 2011; Schlosser, 2012; Marques, 2015; Asma, forthcoming)? These questions, once again, share the assumption that science can help to tell whether the conditions of moral responsibility are ever satisfied.

The examples considered above suggest that the Minimal model has been very influential in guiding the incorporation of scientific results into the philosophical literature on free will and moral responsibility. One of the model's main strengths is that it transfers to science

those questions that cannot be answered by philosophy alone.[2] But as I will argue in the following sections, the Minimal model fails to include a number of scientific questions that may also be relevant in an investigation of moral responsibility. In other words, even if it is granted that the Minimal model covers relevant and legitimate scientific *questions*, it may still fail as a *framework* for the investigation of moral responsibility as a whole.[3] The next section considers a second model that covers some aditional questions.

3 The Folk intuitions model

According to the Minimal model, the only goal of a science of moral responsibility is to help to determine whether the conditions of moral responsibility are ever satisfied. A different model arises within the experimental philosophy movement. Because the term 'experimental philosophy' itself is associated with a number of different projects (Alfano & Loeb, 2016, section 2), I reserve the term 'Folk intuitions model' to refer to one of them in particular. In this section I describe the envisaged model and argue that it also fails to capture the full range of questions that matter for a science of moral responsibility.

The methodology of experimental philosophy was first applied to questions related to moral responsibility in an attempt to advance the traditional dispute among compatibilists and incompatibilists (Nahmias et al., 2005). As I said earlier, this is a dispute about what the conditions of moral responsibility are—incompatibilits affirm, while compatibilists deny, that the falsity of determinism is among such conditions. Nahmias et al. (2005) noted that "so many philosophers [compatibilists and incompatibilists alike] claim that their own position has the most intuitive appeal and best fits our ordinary conception of free will and our practices of responsibility attribution" (p. 563). Because compatibilists and incompatibilists cannot be both right in their claims, Nahmias et al. proposed that surveying non-experts' actual ascriptions of responsibility when exposed to relevant thought experiments could help to better assess the intuitive appeal of each view. The results—the authors suggested—would help not to solve

2 It should be noted that the Minimal model is consistent with the possibility that *none* of the conditions for the existence of moral responsibility can be assessed scientifically. A defense of this thesis would involve showing that science cannot help to assess the truth of any of the suggested conditions of moral responsibility. Although this position has been defended regarding *some* of the conditions discussed—for example, observations have been said to be insufficient to justify a choice between stochastic and deterministic models in some cases (Suppes,1993, but see Werndl, 2013, 2016)—an argument would still be needed to show that the same holds for all of them.

3 Similar remarks apply to the other models to be discussed below. Accordingly, while there is no conflict among the scientific questions that arise within each of the models, there may be conflict among the integral views about the investigation of moral responsibility the models assume.

the old philosophical disputes, but to better assign the burden of proof, in such a way that more intuitive views could start off with an advantage or "squatters' rights" (p. 564). In addition, the results could help to assess how revisionary different views about free will and moral responsibility are (pp. 564-565).

Nahmias et al.'s (2005) own results suggested that most people are willing to attribute free will and moral responsibility to agents in deterministic scenarios. Nichols and Knobe (2007) presented the first evidence that most people deny that agents in deterministic scenarios are morally responsible, at least if the scenario is described in more abstract terms. The debate has remained inconclusive since then. While some results and interpretations support the claim that most people are naturally compatibilists (Nahmias, Coates & Kvaran, 2007; Murray & Nahmias, 2014; Andow & Cova, 2016; Monroe, Brady & Malle, 2017), others support the claim of a natural incompatibilism (Rose & Nichols, 2013; Feltz, 2015; Bear & Knobe, 2016), and some other results are mixed (Roskies & Nichols, 2008; Nadelhoffer et al., 2014). As things stand now, the question remains open (Schooler et al., 2015) and some have suggested that it may not be even worth looking for a unified ordinary view in this domain (Feltz, Cokely & Nelson, 2016).

I take the research program just reviewed to instantiate what I call the Folk intuitions model. The key assumption in this model is that science can help to determine the conditions for the existence of morally responsible agents. The Folk intuitions model, therefore, rejects the Minimal model's assumption that science can only help to tell which of the conditions for moral responsibility are actually satisfied. One of the strengths of the Folk intuitions model is that it helps to keep philosophical theories close to those human activities they are supposed to be about. Alfred Mele claimed that a risk for the philosophical discussion of free will is to have "nothing more than a philosophical fiction as its subject matter'' (Mele, 2001, p. 27). The Folk intuitions model can help to avoid the risk by providing a more accurate portrait of folk concepts and judgments. The Folk intuitions model, moreover, can also incorporate some of the scientific questions assigned to science within the Minimal model, even though the models themselves diverge. In this enlarged version, science contributes with both the specification of the conditions of moral responsibility and the assessment of their satisfaction. However, the Minimal model and the Folk intuitions model (even in its enlarged version) still fail to give science the full role it can have in an investigation of moral responsibility. Both models arise out of a philosophical agenda which does not (and does not have to) represent all of

the scientific questions about moral responsibility we may be interested in answering. In the next section, I describe some of the questions the Minimal model and the Folk intuitions model leave behind. These further questions set the stage for developing an alternative view about the science of moral responsibility.

4 The Enhancement model

Although the models so far considered propose legitimate scientific questions about moral responsibility, they are far from complete. At best, they give a complete picture of the scientific questions that are relevant for the kind of philosophical work on moral responsibility that has been predominant. In this section I describe an alternative model—the Enhancement model—which gives a unified shape to an investigation aiming at the understanding and improvement of the social practices associated with moral responsibility. I begin by describing the general features of the Enhancement model, and then I illustrate how it integrates a broader variety of results, some of which have not been recognized so far as pertaining to a broader and shared enterprise on moral responsibility.

It was implicit in the discussion thus far that a central objective in an investigation of moral responsibility is to specify under what conditions it is acceptable or appropriate to respond to an agent in certain ways. Of course it is open to discussion what exactly 'appropriate' means in this context, as well as what the intended responses are (see, e.g., Rosen, 2015; Zimmerman, 2015), but the same general objective could be set for various candidates. The Enhancement model proposes that pursuing the general objective involves at least three broad types of tasks—namely, conceptual, descriptive, and normative tasks—and that science's primary role is to execute the descriptive tasks. Within the Enhancement model, however, this descriptive task goes beyond the description of responsibility practices as they currently stand. The model leaves open the possibility that a normative assessment can find current practices problematic and in need of revision. In such cases, science has the further task of describing the available routes for change.

An example of a conceptual task is the definition of terms used in the investigation of moral responsibility. This is a traditional business of philosophers, and key questions include what it is to be an agent, what responsibility involves, what it is to blame, punish, or praise someone, and so on. Conceptual assumptions are involved in the very definition of the central objective of the investigation of moral responsibility mentioned in the previous paragraph.

Once this initial conceptualization of key notions is available (even if provisional), the Enhancement model gives science its first task, which is to describe how responsibility practices actually operate. This description focuses on what I will call 'responsibility responses', i.e., the kinds of responses or reactions that are characteristic of moral responsibility, which include praise, blame, and punishment as paradigmatic cases. And the description should aim at providing us with a comprehensive understanding of the causes and effects of those responses. I take this descriptive enterprise to be a task for a plurality of scientific disciplines and fields —certainly including many fields of psychology, neuroscience, and the social sciences—and for a plurality of methodological approaches within those disciplines and fields.

A second task science gets within the Enhancement model is conditional on a normative assessment of the workings of ordinary responsibility practices as revealed in the first task. The Enhancement model assumes that any aspect of actual responsibility responses may be found faulty upon normative scrutiny. Like the descriptive tasks assigned to science, this normative assessment is not supposed to be restricted to a single field or approach—potential candidates include, but are not limited to ethical, social, political, and legal normative theories. For the purposes of this paper, the most important consequence of this normative assessment is that it generates a further potential role for science. If ordinary responsibility practices turn out to have any problematic aspects, then it is desirable to find ways for improvement. Therefore, the second task science gets within the Enhancement model is to describe the possible routes for effectively changing those problematic aspects. And, because interventions are likely to also involve potential risks or costs, the choice to pursue change in any of the available ways is likely to involve further normative assessments.[4]

In sum, the Enhancement model gives the science of moral responsibility two main tasks. The first task is to describe the causes and effects of responsibility responses. And the second task is to describe possible ways for changing problematic aspects associated with the workings of those responses. In the remainder of this section I illustrate how these tasks can be and have already been carried on. But I emphasize from the start that although the philosophical and scientific tasks (and the two scientific tasks themselves) can be clearly distin-

4 The influence of values is often considered in discussions about the objectivity of science. The Enhancement model allows for research questions and applications of scientific results to be affected by those values assumed in the normative assessment of responsibility practices. Most philosophers, however, do not take this sort of interference to be a threat to scientific objectivity. As Fischer et al. (2007, sec. 3.1) note, "[t]he real debate is about whether or not the 'core' of scientific reasoning—the gathering of evidence and the assessment and acceptance of scientific theories—is, and should be, value-free.".

guished for the purposes of reflection, they are often mixed in actual investigation. In illustrating how the proposed tasks can be realized, I also hope to show how the Enhancement model integrates within the science of moral responsibility a number of studies that so far have been largely ignored in more philosophically-oriented work on the topic.

The Path model of blame developed by Malle, Guglielmo and Monroe (2014) exemplifies what the first descriptive task can look like. This model hypothesizes that ordinary attributions of blame are guided by a process that involves the evaluation of certain conditions. First, blame only emerges if an agent is judged to have caused an event or outcome that violates a norm (2014, p. 151). If this condition is met, then a second step involves judging "whether the agent brought about the event intentionally" (p. 151). Once intentionality is assessed, two routes may follow. If the event was intentionally brought about, then the amount of blame depends on whether the agent had good reasons for the action, resulting in "minimal blame if the agent was justified in acting this way; maximal blame if the agent was not justified" (p. 151). If no intentionality is detected, then the amount of blame is guided by an assessment of the agent's obligations and capacity to prevent the undesirable event, in such a way that the existence of such obligation and capacity results in more blame, and their absence results in low or no blame (see 2014, p. 151, figure 2).

The Path model of blame contrasts with the kind of scientific result that is emphasized within the Minimal and the Folk intuitions models. Granted, *philosophical* theories of moral responsibility propose conditions under which blame is appropriate, many of which are roughly in accordance with the Path model of blame. But the scientific questions within the Minimal model concern whether or not those conditions are satisfied in most or in special cases, regardless of whether considerations about those conditions are causally operative or not in ordinary blame ascriptions. Similarly, even though the Folk intuitions model assumes that proposed conditions of moral responsibility should approximate ordinary ascriptions of blame, the questions assigned to science have been largely concerned with solving controversial philosophical disputes about what is intuitive (as in the compatibilism versus incompatibilism issue), and not with providing a comprehensive explanation of blame ascriptions. The Enhancement model emphasizes the goal of describing the causes and effects of ordinary responsibility reactions in a comprehensive and accurate way whether or not such causes and effects capture what has been emphasized in more traditional philosophical theories.

Similar empirical questions can be raised about any other kind of response associated with moral responsibility. Consider punishment. Throughout adult life, the primary use of punishment is supposed to be regulated by criminal justice systems. Despite this fact, there is evidence that external factors—such as the gender, race, and age of punished individuals—influence punishment patterns (Steffensmeier, Ulmer & Kramer, 1998; Spohn & Holleran, 2000). Different forms of punishment are also often used by parents in the process of rearing their children. Regarding *physical* punishment, in particular, studies indicate that predictive factors of its use include the "perception of the seriousness and intent of the child misbehavior" (Ateah & Durrant, 2005). On the positive side of responsibility responses, gratitude, for example, has been shown to arise when one has benefited from someone else's intentional behavior (McCullough, Kimeldorf & Cohen, 2008). The Enhancement model takes all these studies as contributions to the task of describing the workings of ordinary responsibility responses.

Now, as noted earlier, the causes and effects of responsibility responses may not coincide with the conditions for moral responsibility postulated by philosophical theories, nor need all of them be accepted as desirable. In more general terms, we may have normative reasons to think that the way responsibility responses operate needs modification. And this possibility sets the stage for the second task assigned to science within the Enhancement model, namely, the investigation about how to effect changes in problematic aspects of responsibility practices.

The Path model of blame predicts that blame arises when someone violates a norm intentionally. But, under certain circumstances, it seems disputable that blaming responses arising in this way are acceptable, much less recommended. Hanna Pickard (2011; 2013), for instance, has developed a notion of responsibility without blame. In certain therapeutic settings, she argues, blame is counterproductive for recovery even though capacities that are required for moral responsibility are significantly operative. But it is beneficial for treatment that patients see themselves as responsible for their actions. How then can responsibility and blame be separated? Pickard's proposal is that therapists need to withhold the expression of negative emotions and judgments that give blame its characteristic 'sting'. Instead, they should appeal to what she calls 'detached blame': "Detached blame consists in judgments of blameworthiness, and may further involve correspondingly appropriate revisions of intentions, the imposition of negative consequences, and accountability and answerability" (Pickard, 2013, p.

1146). But this is to be distinguished from an ‘affective’ form of blame, which characteristically involves the blamer feeling herself entitled to express certain negative reactive attitudes and emotions. The affective form of blame, she says, is counterproductive for treatment and active steps should be taken to prevent its manifestation in therapeutic contexts. Pickard’s proposal is in accordance with what I call a normative assessment of responsibility practices as well as with the scientific description of alternatives. Of course, it can be replied that therapeutic contexts are too peculiar to imply any general verdict about responsibility norms, and so it may be worth looking at more common contexts as well.

Research on relationships and conflict resolution provides a more ordinary case for reflection. Counselors working in schools are sometimes advised to seek conflict resolution in ways that make students stop “from blaming one another or pointing fingers at whose fault it is” (Brinson, Kottler & Fisher, 2004, p. 300). In workplace settings, blaming was found positively associated with seeking revenge and negatively associated with reconciliation (Aquino, Tripp & Bies, 2001). And marital satisfaction can be negatively affected when spouses keep blaming one another for their problems and faults (Madden & Janoff-Bulman, 1981; Kubany et al., 1995). Malle, Guglielmo and Monroe (2014, pp. 173-174) themselves also describe a “darker side” of blame, which includes blame expressed with exaggerated emotional intensity, blame followed by complaints and countercomplaints that increase the odds of conflict, and situations where the value of repairing a relationship is missing, or a powerful social group creates “scapegoats” to be blamed for the larger community’s problems. The physical punishment of children raises similar concerns. The practice has become increasingly condemned world-wide due to its long-lasting effects on children’s behavior, relationships, and mental health (Ateah & Durrant, 2005; Afifi et al., 2006; Lansford et al., 2007; Teicher, 2010; Durrant & Ensom, 2012; Initiative, 2015; Kish & Newcombe, 2015).

Rather than making specific judgments about the acceptability of blame or punishment in such a variety of contexts, my point is simply that studies like the ones just mentioned should all be seen as part of a comprehensive science of moral responsibility. Science can provide us with an accurate view of current responsibility practices and guidance about how to avoid problematic aspects. These are both essential tasks for the science of moral responsibility as defined by the Enhancement model.

While in the cases just considered the suggestion is that science can help to improve responsibility practices by finding ways to *prevent* certain responsibility responses, the prac-

tices can also be improved by having other responses *promoted*. This possibility arises from the fact that some factors can influence responsibility responses in an inhibitory way. Some studies about violence against women exemplify this situation.

Since at least the late 1980s, researchers have been aware of a serious problem of sexual violence against women in the university context. In a study in the United States by Koss, Gidycz and Wisniewski (1987), 28% of women in college reported being a victim of a crime satisfying the legal definition of rape since the age of fourteen. The problem was neither restricted to the higher education context, nor to the United States. In a population-based survey in the United States, 18% of women reported having been raped, and 45% reported having been victims of other forms of sexual violence (Black et al., 2011). And the problem has been documented world-wide (WHO, 2013).

From the standpoint of the Enhancement model, it matters to note that proposed strategies to reduce those sorts of violence have included a promotion of responsibility responses. Black et al. (2011, p. 4) say that "[a]n important part of any response […] is to hold perpetrators accountable". Responsibility responses are also promoted within so-called 'bystander approaches' to the prevention of sexual violence:

> Bystander models share a literature that provides guidance on which factors increase the likelihood that a bystander will intervene to prevent violence. Briefly, the objective of bystander intervention is to involve both men and women to change the context or environment that may tacitly support violence against women. (Coker et al., 2011, p. 779)

The Green Dot Intervention Program is one such bystander program that was found effective in promoting bystander interventions (Coker et al., 2011). In assessing the program, Coker et al. used a modified version of the Bystander Behaviors Scale (Banyard, Plante & Moynihan, 2005), in which participants were asked to inform things such as how often they "spoke up if somebody said that someone deserved to be raped or to be hit by their partner" (pp. 785-786) and how often they "spoke up to someone who was bragging or making excuses for forcing someone to have sex with them" (p. 786). These 'speaking-ups' clearly express disapprobation and censure, and as such, they count as expressions of blame. Thus, while in the cases earlier considered responsibility practices could be enhanced by having certain responses prevented, in the case of sexual violence the practices can be enhanced by having some responses promoted.

The Enhancement model, therefore, improves on the Minimal model and the Folk intuitions model by giving the science of moral responsibility a set of questions that is broader

in scope and less tied to the traditional philosophical agenda. As the Folk intuitions model is intended to do, the Enhancement model also keeps a tight connection with real-world practices. But it takes the practices as they currently stand as just a starting point for an investigation whose ultimate goal is their improvement. After all, actual practices can be suboptimal in a number of ways, in which case we would all benefit from acknowledging the imperfections and finding ways to fix them. Hence, according to the Enhancement model science should describe the current state of the practices as well as the available ways for pursuing their improvement.

5 Some related views

The many scientific studies mentioned in the previous section instantiate different aspects of the investigation of moral responsibility described by the Enhancement model. In this respect, the Enhancement model is just a framework that integrates a variety of available studies, making explicit their underlying assumptions and relations. This section describes some views that approximate aspects of the Enhancement model as such, i.e., as an explicit framework for the investigation of moral responsibility that integrates descriptive and evaluative work with the aim of improving responsibility practices.

The approach to moral responsibility originally put forth by P.F. Strawson (1962, p. 152) and further developed by Adina Roskies and Bertram Malle (2013) has a lot in common with the Enhancement model. One of the insights in Strawson's landmark essay is the idea that we should look at how responsibility practices figure in ordinary interpersonal relations in order to establish what the conditions for moral responsibility are.[5] Although Strawson sought to describe the practices based on armchair considerations, Roskies and Malle (2013) updated the method by turning to empirical studies, including an earlier version of the Path model of blame (Guglielmo, Monroe & Malle, 2009; Malle, Guglielmo & Monroe, 2012). In this first aspect, Roskies and Malle's approach is very similar to the first descriptive task assigned to science within the Enhancement model. Roskies and Malle also mention that ordinary responsibility practices could be "measured against deeper values such as fairness and justice (and [that] perhaps some cultures have more just practices of blame and punishment than others)" (2013, p. 147). This is close to the Enhancement model's proposed normative assessment of

5 Although it is open to debate how exactly Strawson's point is to be understood, it seems to involve the assumption that people are responsible because they are *held* responsible in interpersonal relations, and not the other way around—hence the claim that the approach involves some sort of 'reversal' (see Todd, 2016, for an in-depth discussion of the point).

responsibility practices, although Roskies and Malle do not describe the search for improvement that such an assessment can motivate neither the role science can have in it. On the other hand, when briefly discussing certain studies in psychology, Strawson did acknowledge "the possibility and desirability of redirection and modification of our human attitudes in the light of these studies" (1962, p. 170). This largely ignored aspect of Strawson's contribution approximates the Enhancement model's evaluative and descriptive tasks directed toward the improvement of ordinary practices, even though Strawson himself did not distinguish between the normative and scientific elements involved. Finally, Strawson as well as Roskies and Malle were very interested in assessing the relevance of determinism in ordinary practices, and this is a point where both views depart from the Enhancement model. For within the Enhancement model, determinism may become relevant only if it is explanatorily relevant—for example, if beliefs about determinism play a causal role in current responsibility practices—or if determinism is thought to be both true and normatively relevant for an assessment of the practices. If none of these conditions is met, then the topic of determinism is at best secondary within the Enhancement model.

Hagop Sarkissian's (2010) work on the situationist literature also has elements in common with the Enhancement model. As mentioned earlier (section 2), research in social psychology provides evidence that seemingly minor environmental factors can have a major influence on one's subsequent moral behavior. Although the situationist literature may seem troubling, Sarkissian argues that it also opens the possibility that minor changes in our actions and attitudes—such as "choice of words, emotional expressions, mannerisms, tone of voice, posture" (2010. p. 10)—can positively impact the situations we find ourselves in: "By proactively introducing signals that foster an environment amenable to cooperation, one can enhance the probabilities of positive outcomes emerging." (p. 10). This concern for improving the moral quality of human interactions is something the Enhancement model shares with Sarkissian's approach. But there are also differences in scope. First, Sarkissian focuses on improving the outcomes of human interactions in general, while the Enhancement model is particularly concerned with interactions that involve responsibility responses. And, second, Sarkissian is specifically concerned with the impact of "minor tweaks" in our behavior and attitudes, while the Enhancement model leaves open the type of adjustment to be considered.

Finally, it may be instructive to distinguish between the Enhancement model and Manuel Vargas' (2007; 2013) revisionist approach to moral responsibility. Vargas assumes that

most people take responsibility practices to rely on incompatibilist commitments (2013, p. 40) but contends that libertarian theories assuming the existence of indeterministic processes at "particular places and times along the pathway to human decisions" (p. 67) fail to meet a standard of naturalistic plausibility. Vargas thinks this situation justifies not skepticism—the view that no one is morally responsible—but revisionism, a "repair to commonsense thinking, one that strips away the metaphysically demanding elements of libertarianism and preserves the justifiable core of our attitudes and practices." (2007, p. 160). The Enhancement model shares Vargas' idea that responsibility practices can be justified in part because they help to cultivate a valuable form of agency. But the two views have different targets when it comes to their proposed revisions. Vargas' approach is mainly concerned with the revision of ordinary commitments or beliefs (particularly the incompatibilist ones). The Enhancement model, in contrast, focuses on possible revisions to the very responsibility responses that may or may not depend on those beliefs.

6 Conclusion

This paper has considered three models that provide general frameworks for specifying the scientific questions that matter for an investigation of moral responsibility. While the Minimal model only gives science the task of telling whether specific conditions of moral responsibility are satisfied, the Folk intuitions model gives science a role in the very specification of those conditions. Both models, however, are incomplete and leave out further relevant questions about moral responsibility. The Enhancement model gives voice to a group of such additional questions, one that arises out of an interest to understand and improve the actual workings of responsibility responses. Within this framework, science gets the tasks of explaining responsibility responses and describing how normatively problematic aspects of the responses can be effectively changed. The Enhancement model, therefore, gives the science of moral responsibility both purely descriptive tasks and tasks that pertain to a more revisionist enterprise whose ultimate goal is the improvement of responsibility practices.[6]

References

Afifi, T. O.; Brownridge, D. A.; Cox, B. J. and Sareen, J. (2006). "Physical punishment, child abuse and psychiatric disorders", *Child Abuse and Neglect* 30: 1093-1103.

6 I would like to thank Frank Sautter, Adam Bear, and two anonymous reviewers for this journal for their comments on earlier versions of the paper. Thanks also to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support.

Alfano, M. and Loeb, D. (2016). “Experimental moral philosophy”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.

Andow, J. and Cova, F. (2016). “Why compatibilist intuitions are not mistaken: A reply to Feltz and Millan”, *Philosophical Psychology* 29: 550-566.

Aquino, K.; Tripp, T. M. and Bies, R. J. (2001). “How employees respond to personal offense: The effects of blame attribution, victim status, and offender status on revenge and reconciliation in the workplace”, *Journal of Applied Psychology* 86: 52-59.

Asma, L. (forthcoming). “There is no free won’t: The role definitions play”, *Journal of Consciousness Studies* .

Ateah, C. A. and Durrant, J. E. (2005). “Maternal use of physical punishment in response to child misbehavior: Implications for child abuse prevention”, *Child Abuse & Neglect* 29: 169–185.

Balaguer, M. (2009). “Why there are no good arguments for any interesting version of determinism”, *Synthese* 168: 1-21.

Banyard, V. L.; Plante, E. G. and Moynihan, M. M. (2005). *Rape prevention through bystander education: Bringing a broader community perspective to sexual violence prevention*. Washington, DC: Report submitted to the U.S. Department of Justice.

Baumeister, R. F.; Masicampo, E. J. and Vohs, K. D. (2011). “Do conscious thoughts cause behavior?”, *Annual Review of Psychology* 62: 331-361.

Bear, A. and Knobe, J. (2016). “What do people find incompatible with causal determinism?”, *Cognitive Science* 40: 2025-2049.

Black, M. C.; Basile, K. C.; Breiding, M. J.; Smith, S. G.; Walters, M. L.; Merrick, M. T.; Chen, J. and Stevens, M. R. (2011). *The National Intimate Partner and Sexual Violence Survey: 2010 summary report*. Atlanta: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Brinson, J. A.; Kottler, J. A. and Fisher, T. A. (2004). “Cross-cultural conflict resolution in the schools: Some practical intervention strategies for counselors”, *Journal of Counseling & Development* 82: 294-301.

Coker, A. L.; Cook-Craig, P. G.; Williams, C. M.; Fisher, B. S.; Clear, E. R.; Garcia, L. S. and Hegge, L. M. (2011). “Evaluation of Green Dot: An active bystander intervention to reduce sexual violence on college campuses”, *Violence Against Women* 17: 777-796.

Durrant, J. and Ensom, R. (2012). “Physical punishment of children: lessons from 20 years of research”, *Canadian Medical Association Journal* 184: 1373-1377.

Feltz, A. (2015). “Experimental philosophy of actual and counterfactual free will intuitions”, *Consciousness and Cognition* 36: 113-130.

Feltz, A.; Cokely, E. T. and Nelson, B. (2016). “Experimental philosophy needs to matter: Reply to Andow and Cova”, *Philosophical Psychology* 29: 567-569.

Fischborn, M. (2016). “Libet-style experiments, neuroscience, and libertarian free will”, *Philosophical Psychology* 29: 494-502.

Fischborn, M. (2017). “Neuroscience and the possibility of locally determined choices: Reply to Adina Roskies and Eddy Nahmias”, *Philosophical Psychology* 30: 198-201.

Fischer, J. M.; Kane, R.; Pereboom, D. & Vargas, M. (Ed.) (2007). *Four views on free will*. Malden: Wiley-Blackwell.

Fischer, J. M. and Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Gomes, G. (2007). “Free will, the self, and the brain”, *Behavioral Sciences and the Law* 25: 221-234.

Guglielmo, S.; Monroe, A. E. and Malle, B. F. (2009). “At the heart of morality lies folk psychology”, *Inquiry* 52: 449-466.

Haynes, J.-D. (2011). “Beyond Libet: Long-term prediction of free choices from neuroimaging signals”. In: Sinnott-Armstrong, W. & Nadel, L. (Ed.), *Conscious will and responsibility: A tribute to Benjamin Libet*, pp. 85-96. Oxford: Oxford University Press.

Hoefer, C. (2016). “Causal determinism”. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.

Initiative (2015). “Global initiative to end all corporal punishment of children”, http://www.endcorporalpunishment.org/.

Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.

Kish, A. M. and Newcombe, P. A. (2015). ““Smacking never hurt me!” Identifying myths surrounding the use of corporal punishment”, *Personality and Individual Differences* 87: 121-129.

Koss, M. P.; Gidycz, C. A. and Wisniewski, N. (1987). “The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of higher education students”, *Journal of Counseling and Clinical Psychology* 55: 162-170.

Kubany, E. S.; Bauer, C. B.; Muraoka, M. Y.; Richard, D. C. and Read, P. (1995). “Impact of labeled anger and blame in intimate relationships”, *Journal of Social and Clinical Psychology* 14: 53-60.

Lansford, J. E.; Miller-Johnson, S.; Berlin, L. J.; Dodge, K. A.; Bates, J. E. and Pettit, G. S. (2007). “Early physical abuse and later violent delinquency: A prospective longitudinal study”, *Child Maltreatment* 12: 233-245.

Libet, B.; Gleason, C. A.; Wright, E. W. and Pearl, D. K. (1983). “Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act”, *Brain* 106: 623-642.

Madden, M. E. and Janoff-Bulman, R. (1981). “Blame, control, and marital satisfaction: Wives' attributions for conflict in marriage”, *Journal of Marriage and Family* 43: 663-674.

Malle, B. F.; Guglielmo, S. and Monroe, A. E. (2012). “Moral, cognitive, and social: The nature of blame”. In: Forgas, J.; Fiedler, K. & Sedikides, C. (Ed.), *Social thinking and interpersonal behaviour*, pp. 313-331. New York: Psychology Press.

Malle, B. F.; Guglielmo, S. and Monroe, A. E. (2014). “A theory of blame”, *Psychological Inquiry* 25: 147-186.

Marques, B. S. (2015). “Different kinds of decisions and an experiment on unconscious generation of free decisions: A conceptual analysis”, *Filosofia Unisinos* 16: 44-57.

McCullough, M. E.; Kimeldorf, M. B. and Cohen, A. D. (2008). “An adaptation for altruism? The social causes, social effects, and social evolution of gratitude ”, *Current Directions in Psychological Science* 17: 281-285.

Mele, A. (2001). “Acting intentionally: Probing folk notions”. In: Malle, B.; Moses, L. & Baldwin, D. (Ed.), *Intentions and intentionality: Foundations of social cognition*, pp. 27-43. Cambridge, MA: MIT Press.

Mele, A. (2009). *Effective intentions: The power of conscious will*. Oxford: Oxford University Press.

Misirlisoy, E. and Haggard, P. (2014). “A neuroscientific account of the human will”. In: Sinnott-Armstrong, W. (Ed.), *Moral psychology: Free will and moral responsibility*, pp. 37-42. Cambridge, MA: MIT Press.

Monroe, A. E.; Brady, G. and Malle, B. F. (2017). “This isn’t the free will worth looking for: General free will beliefs do not influence moral judgments; agent-specific choice ascriptions do”, *Social Psychological and Personality Science* 8: 191-199.

Murray, D. and Nahmias, E. (2014). “Explaining away incompatibilist intuitions”, *Philosophy and Phenomenological Research* 88: 434-467.

Nadelhoffer, T.; Shepard, J.; Nahmias, E.; Sripada, C. and Ross, L. T. (2014). “The free will inventory: Measuring beliefs about agency and responsibility”, *Consciousness and Cognition* 25: 27-41.

Nahmias, E. (2002). “When consciousness matters: A critical review of Daniel Wegner's The illusion of conscious will”, *Philosophical Psychology* 15: 527-541.

Nahmias, E. (2014). “Is free will an illusion? Confronting challenges from the modern mind sciences”. In: Sinnott-Armstrong, W. (Ed.), *Moral psychology: Free will and moral responsibility*, pp. 1-25. Cambridge, MA: MIT Press.

Nahmias, E.; Coates, D. J. and Kvaran, T. (2007). “Free will, moral responsibility, and mechanism: Experiments on folk intuitions”, *Midwest Studies in Philosophy* 31: 214-242.

Nahmias, E.; Morris, S.; Nadelhoffer, T. and Turner, J. (2005). “Surveying freedom: Folk Intuitions about free will and moral responsibility”, *Philosophical Psychology* 18: 561-584.

Nelkin, D. K. (2005). “Freedom, responsibility and the challenge of situationism”, *Midwest Studies in Philosophy* 29: 181-206.

Nichols, S. and Knobe, J. (2007). “Moral responsibility and determinism: The cognitive science of folk intuitions”, *Noûs* 41: 663-685.

Pickard, H. (2011). “Responsibility without blame: Empathy and the effective treatment of personality disorder”, *Philosophy, Psychiatry, Psychology* 18: 209-223.

Pickard, H. (2013). “Responsibility without blame: Philosophical reflections on clinical practice”. In: Fulford, K. W. M.; Davis, M.; Gipps, R. G. T.; Graham, G.; Sadler, J. Z.; Stanghellini, G. & Thorton, T. (Ed.), *The Oxford handbook of philosophy and psychiatry*, pp. 1134-1154. Oxford: Oxford University Press.

Pockett, S.; Banks, W. & Gallagher, S. (Ed.) (2006). *Does consciousness cause behavior?*. Cambridge, MA: MIT Press.

Rose, D. and Nichols, S. (2013). “The lesson of bypassing”, *Review of Philosophy and Psychology* 4: 599-619.

Rosen, G. (2015). “The alethic conception of moral responsibility”. In: Clarke, R.; McKenna, M. & Smith, A. M. (Ed.), *The nature of moral responsibility: New essays*, pp. 65-87. New York: Oxford University Press.

Roskies, A. (2006). “Neuroscientific challenges to free will and responsibility”, *Trends in Cognitive Science* 10: 419-423.

Roskies, A. and Nahmias, E. (2017). ““Local determination”, even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn”, *Philosophical Psychology* 30: 185-197.

Roskies, A. L. (2011). “Why Libet's studies don't pose a threat to free will”. In: Sinnott-Armstrong, W. & Nadel, L. (Ed.), *Conscious will and responsibility: A tribute to Benjamin Libet*, pp. 11-22. Oxford: Oxford University Press.

Roskies, A. L. (2014). “Can neuroscience resolve issues about free will?”. In: Sinnott-Armstrong, W. (Ed.), *Moral psychology: Free will and moral responsibility*, pp. 103-126. Cambridge, MA: MIT Press.

Roskies, A. L. and Malle, B. F. (2013). “A Strawsonian look at desert”, *Philosophical Explorations* 16: 133-152.

Roskies, A. L. and Nichols, S. (2008). “Bringing moral responsibility down to earth”, *The Journal of Philosophy* 105: 371-388.

Sarkissian, H. (2010). “Minor tweaks, major payoffs: The problems and promise of situationism in moral philosophy”, *Philosophers' Imprint* 10: 1-15.

Schlosser, M. E. (2012). “Free will and the unconscious precursors of choice”, *Philosophical Psychology* 25: 365-384.

Schlosser, M. E. (2013). “Conscious will, reason-responsiveness, and moral responsibility”, *The Journal of Ethics* 17: 205-232.

Schooler, J.; Nadelhoffer, T.; Nahmias, E. and Vohs, K. D. (2015). “Measuring and manipulating beliefs and behaviors associated with free will: The good, the bad, and the ugly”. In: Mele, A. R. (Ed.), *Surrounding free will: Philosophy, Psychology, Neuroscience*, pp. 72-94. New York: Oxford University Press.

Shepherd, J. (2015). “Scientific challenges to free will and moral responsibility”, *Philosophy Compass* 10: 197-207.

Soon, C. S.; Brass, M.; Heinze, H.-J. and Haynes, J.-D. (2008). “Unconscious determinants of free decisions in the human brain”, *Nature Neuroscience* 11: 543-545.

Spohn, C. and Holleran, D. (2000). “The imprisonment penalty paid by the young, unemployed black and hispanic male offenders”, *Criminology* 38: 281-306.

Steffensmeier, D.; Ulmer, J. and Kramer, J. (1998). “The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male”, *Criminology* 36: 763-798.

Strawson, P. F. (1962). “Freedom and resentment”. In: Pereboom, D. (Ed.), *Free will*, pp. 148-171. Indianapolis: Hackett.

Suppes, P. (1993). “The transcendental character of determinism”, *Midwest Studies in Philosophy* 18: 242-257.

Teicher, M. H. (2010). “Commentary: Childhood abuse: New insights into its association with posttraumatic stress, suicidal ideation, and aggression”, *Journal of Pediatric Psychology* 35: 578-580.

Todd, P. (2016). “Strawson, moral responsibility, and the “order of explanation”: An intervention”, *Ethics* 127: 208-240.

Vargas, M. (2007). “Revisionism”. In: Fischer, J. M.; Kane, R.; Pereboom, D. & Vargas, M. (Ed.), *Four views on free will*, pp. 126-165. Malden: Wiley-Blackwell.

Vargas, M. (2013b). *Building better beings: A theory of moral responsibility*. Oxford: Oxford University Press.

Vargas, M. (2013a). “Situationism and moral responsibility: Free will in fragments”. In: Clark, A.; Kiverstein, J. & Vierkant, T. (Ed.), *Decomposing the will*, pp. 325-349. Oxford: Oxford University Press.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Werndl, C. (2013). “On choosing between deterministic and indeterministic models: Underdetermination and indirect evidence”, *Synthese* 190: 2243-2265.

Werndl, C. (2016). “Determinism and indeterminism”. In: Humphreys, P. (Ed.), *The Oxford handbook of philosophy of science*, pp. 210-232. New York: Oxford University Press.

WHO (2013). *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. World Health Organization.

Zimmerman, M. (2015). “Varieties of moral responsibility”. In: Clarke, R.; McKenna, M. & Smith, A. M. (Ed.), *The nature of moral responsibility: New essays*, pp. 45-64. New York: Oxford University Press.

DISCUSSÃO

Afirmei na Introdução que a tese unificadora deste trabalho diz que a discussão sobre a existência da responsabilidade moral é insuficiente para promover o aprimoramento das práticas cotidianas de responsabilidade. Os artigos que antecederam esta seção fornecem os elementos principais para um argumento em favor dessa tese. No que segue, busco dar forma a esse argumento. Considero, inicialmente, a proposta de aprimoramento das práticas de responsabilidade moral embutida nas propostas céticas de Pereboom e Caruso, e discuto suas limitações. Em seguida, considero as limitações da discussão sobre a existência da responsabilidade moral em geral relativamente à pretensão de se aprimorar as práticas cotidianas de responsabilidade. Ao final, mostro como a investigação interdisciplinar sugerida pelo modelo de aprimoramento supera essas limitações.

O interesse no valor das práticas cotidianas de responsabilidade presente na literatura sobre a existência da responsabilidade moral é particularmente bem ilustrado, ainda que de maneira radical, nas propostas céticas de Pereboom e Caruso (ver o artigo 3). Por defender que ninguém jamais é moralmente responsável no sentido de merecimento básico, o ceticismo sobre a responsabilidade moral implica que as práticas de responsabilidade cotidianas em seu todo apresentam um defeito valorativo. Em outras palavras, se a qualidade valorativa das práticas cotidianas de responsabilidade (incluindo aí as práticas instituídas nos diversos sistemas penais) depende de um pressuposto que não é satisfeito, então parece haver, no mínimo, um toque de injustiça ou imoralidade envolvido em seu funcionamento. Pereboom e Caruso não são tão radicais a ponto de defenderem a eliminação pura e simples das práticas de responsabilidade. Mas eles se dedicaram a propor alternativas viáveis. No artigo 3, distingui dois componentes dessas propostas. Por um lado, há uma proposta que aponta na direção de se reduzir a severidade (ou o grau de dano) envolvido nas punições que são atribuídas àqueles que realizam algum crime. Essa parte da proposta deriva propriamente do ceticismo sobre a responsabilidade moral. Por outro lado, ambos abrem espaço para versões (análogas) da punição que são justificadas em termos outros que o merecimento e a retribuição. Pereboom falou originalmente em formas de detenção (menos danosas que as formas de detenção atuais) que são justificadas com base no direito de autodefesa e por analogia à quarentena para indivíduos com doenças contagiosas (sob certas condições). Caruso desenvolveu a proposta de Pereboom buscando justificar maneiras de lidar com a criminalidade por analogia a modelos de ética da

saúde pública. Sua proposta enfatiza a prevenção da criminalidade—a partir do tratamento de suas causas biológicas, sociais, psicológicas etc.—de maneira que a detenção baseada no direito de autodefesa tenha seu uso reduzido ao máximo.

Os artigos 1–4 desta tese fornecem elementos para mostrar a insuficiência da discussão da existência da responsabilidade moral para promover o aprimoramento das práticas de responsabilidade no caso particular do ceticismo. Em primeiro lugar, a tese de que não há livre-arbítrio e responsabilidade moral é teoricamente questionável. No artigo 1, rejeitei uma das maneiras de se defender a inexistência do livre-arbítrio. Segundo alguns autores, principalmente neurocientistas, resultados de experimentos como os de Libet mostrariam que nossas escolhas são determinadas e que, por isso, não somos livres.[1] Argumentei que os resultados desses experimentos não são suficientes para mostrar que quaisquer escolhas sejam *determinadas*—em um sentido que seria relevante para posições que tomam o livre-arbítrio como incompatível com o determinismo.[2] Esse argumento se soma a uma extensa literatura que rejeita que a inexistência do livre-arbítrio tenha sido estabelecida com base em estudos em neurociência e psicologia ou em algum outro tipo de argumento (ver, por exemplo, o artigo 1, n. 1 e o artigo 3, n. 2). Nesse contexto, é razoável concluir que o ceticismo sobre a responsabilidade moral é no mínimo altamente controverso e que, por isso, é frágil a justificação de sua proposta de modificação das práticas de responsabilidade.

Um problema adicional para a proposta cética de alteração das práticas de responsabilidade diz respeito à implausibilidade psicológica de sua adoção pelas pessoas em geral. No artigo 3, argumentei que estudos empíricos sobre o funcionamento da crença no livre-arbítrio e do desejo de que criminosos sejam punidos tornam implausível a implementação de pelo menos parte da proposta cética. A parte em questão diz respeito àquilo que o ceticismo sobre a responsabilidade estritamente *exige*—i.e., uma preocupação em reduzir tanto quanto possível a severidade da punição. Essa parte da proposta deve ser distinguida de outros componentes que os céticos poderiam simplesmente acrescentar à sua proposta, mas por razões independentes do ceticismo—por exemplo, detenção com base no direito de autodefesa ou medidas pre-

1 Enfatizo que esse argumento depende de se aceitar que o livre-arbítrio (e, por isso, também a responsabilidade moral) são incompatíveis com a tese do determinismo, e que os defensores do compatibilismo rejeitam essa tese (para exemplos, ver a Introdução desta tese).

2 Apesar de negar que os estudos em questão tenham *mostrado* que algumas de nossas escolhas sejam determinadas pela atividade neural, deixei em aberto que outros estudos em neurociência ou psicologia *possam mostrar* que haja determinação no sentido discutido. Embora concordem com a primeira tese, Roskies e Nahmias (2017) criticaram meus argumentos a favor da segunda tese. O artigo 2 é minha resposta a essas críticas.

ventivas. A literatura da psicologia social revisada no artigo 3 sugere que a crença no livre-arbítrio e na adequação de se culpar ou punir o autor de um crime é forte e robusta, e que se reforça diante da ocorrência do comportamento imoral ou criminoso. Por essa razão, é implausível que alguma sociedade venha a reduzir a severidade das formas de punição que adota em virtude de uma redução generalizada da crença na responsabilidade moral e no livre-arbítrio. No entanto, sugeri também que a aspiração do cético a punições menos severas poderia vir a ser alcançada por uma via alternativa. Em primeiro lugar, o ponto de partida da mudança deveria focar no desejo natural de punir, e não na crença no livre-arbítrio. Em segundo lugar, seriam necessárias estratégias de redução da criminalidade que sejam incompatíveis com a adoção simultânea de penas severas.[3] Independentemente dessa alternativa, a observação relevante neste momento é simplesmente que a proposta de modificação das práticas de responsabilidade (neste caso, as práticas de punição) que estritamente se segue do ceticismo sobre a responsabilidade moral é psicologicamente implausível, pelo menos na medida em que sua realização depende da aceitação coletiva do ceticismo.

O artigo 4, por fim, sugere um terceiro problema para a alternativa cética. Os resultados experimentais apresentados são consistentes com a tese de que, além de psicologicamente implausível, a rota cética para reduzir a severidade da punição é psicologicamente desnecessária. No experimento realizado, a quantidade de punição recomendada para o autor de um crime fictício variou em função da disponibilidade e eficácia de medidas alternativas visando a prevenção de sua reincidência no crime. A pena média sugerida foi de 3,25 anos de prisão na condição em que a medida alternativa era muito mais eficaz que a punição, comparada a 5,63 anos na condição em que a medida alternativa foi dita ter eficácia preventiva equivalente à da punição. Ambos os resultados foram significativamente menores do que a punição média de 8,12 anos de prisão recomendada na condição de controle em que não houve menção à possibilidade de medidas alternativas. Um segundo resultado igualmente importante foi que a essas diferenças na quantidade de punição recomendada não correspondeu qualquer diferença significativa nos níveis de crença na existência do livre-arbítrio em geral, e nem quanto ao livre-arbítrio e culpa atribuídos especificamente ao criminoso fictício considerado. Embora seja preciso cautela na interpretação desse resultado nulo devido ao tamanho da amostra utilizada no experimento (ver o artigo 4 para mais detalhes), o resultado é compatível com a hipótese de que o apoio à redução da punição exigida pelo ceticismo poderia ser alcançado mais facil-

3 Noto que os resultados apresentados no artigo 4, discutidos a seguir, indicam que essa afirmação precisa ser qualificada.

mente por vias que não exijam a aceitação do ceticismo. Em outras palavras, além de implausível que as pessoas possam apoiar a redução da punição por desacreditarem no livre-arbítrio, também é desnecessário que assim o façam, já que poderiam apoiar essa redução por outras razões que são compatíveis com a manutenção da crença no livre-arbítrio, tais como a crença na eficácia de medidas alternativas.

O resultado geral da avaliação da rota para redução da severidade da punição incluída na visão cética é que, além de teoricamente questionável, é psicologicamente implausível e desnecessária. Mas não gostaria de desvalorizar, com essa crítica, as contribuições que os céticos sobre a existência do livre-arbítrio têm oferecido ao tema do aprimoramento das práticas de responsabilidade. A crítica apresentada ajuda a trazer à tona um dado importante sobre o que o aprimoramento das práticas de responsabilidade exige. Trata-se justamente do fato de que essa busca de aprimoramento depende de considerações que vão além da discussão estritamente focada na existência da liberdade e da responsabilidade moral. Desenvolvo esta observação a seguir.

A sugestão de que as práticas de responsabilidade cotidianas possam ser aprimoradas envolve pelo menos duas suposições centrais. Uma delas é que se pode *avaliar* o valor dessas práticas e, como resultado dessa avaliação, constatar que sua qualidade esteja aquém de algum padrão valorativo concebível. A segunda suposição é que, pelo menos em princípio, pode-se implementar modificações das práticas de responsabilidade que visem *aumentar* o seu valor ou aproximá-las de algum padrão valorativo mais elevado. Para mostrar que a discussão sobre a existência da responsabilidade moral e do livre-arbítrio é insuficiente para o propósito de aprimorar as práticas de responsabilidade cotidianas, considerarei suas limitações no âmbito de cada uma dessas suposições.

Vejamos, primeiramente, o tipo de consideração valorativa sobre as práticas de responsabilidade que a discussão sobre a existência do livre-arbítrio, especificamente, pode gerar. Os dois resultados principais a que essa discussão pode levar são que o livre-arbítrio existe ou que não existe (em geral ou sob algumas condições). O fato de que o livre-arbítrio é uma condição necessária, mas não suficiente, para a responsabilidade moral implica que, se o livre-arbítrio não existe, então tampouco existe a responsabilidade moral. A consideração avaliativa que resulta nesse caso aponta na direção de que as práticas de responsabilidade sofrem de algum nível de inadequação. Digo “aponta” (e não algo mais forte) porque, como vimos no caso das teorias céticas de Pereboom e Caruso, não se precisa entender a inexistência da responsa-

bilidade moral como uma razão definitiva para se abandonar as práticas de responsabilidade em seu todo. Em particular, é em princípio possível que razões valorativas independentes do livre-arbítrio ou da responsabilidade superem as considerações derivadas da inexistência do livre-arbítrio e resultem, no fim das contas, numa avaliação em que as práticas de responsabilidade (ou alguma versão modificada delas) sejam consideradas aceitáveis ou mesmo recomendadas. Já no caso em que se afirma a existência do livre-arbítrio, a consequência é que a responsabilidade moral *pode* existir, uma vez que o livre-arbítrio é apenas uma das condições necessárias para a responsabilidade moral. Diferentemente do caso cético, nenhuma avaliação pode ainda resultar, nem mesmo parcial, sem que o estado de outras condições ainda pertencentes ao âmbito da responsabilidade moral tenha sido estabelecido.

Vejamos agora o tipo de consideração valorativa que a discussão sobre a existência da responsabilidade moral (da qual a discussão sobre a existência do livre-arbítrio é uma parte) pode gerar. Aqui há espaço para variação, dependendo de como se entende a noção de responsabilidade moral. Concentrarei minha discussão, novamente, no entendimento de que ser moralmente responsável é merecer, no sentido básico, alguma resposta característica da responsabilidade. Consideremos o que acontece quando se conclui que há agentes moralmente responsáveis nesse sentido. O que essa conclusão implica a respeito do valor das práticas de responsabilidade? Se o agente é moralmente responsável, segue-se que merece alguma resposta. Mas podemos perguntar, se queremos avaliar as práticas de responsabilidade, se esse merecimento implica ou não que o agente *deva* receber a resposta merecida. Parece-me razoavelmente evidente que merecer certa resposta não implica que essa resposta deva, em definitivo, ser recebida. No artigo 5, mencionei alguns estudos que sugerem que a atribuição de culpa em certos contextos pode ter consequências negativas para a qualidade de relações, incluindo relações entre colegas de escola ou trabalho e relações conjugais. Podemos levar esse tipo de fato em conta ao buscar decidir se se deve ou não atribuir culpa nesses contextos em alguma ocasião, mesmo se acreditarmos que o alvo potencial da culpa foi moralmente responsável pela ação em questão. E, se podemos levar essas considerações adicionais em conta, então mesmo nos casos em que concluímos que a culpa merecida deva ser atribuída, não parece que o juízo de dever siga-se apenas da avaliação do merecimento. Se for assim, então o fato de alguém ser moralmente responsável em certa ocasião não é suficiente para se chegar a uma avaliação final sobre se uma resposta merecida deve ou não ser dada. No caso do cético, por sua vez, é ainda mais evidente que precise recorrer a considerações além do merecimento em sua avalia-

ção das práticas de responsabilidade. Diferenciei acima aquilo que o ceticismo sobre a responsabilidade moral implica (preocupação com a redução da punição) daquilo que não implica mas pode aceitar por outras razões (por exemplo, medidas de prevenção do crime). No caso das considerações do segundo tipo, os céticos levam em conta uma gama variada de considerações valorativas, incluindo a segurança a nível social, direitos das vítimas, o bem-estar daqueles que cometem crimes mas não merecem nenhum tipo de punição, entre outros. Isso mostra que, embora de maneiras diferentes, nem a atribuição e nem a negação de merecimento implica que alguma resposta característica da responsabilidade deva ou não ser dada. Em ambos os casos, considerações valorativas além do merecimento podem ser levadas em conta. E, portanto, a discussão sobre a existência da responsabilidade é, por si só, insuficiente para se avaliar se aspectos das práticas de responsabilidade devem ou não ser mantidos ou modificados.[4]

A discussão sobre a existência da responsabilidade é insuficiente para se avaliar as práticas de responsabilidade também em um segundo aspecto. Avaliar as práticas pressupõe não apenas que se tenha padrões valorativos a serem levados em conta, mas também que se tenha uma descrição adequada das práticas a serem avaliadas. E a discussão sobre a existência da responsabilidade moral, assim como não gera toda a gama de considerações valorativas relevantes para a avaliação das práticas cotidianas, tampouco gera uma descrição dessas práticas que permita sua submissão a tal avaliação. Consideremos novamente a sugestão de que os efeitos da atribuição de culpa em determinadas circunstâncias possam ser relevantes para se avaliar se devemos ou não culpar alguém. Só podemos avaliar se esses efeitos são desejáveis ou indesejáveis—e portanto se as práticas de atribuição de culpa deveriam ser mantidas ou modificadas—se tivermos uma descrição suficientemente precisa e detalhada de quais são esses efeitos, qual sua magnitude, se são evitáveis e assim por diante. Oferecer esse tipo de descrição está além do escopo da discussão sobre a existência da responsabilidade moral. E, assim, temos aqui mais uma limitação dessa discussão no que diz respeito à tarefa de avaliar a qualidade de nossas práticas cotidianas de responsabilidade.

Afirmei anteriormente que a possibilidade de que as práticas cotidianas de responsabilidade sejam aprimoradas pressupõe que essas práticas possam ser avaliadas e que possam ser modificadas de modo a terem seu valor aumentado. Os parágrafos anteriores mostram as limitações da discussão sobre a existência da responsabilidade moral no que diz respeito à avalia-

4 Os resultados do artigo 4 também sugerem que as pessoas de fato levam em conta considerações além do merecimento, como as consequências da punição e alternativas.

ção das práticas de responsabilidade. Minhas considerações sobre a viabilidade de se implementar as mudanças na lei exigidas pelo ceticismo sobre a responsabilidade moral no artigo 3 sugere uma limitação adicional no que diz respeito à implementação de mudanças de acordo com as avaliações. O artigo 3 considera dificuldades prováveis para se reduzir a severidade da punição legalmente autorizada e recomendada. Essas dificuldades dizem respeito à oposição provável oriunda das crenças e atitudes comuns entre a maioria da população (crença no livre-arbítrio e desejo de punir) bem como aos efeitos possíveis da modificação considerada (por exemplo, aumento da criminalidade). Levar em conta esse tipo de dificuldade reflete o fato mais geral de que qualquer intervenção a nível social depende, para seu sucesso, de se conhecer os efeitos esperados, avaliar seus possíveis riscos e benefícios, entre outras coisas. Essas considerações, uma vez mais, não fazem parte da discussão sobre a existência da responsabilidade moral, que novamente precisa ser suplementada se o objetivo for a promoção do valor das práticas cotidianas de responsabilidade.

Os parágrafos precedentes enumeraram limitações no potencial da discussão que versa exclusivamente sobre a existência da responsabilidade moral de promover o aprimoramento das práticas cotidianas de responsabilidade. Reconhecer essas limitações prepara o caminho para mostrar como o modelo de aprimoramento, proposto no artigo 5, avança nesse âmbito. O modelo de aprimoramento entende a ciência da responsabilidade moral como parte de uma investigação interdisciplinar que busca tanto entender o funcionamento de nossas práticas cotidianas de responsabilidade quanto apontar maneiras em que essas práticas poderiam ser aprimoradas. Nessa investigação, a ciência—incluindo, em especial, setores da psicologia, da neurociência, da criminologia, entre outras—recebe questões que pertencem a dois grandes grupos. No primeiro grupo estão questões sobre as causas e efeitos das práticas cotidianas de responsabilidade em seus vários aspectos. Pode-se perguntar aqui pelas causas e efeitos da atribuição de culpa, do elogio, da punição e assim por diante. Essa é obviamente uma apresentação bastante geral da proposta, já que as questões podem ser refinadas tanto quanto for conveniente para a investigação. No quinto artigo, ofereci vários exemplos de estudos já realizados que permitem responder a questões desse tipo, incluindo o modelo sobre as causas da atribuição de culpa proposto por Malle, Guglielmo e Monroe (2014) e alguns estudos sobre os efeitos negativos da atribuição de culpa no contexto familiar ou escolar (Madden & Janoff-Bulman, 1981; Aquino, Tripp & Bies, 2001) e da punição física de crianças (Gershoff, 2002; Ateah & Durrant, 2005; Afifi et al., 2006). Ao apontarem as causas e efeitos de aspectos das práti-

cas cotidianas de responsabilidade, esses estudos exemplificam o tipo de investigação associado ao primeiro grupo de questões científicas proposto no modelo de aprimoramento.

O segundo grupo de questões científicas que o modelo de aprimoramento propõe é derivado de uma investigação normativa sobre as práticas cotidianas de responsabilidade. Essa investigação normativa—entendida como tarefa de disciplinas como ética normativa, filosofia política e social, direito, entre outras—parte da suposição de que, em princípio, qualquer aspecto do funcionamento das práticas de responsabilidade pode ser considerado problemático.[5] Um exemplo claro desse tipo de avaliação é opinião, adotada por um número crescente de países desde a década de 1980, de que a punição física de crianças deveria, em função dos efeitos descritos anteriormente, ser abolida (Durrant & Ensom, 2012; Initiative, 2015). No artigo 5, dei vários outros exemplos de casos em que se pode suspeitar de que as práticas de responsabilidade funcionem de maneira menos que ideal. O segundo grupo de questões que a ciência recebe no modelo de aprimoramento diz respeito a como implementar mudanças nas práticas cotidianas de responsabilidade que sejam capazes de atenuar ou eliminar seus aspectos valorativamente problemáticos. Para continuar com o exemplo da punição física de crianças, alguns pesquisadores têm buscado conhecer os fatores que são preditivos do seu uso, e que tipo de intervenções podem contribuir para sua superação (ver, por exemplo, Afifi & Romano, 2017).[6] Esse segundo passo da investigação científica proposta pelo modelo de aprimoramento, portanto, busca encontrar maneiras eficazes de se implementar as sugestões sobre como melhorar as práticas cotidianas de responsabilidade oriundas da avaliação normativa dessas práticas.

Os traços do modelo de aprimoramento recém apresentados permitem ver como ele supera as limitações anteriormente apontadas na discussão sobre a existência da responsabilidade moral. A primeira limitação deriva-se do fato de que as discussões sobre a existência do livre-arbítrio e da responsabilidade moral focam-se exclusivamente nas condições para que as respostas características da responsabilidade sejam ou não merecidas, deixando de lado outras

5 Ao distinguir esses dois grupos de disciplinas, não pretendo sugerir, por um lado, que a avaliação de aspectos de nossas práticas de responsabilidade cotidianas não dependa também de sua descrição ou, por outro, que a investigação científica que busca descrever o funcionamento dessas práticas não seja permeada ela mesma por valores diversos. Notei no quinto artigo que essas duas atividades estão frequentemente misturadas na prática investigativa. Nesse sentido, o modelo de aprimoramento pode ser visto como uma proposta específica sobre como estabelecer essas relações.

6 Para os casos possíveis em que as práticas de responsabilidade cotidianas sejam problemáticas por envolverem punição excessivamente severa, a investigação conduzida nos artigos 3 e 4 também exemplifica a investigação científica de questões do segundo grupo. Noto também que—como exemplifica a discussão sobre violência contra mulheres no artigo 5—o modelo de aprimoramento é compatível com a sugestão de que as práticas cotidianas de responsabilidade possam ser aprimoradas através da promoção de respostas como a atribuição de culpa e punição.

considerações valorativas que também podem ser relevantes para sua avaliação. A investigação normativa proposta no modelo de aprimoramento permite superar essa limitação por deixar em aberto que considerações valorativas de todas as ordens sejam levadas em conta, incluindo o impacto das práticas de responsabilidade sobre as pessoas envolvidas, sobre as relações que essas pessoas mantêm e assim por diante. Uma segunda limitação apontada, também sobre a avaliação das práticas cotidianas, foi que a investigação sobre a existência da responsabilidade moral não inclui uma descrição do funcionamento das práticas de responsabilidade que apresente todos os aspectos que possam ser relevantes para sua avaliação. O modelo de aprimoramento supera essa limitação por ter como ponto de partida justamente uma descrição integral do funcionamento das práticas de responsabilidade tal como se encontram em nossas vidas cotidianas. Finalmente, a terceira limitação atribuída à discussão da existência da responsabilidade moral foi que não inclui uma preocupação sobre a viabilidade de se implementar modificações que visem ao aprimoramento das práticas cotidianas. O modelo de aprimoramento contorna essa limitação em seu segundo grupo de questões científicas, que tem justamente a tarefa de apontar maneiras eficazes de se implementar as modificações das práticas cotidianas de responsabilidade sugeridas a partir da investigação normativa.

Entendo o modelo de aprimoramento e as demais propostas desenvolvidas ao longo desta tese como contribuições à vasta literatura sobre o livre-arbítrio, a responsabilidade moral e a interação entre filosofia e ciência na investigação desses temas. Preciso notar, no entanto, que essas contribuições deixam várias questões em aberto e que também cobrem apenas uma ínfima porção de tudo o que já escreveu sobre esses temas. Para dar apenas alguns exemplos, este trabalho se concentrou em desenvolvimentos recentes na discussão sobre o livre-arbítrio e a responsabilidade moral e, ao fazê-lo, precisou deixar de fora a rica história dessa discussão e as contribuições dos grandes clássicos da filosofia. Ademais, minha consideração da literatura recente também é seletiva. Duas questões especialmente relevantes que ficam em aberto dizem respeito ao papel da filosofia na investigação proposta no modelo de aprimoramento e sua relação com a investigação filosófica mais tradicional a respeito da existência da responsabilidade moral. Gostaria de finalizar esta seção de discussão comentando brevemente sobre essas questões.

Sugeri, no quinto artigo, que a avaliação normativa das práticas cotidianas de responsabilidade é tarefa para um grupo de disciplinas que poderia incluir, sem a elas se restringir, a ética normativa, a filosofia política e o direito. Detalhar o que esse tipo de avaliação envolve é

uma das tarefas que essa tese deixa por fazer, mas alguns elementos podem antecipados neste momento. Ao envolver disciplinas comumente ditas normativas, pode-se falar do modelo de aprimoramento não apenas como um modelo para a *ciência* da responsabilidade moral, mas também como um modelo para uma *investigação interdisciplinar* que envolve tarefas tanto descritivas quanto normativas. Costuma-se distinguir a ética normativa teórica da ética normativa aplicada (Darwall, 2003; Copp, 2006). Segundo essa distinção, a ética normativa teórica encarrega-se da investigação mais abstrata de princípios que possam estar na base de nossas avaliações morais, enquanto a ética normativa aplicada busca estabelecer e justificar avaliações éticas sobre casos particulares, como a permissibilidade do aborto, da eutanásia ou da pena de morte (Darwall, 2003). Levando em conta essa distinção, pode-se ver que a investigação normativa proposta no modelo de aprimoramento apresenta bastante afinidade com o tipo de investigação em ética aplicada. Assim, pode-se dizer que uma das contribuições que a filosofia tem a dar à investigação que segue o modelo de aprimoramento tem a forma de uma ética aplicada às práticas cotidianas de responsabilidade. Um resultado dessa sugestão é que a investigação filosófica da responsabilidade moral—que é variavelmente situada no interior da metafísica, da filosofia da ação, da metaética ou mesmo da filosofia da neurociência ou ciências cognitivas—também faz parte da ética aplicada.

Finalmente, pode-se perguntar sobre como o modelo de aprimoramento se relaciona com a discussão mais tradicional sobre a existência da responsabilidade moral e do livre-arbítrio. Seguindo o que foi dito no parágrafo anterior, a relevância da discussão tradicional reside primariamente nas considerações que gera para a avaliação das práticas cotidianas de responsabilidade. Essas considerações, como disse anteriormente, dizem respeito à presença ou ausência de merecimento nos vários aspectos das práticas cotidianas de responsabilidade. Mas é importante enfatizar que considerações sobre merecimento são apenas um dos tipos de consideração que se pode levar em conta na avaliação das práticas de responsabilidade proposta pelo modelo de aprimoramento. Consequentemente, os resultados da investigação sobre a existência da responsabilidade moral poderão ter maior ou menor peso na avaliação final de algum aspecto específico das práticas de responsabilidade dependendo do contexto. Nada impede, por exemplo, que considerações sobre merecimento se sobreponham a outras considerações em alguns casos, mas que sejam sobrepostas por outras considerações em outros. Vale a pena destacar também que, embora partes desta tese sejam críticas do ceticismo generalizado sobre o livre-arbítrio ou a responsabilidade moral, o modelo de aprimoramento por si só é

compatível tanto com a afirmação quanto com a negação da existência da responsabilidade moral em absoluto ou em casos específicos.

CONCLUSÃO

Se minha tarefa ao escrever esta tese fosse vista como a tarefa do ourives que produz uma joia, então tenho certeza de que teria falhado completamente em meu empreendimento. Os artigos que compõem a tese deixam tantas questões em aberto e tantos temas adjacentes sem tratamento que o produto final seria feio e inutilizável. Uma metáfora alternativa é ver o trabalho realizado como a tarefa do garimpeiro. Diferentemente do ourives, que precisa compor um produto harmonioso e completo para ser bem-sucedido, o garimpeiro pode cumprir bem sua tarefa se pelo menos parte dos itens que encontrar forem valiosos. Nesta tese, reuni alguns achados do garimpo. Defendi, primeiramente que os resultados de experimentos de tipo Libet disponíveis não mostram que nossas escolhas sejam determinadas por eventos no cérebro de uma maneira que poderia ameaçar o libertismo. Também apontei limitações da proposta de aprimoramento das práticas cotidianas de responsabilidade derivada do ceticismo sobre a responsabilidade moral e o livre-arbítrio. E, mais centralmente, defendi que a discussão filosófica tradicional sobre a existência da responsabilidade moral é insuficiente para se promover o aprimoramento das práticas cotidianas de responsabilidade e propus o modelo de aprimoramento como uma alternativa. O modelo de aprimoramento oferece direções para uma investigação interdisciplinar que defendo estar mais bem equipada para realizar o objetivo de aprimorar as práticas cotidianas de responsabilidade. Nele, a ciência recebe a tarefa de descrever as causas e efeitos das práticas de responsabilidade tal como se encontram em nossas vidas cotidianas, e também a tarefa de descrever maneiras pelas quais se pode efetivar mudanças que visem à modificação dos aspectos problemáticos. A detecção desses aspectos problemáticos, no entanto, cabe a uma investigação de tipo normativo que, na seção anterior, sugeri poder ser entendida como um tipo de investigação em ética aplicada. Desenvolver essa sugestão em maiores detalhes é certamente uma das várias tarefas que esta tese deixa para investigações futuras. Por ora, estarei contente se algumas das propostas aqui reunidas forem de algum valor.

REFERÊNCIAS

Afifi, T. O.; Brownridge, D. A.; Cox, B. J.; Sareen, J. Physical punishment, child abuse and psychiatric disorders. *Child Abuse and Neglect* v. 30, p. 1093-1103, 2006.

Afifi, T. O.; Romano, E. Ending the spanking debate. *Child Abuse & Neglect* v. 71, n. Supplement C, p. 3-4, 2017.

Alexander, P.; Schlegel, A.; Sinnott-Armstrong, W.; Roskies, A.; Tse, P. U. and Wheatley, T., Dissecting the readiness potential: An investigation of the relationship between readiness potentials, conscious willing, and action. In: Mele, A. R. (Ed.), *Surrounding free will: Philosophy, Psychology, Neuroscience*, p. 203-230. New York: Oxford University Press, 2015.

Aquino, K.; Tripp, T. M.; Bies, R. J. How employees respond to personal offense: The effects of blame attribution, victim status, and offender status on revenge and reconciliation in the workplace. *Journal of Applied Psychology* v. 86, n. 1, p. 52-59, 2001.

Aristóteles. *Ética a Nicômaco*. São Paulo: Abril Cultural, 1973.

Ateah, C. A.; Durrant, J. E. Maternal use of physical punishment in response to child misbehavior: Implications for child abuse prevention. *Child Abuse & Neglect* v. 29, n. 2, p. 169–185, 2005.

Bedau, H. A.; Kelly, E. Punishment. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, 2015.

Brooks, T. *Punishment*. London: Routledge, 2012.

Caruso, G. D. Free will skepticism and criminal behavior: A public health-quarantine model. *Southwest Philosophy Review* v. 32, n. 1, p. 25-48, 2016.

Caruso, G. D.; Morris, S. G. Compatibilism and retributivist desert moral responsibility: On what is of central philosophical and practical importance, *Erkenntnis* v. 82, n. 4, p. 837-855, 2017.

Chisholm, R., Human freedom and the self. In: Pereboom, D. (Ed.), *Free will*. Indianapolis: Hackett, 1964, p. 172-184.

Clark, C. J.; Luguri, J. B.; Ditto, P. H.; Knobe, J.; Shariff, A. F.; Baumeister, R. F. Free to punish: A motivated account of free will belief, *Journal of Personality and Social Psychology* v. 106, n. 4, p. 501-513, 2014.

Coates, D. J.; Tognazzini, N. A. The nature and ethics of blame, *Philosophy Compass* v. 7, n. 3, p. 197-207, 2012.

Copp, D. Introduction: Metaethics and normative ethics. In: Copp, D. (Ed.), *The Oxford handbook of ethical theory*. Oxford: Oxford University Press, p. 3-35, 2006.

Darwall, S. L. Theories of ethics. In: Frey, R. G.; Wellman, C. H. (Ed.), *A companion to applied ethics*. Malden: Blackwell, 2003, p. 17-37.

Davidson, D., Mental events. In: (Ed.), *Essays on actions and events*, p. 207-227. Oxford: Clarendon Press, 1970.

Dennett, D. C. *Elbow room: The varieties of free will worth wanting*. Oxford: Oxford University Press, 1984.

Duff, A., Legal and moral responsibility. *Philosophy Compass* v. 4, n. 6, p. 978-986, 2009.

Durrant, J.; Ensom, R. Physical punishment of children: lessons from 20 years of research, *Canadian Medical Association Journal* v. 184, n. 12, p. 1373-1377, 2012.

Earman, J., *A primer on determinism*. Dordrecht: D. Reidel, 1986.

Eshleman, A. Moral responsibility. In: Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2016.

Fischborn, M., Monismo anômalo: Uma reconstrução e revisão da literatura. *Principia* v. 18, n. 1, p. 53-66, 2014.

Fischer, J. M. *The metaphysics of free will*. Oxford: Blackwell Publishers, 1994.

Fischer, J. M.; Ravizza, M. *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press, 1998.

Gershoff, E. T. Corporal punishment by parents and associated child behaviors and experiences: a meta-analytic and theoretical review. *Psychological Bulletin* v. 128, n. 4, p. 539-579, 2002.

Global Initiative to End All Corporal Punishment of Children, http://www.endcorporalpunishment.org/, 2017.

Gomes, G., Preparing to move and deciding not to move. *Consciousness and Cognition* v. 19, n. 1, p. 457-459, 2010.

Hart, H. L. A., *Punishment and responsibility: Essays in the philosophy of law (Second Edition)*. Oxford: Oxford University Press, 2008.

Haynes, J.-D. Beyond Libet: Long-term prediction of free choices from neuroimaging signals. In: Sinnott-Armstrong, W. & Nadel, L. (Ed.), *Conscious will and responsibility: A tribute to Benjamin Libet*, p. 85-96. Oxford: Oxford University Press, 2011.

Kane, R. *The significance of free will*. Oxford: Oxford University Press, 1996.

de Keijser, J. W. and Elffers, H., Punitive public attitudes: A threat to the legitimacy of the criminal justice system?. In: Oswald, M. E.; Bieneck, S. & Hupfeld-Heinemann, J. (Ed.), *Social psychology of punishment of crime*, p. 55-74. Chichester: Wiley-Blackwell, 2009.

Libet, B., Do we have free will?. *Journal of Consciousness Studies* v. 6, n. 8-9, p. 47-57, 1999.

Libet, B.; Gleason, C. A.; Wright, E. W. and Pearl, D. K., Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain* v. 106, n. , p. 623-642, 1983.

Lycan, W., Dando ao dualismo o que lhe é devido. Tradução de Bruno Faez. *Cognitio-Estudos* v. 11, n. 2, p. 271-286, 2015.

Madden, M. E.; Janoff-Bulman, R. Blame, control, and marital satisfaction: Wives' attributions for conflict in marriage. *Journal of Marriage and Family* v. 43, n. 3, p. 663-674, 1981.

Malle, B. F.; Guglielmo, S.; Monroe, A. E. A theory of blame. *Psychological Inquiry* v. 25, p. 147-186, 2014.

Marques, B. S., An issue for Wegner's theory about the conscious will: The readiness potential does not conclusively represent preparation for an action. *Veritas* v. 62, n. 3, p. 860-876, 2017.

Mele, A. *Free will and luck*. Oxford University Press: Oxford University Press, 2006.

Misirlisoy, E.; Haggard, P. A neuroscientific account of the human will. In: Sinnott-Armstrong, W. (Ed.), *Moral psychology: Free will and moral responsibility*. Cambridge, MA: MIT Press, 2014, p. 37-42.

Nahmias, E., Is free will an illusion? Confronting challenges from the modern mind sciences. In: Sinnott-Armstrong, W. (Ed.), *Moral psychology: Free will and moral responsibility*. Cambridge, MA: MIT Press, 2014, p. 1-25.

Nelkin, D. K. *Making sense of freedom and responsibility*. Oxford: Oxford University Press, 2011.

O'Connor, T. *Persons and causes: The metaphysics of free will*. New York: Oxford University Press, 2002.

Pereboom, D. (Ed.). *Free will*. Indianapolis: Hackett, 2009.

Pereboom, D. *Living without free will*. Cambridge: Cambridge University Press, 2001.

Pereboom, D. *Free will, agency, and meaning in life*. New York: Oxford University Press, 2014.

Robichaud, P.; Wieland, J. W. *Responsibility: The epistemic condition*. New York: Oxford University Press, 2017.

Roskies, A. Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Science* v. 10, n. 9, p. 419-423, 2006.

Roskies, A.; Nahmias, E. “Local determination”, even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn. *Philosophical Psychology* v. 30, n. 1-2, p. 185-197, 2017.

Schurger, A.; Sitt, J. D. and Dehaene, S., An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences* v. 109, n. 42, p. E2904-E2913, 2012.

Soon, C. S.; Brass, M.; Heinze, H.-J.; Haynes, J.-D. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* v. 11, n. 5, p. 543-545, 2008.

Steward, H., Moral responsibility and the irrelevance of physics: Fischer’s semi-compatibilism vs. anti-fundamentalism. *The Journal of Ethics* v. 12, n. 2, p. 129-145, 2008.

Trevena, J. and Miller, J., Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and Cognition* v. 19, n. 1, p. 447-456, 2010.

Vargas, M., *Building better beings: A theory of moral responsibility*. Oxford: Oxford University Press, 2013.

Wolf, S. Asymmetrical freedom. *Journal of Philosophy* v. 77, n. March, p. 151-166, 1980.

Zimmerman, M. Varieties of moral responsibility. In: Clarke, R.; McKenna, M. & Smith, A. M. (Ed.), *The nature of moral responsibility: New essays*. New York: Oxford University Press, 2015, p. 45-64.

APÊNDICES

Apêndice A – Materiais do Artigo 3

1 Descrição do caso

Em abril de 2016, um homem de iniciais M.C.D., então com 32 anos, realizou um assalto à mão armada. Ele roubou a motocicleta de uma pessoa que se preparava para deixar o estacionamento do supermercado. Com um revólver, M.C.D. ameaçou a vítima, que acabou entregando a motocicleta.

A punição prevista para crimes como esse é de 5 a 15 anos de prisão.

2 Manipulação

Condição A:

Como parte de um programa que testa alternativas para a prevenção de crimes, o juiz solicitou uma avaliação do caso a uma equipe de criminologistas, psicólogos e psiquiatras. Com base na história e personalidade de M.C.D., a equipe avaliou se ele satisfazia as exigências para participar de um programa de reintegração social de 8 meses que se mostrou muito eficaz em outros casos.

A equipe concluiu que, participando do programa, as chances de M.C.D. repetir o crime seriam praticamente nulas (cerca de 10%), mas que o tratamento seria ineficaz se acompanhado de punição. Segundo a equipe, a aplicação de punição aumentaria para cerca de 80% as chances de M.C.D repetir o crime.

Pergunta: Você acha que M.C.D. deveria participar do tratamento de 8 meses? {Sim; Não}

Condição B:

Como parte de um programa que testa alternativas para a prevenção de crimes, o juiz solicitou uma avaliação do caso a uma equipe de criminologistas, psicólogos e psiquiatras. Com base na história e personalidade de M.C.D., a equipe avaliou se ele satisfazia as exigências para participar de um programa de reintegração social de 8 meses que se mostrou muito eficaz em outros casos.

A equipe concluiu que no caso de M.C.D. a efetividade do programa seria praticamente a mesma da punição. Participando do programa de reintegração social ou recebendo a punição usual, as chances de M.C.D. repetir o crime seriam de cerca de 40%.
Pergunta: Você acha que M.C.D. deveria participar do tratamento de 8 meses? {Sim; Não}

Condição C:
[Nenhuma informação adicional]

3: Questões sobre a punição recomendada (1), crenças específicas sobre livre-arbítrio (2, 4, 6) e atribuições de culpa (3) e responsabilidade (5)

1. Que punição você acha que M.C.D. deveria receber (em anos de prisão)? {Nenhuma punição; 1–15 anos, em intervalos de um ano}
Para as próximas respostas, considere a escala abaixo:
{1: Discordo plenamente; 2: Discordo; 3: Discordo parcialmente; 4: Nem discordo, nem concordo; 5: Concordo parcialmente; 6: Concordo; 7: Concordo plenamente}
Usando a escala, indique seu grau de acordo ou desacordo com os enunciados abaixo:
2. “M.C.D. exerceu seu livre-arbítrio ao realizar o roubo.”
3. “M.C.D. é culpado pelo roubo.”
4. “M.C.D. poderia ter decidido não realizar o roubo”
5. “M.C.D. é responsável pelo roubo”
6. “M.C.D. decidiu realizar o roubo livremente”

4 Questões sobre crenças gerais sobre livre-arbítrio (Subescala ‘Livre-arbítrio’ do Inventário do Livre-arbítrio)

Para as próximas respostas, considere novamente a escala abaixo:
{1: Discordo plenamente; 2: Discordo; 3: Discordo parcialmente; 4: Nem discordo, nem concordo; 5: Concordo parcialmente; 6: Concordo; 7: Concordo plenamente}
Usando a escala, indique seu grau de acordo ou desacordo com os enunciados abaixo:
7. “As pessoas sempre podem agir de outro modo”
8. “As pessoas sempre têm livre-arbítrio.”
9. “O modo como a vida das pessoas se desdobra é completamente dependente da escolha delas.”

10. “As pessoas têm, essencialmente, controle total sobre suas decisões e ações.”
11. “As pessoas têm livre-arbítrio mesmo quando suas escolhas são completamente limitadas por circunstâncias externas.”

5 Questões demográficas

Nas perguntas abaixo, estamos interessados em algumas informações gerais sobre você.

12. Sexo: {masculino; feminino; não informar}
13. Cor ou raça: {branca; preta; parda; amarela; indígena; outro; não informar}
14. Idade: {menos de 18 anos; 18-20 anos; 21-30 anos; 31-40 anos; 41-50 anos; 51-60 anos; 61-70 anos; 71 anos ou mais}
15. Unidade da federação em que reside: {AC; AL; AP; AM; BA; CE; DF; ES; GO; MA; MT; MS; MG; PA; PB; PR; PE; PI; RJ; RN; RS; RO; RR; SC; SP; SE; TO; outro}
16. Nível mais alto de escolaridade frequentado: {Nenhum; Ensino fundamental; Ensino médio; Ensino superior; Pós-Graduação}
17. Área de estudos: {Ciências Exatas e da Terra; Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Agrárias; Ciências Sociais Aplicadas; Ciências Humanas; Linguística, Letras e Artes; Outro}