

# CAUSAS EXCLUDENTES

ANDRÉ FUHRMANN

Universität Konstanz

WILSON MENDONÇA

Universidade Federal do Rio de Janeiro

## Abstract

*We defend J. Kim's principle of explanatory exclusion from a recent criticism advanced by A. Marras. We show that the principle follows from a less controversial principle of causal exclusion together with the assumption that claims of explanation are factual. We resolve the tension produced by Marras' argument by drawing a distinction between causal and explanatory relevance. In cross-level explanations (mental-to-physical and physical-to-mental) the explanans property is not causally but explanatorily relevant to the explanandum. This calls for an account of how explanatorily relevant properties are grounded in causally relevant properties which in turn ultimately depend on causally efficacious properties.*

## I

A literatura contemporânea contém uma grande variedade de formulações do que continua sendo, talvez, a questão mais vexatória na filosofia da mente, o problema da causação mental. Uma maneira de formular o problema é mediante a apresentação de um argumento cuja conclusão afirma que a eficácia de causas mentais na produção de um efeito físico é excluída logo que uma causa física suficiente esteja disponível. Nesta roupagem, o problema da causação mental se torna o problema da exclusão causal, como Jaegwon Kim o denomina. Se efeitos físicos têm causas físicas completas, as causas mentais não são necessárias para dar forma ao mundo físico. O mental é, portanto,

© *Principia*, 4(2) (2000), pp. 257–75. Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

ou bem causalmente inerte, ou bem causalmente eficaz somente sob a condição de sobredeterminação dos seus efeitos. Ademais, sob a suposição de que efeitos mentais são de algum modo constituídos ou realizados por eventos físicos, a situação permanece inalterada quando se consideram as causas mentais e seus supostos efeitos mentais. Assim, a idéia da causação mental-física, bem como mental-mental, parece ser incoerente ou envolver a idéia de uma sobredeterminação maciça e regular de efeitos físicos e mentais.

Alguns filósofos sugeriram que o problema da causação mental é uma ilusão filosófica que se evapora logo que abandonemos uma certa fixação a propósito do conceito de causação. Deveríamos, como tais "explicativistas" sugerem, considerar o fato de que a nossa compreensão de causação é derivada principalmente da nossa prática de explicação.<sup>1</sup> Explicativistas argumentam, a seguir, que não há um problema de explicação mental: freqüentemente e de maneira não-controvertida, explicamos alguns efeitos em termos de causas mentais. Portanto, a explicação mental não se constitui num problema e o mesmo deve ser dito da causação mental, corretamente compreendida.

Ainda que possamos conceder ao explicativista que as explicações mentais sejam em geral bem sucedidas e que a causação mental de fato ocorra — que a explicação mental e a causação mental sejam um *factum brutum* — podemos ainda insistir na pergunta sobre como são possíveis tais *facta*.<sup>2</sup> A questão parece ser perfeitamente pertinente. E ela não recebe uma resposta quando se muda de assunto. Contudo, há alguns filósofos que estariam relutantes até mesmo a fazer esta concessão. Eles ou elas argumentariam que o problema da causação mental nos força a aceitar uma conclusão epifenomenalista, a saber: que eventos mentais são causalmente inertes.<sup>3</sup>

Kim respondeu de várias maneiras à tentativa explicativista de dissolver o problema da causação mental.<sup>4</sup> Talvez, a resposta mais direta — uma resposta que envolve diretamente o explicativista — seja a que estende o problema da exclusão do âmbito inicial da causação para o âmbito da explicação. De fato, Kim sugere que não há, aqui, dois problemas distintos. Se puder ser considerado como bem sucedido, isto mostraria que a noção de explicação não é menos problemática do que a noção de causação. Eis a formulação

proposta por Kim (1989, 64f.) para o problema da exclusão causal/explanatória:

Quando nos confrontamos com duas possíveis causas (ou duas possíveis explicações causais) de um único evento, podemos contar, inicialmente, com as seguintes alternativas: (a) cada uma delas é uma causa suficiente e o efeito é sobredeterminado; (b) cada uma delas é necessária e, tomadas em conjunto, elas formam uma causa suficiente (isto é, cada uma delas é somente uma “causa parcial”); (c) uma é parte da outra; (d) as causas são de fato uma e a mesma causa, mas dadas sob diferentes descrições; (e) uma (presumivelmente a causa mental) é, de alguma forma, redutível à outra; (f) uma (novamente a causa mental) é uma causa derivada e o seu *status* causal depende, de alguma forma, da causa neural N. (...) O ponto que eu gostaria de salientar é o seguinte: a presença de duas explicações causais, cada uma pretendendo oferecer uma explicação causal completa de um evento dado, cria uma situação instável, requerendo de nós que encontremos uma explicação de como as duas supostas causas estão relacionadas entre si. Este é o problema da “exclusão causal/explanatória.”

Kim declara que “para o nosso caso, a maior parte delas [isto é, das alternativas mencionadas] são fracassadas de início.” No caso de (b), isto é assim porque o problema da exclusão é colocado sob a suposição de que as causas em questão não são coordenadas. As alternativas (c), (d) e (e) vão contra a suposição de que as causas são distintas ou, de qualquer forma, que uma não é redutível à outra. E (f), se não representar mais um caso de redutibilidade, não oferece solução, pois se baseia numa equivocação da noção de causa — noções derivadas ou supervenientes de causação são relações causais somente num sentido muito amplo de causa (causação sem “impacto”).<sup>5</sup>

Por que Kim afirma que (a) é fracassada de início? Algumas vezes, cenários de sobredeterminação são elaborados na literatura sobre a causação (por exemplo, dois projéteis atingindo ao mesmo tempo um alvo comum). Tais casos dependem de circunstâncias altamente improváveis. Mas se a causação mental tivesse de ser explicada como um caso de sobredeterminação coincidente, a coincidência teria de ser não excepcional, mas sim regular. Tal coincidência regular

seria não somente surpreendente (como no caso dos dois projéteis), mas verdadeiramente miraculosa. Em outras palavras, o fato de que consideramos os casos de causação simultânea como mais ou menos incríveis, embora possíveis, mostra que rejeitamos a idéia segundo a qual a causação simultânea possa ser uma característica onipresente do nosso mundo. Em resumo, estamos, aparentemente, assumindo um princípio de exclusão causal: nenhum evento pode ter mais de uma causa completa e independente. Como os casos de causação simultânea não podem ser excluídos *a priori*, o princípio não pode ser uma verdade conceitual. Ele parece, contudo, ter o *status* de um truísmo falível. Este princípio gera o problema da exclusão causal. Para poder gerar o problema análogo no âmbito das explicações, Kim deve pressupor um princípio análogo, isto é, o princípio da exclusão explanatória: “Nenhum evento pode ter mais do que uma explicação completa e independente” (Kim 1989, 79).

## II

Para os nossos propósitos é útil formular estes princípios com algum cuidado. Começemos pela consideração da explicação como uma relação  $Ex(Cc, Ee)$  entre proposições, onde o primeiro elemento,  $Cc$ , afirma que um certo evento  $c$ , tendo a propriedade  $C$ , ocorreu, ao passo que o segundo elemento,  $Ee$ , afirma que o fato  $a$  ser explicado,  $e$  tendo a propriedade  $E$ , ocorreu.<sup>6</sup>

Deve-se notar que a formulação da exclusão explanatória proposta por Kim é ambígua. Por um lado, a expressão “explicação” tal como aparece na passagem citada pode referir-se ao *par* que consiste nas proposições de que o *explanans* ocorreu e que o *explanandum* ocorreu — como em “‘Ela tem sucesso porque é bem-sucedida’ não é uma boa explicação.” Por outro lado, a expressão “explicação” pode referir-se somente ao *explanans* — como em “O fato de que a maçã estava envenenada é uma boa explicação para a sua morte súbita.” Finalmente, há uma terceira versão deflacionária de acordo com a qual os eventos brutos envolvidos nos respectivos *explanantia* (independentemente das descrições possíveis) devem ser independentes entre si. De acordo com isto, há três candidatos para o

## Princípio da exclusão explanatória

XE Se  $Ex(Cc, Ee)$  e  $Ex(Dd, Ee)$  são explicações completas de um dado evento  $e$ , então

- (a) ...  $Cc$  e  $Dd$  não são independentes entre si;
- (b) ...  $Ex(Cc, Ee)$  e  $Ex(Dd, Ee)$  não são independentes entre si;
- (c) ...  $c$  e  $d$  não são independentes entre si.

Uma ambigüidade análoga não surge no

## Princípio da exclusão causal

XC Se  $c$  e  $d$  são causas completas de um dado evento  $e$ , então  $c$  e  $d$  não são independentes entre si.

Obviamente, o conteúdo exato destes princípios depende do sentido dado à noção de (in)dependência. Em XE(c) e em XC trata-se de independência de causas. Em XE(b) precisamos de alguma noção de independência de explicações que, por sua vez, dependerá provavelmente de uma noção de *explanantia* independentes, tal como ela ocorre em XE(a).

No que diz respeito à independência de causas, a caracterização negativa indicada pelas alternativas (c) — (e) na passagem já citada de Kim será suficiente para o presente propósito. Assim a independência de causas impede, em particular, que as causas sejam arranjadas seqüencialmente numa única cadeia causal ou que uma constitua, de alguma maneira, a outra.<sup>7</sup>

Consideremos agora o par:

$Cc$  = o evento que causou o acidente ocorreu;

$Ee$  = o acidente  $e$  ocorreu.

Se a questão for “Por que o acidente ocorreu?” a resposta “Porque o evento que causou o acidente ocorreu” é claramente destituída de valor explanatório.  $Cc$  não é um *explanans* para  $Ee$ . Portanto,  $Cc$  não está numa relação de explicação com  $Ee$ . Em outras palavras, a afirmação que  $Cc$  explica  $Ee$  é falsa. Todavia,  $c$ , o evento referido em  $Cc$ , causou o evento  $e$ , referido em  $Ee$ . Portanto, a verdade da proposição

causal (que  $c$  causou  $e$ ) não é suficiente para que possamos afirmar verdadeiramente de qualquer descrição  $D$  do evento  $c$  que  $D$  explica a ocorrência de  $e$ . Um argumento similar mostraria que o valor explanatório depende também de como o evento causado é descrito. O que dizer do sentido oposto? A proposição que a ocorrência de  $c$  (sob qualquer descrição) explica causalmente a ocorrência de  $e$  (sob qualquer descrição) implica a proposição que  $c$  causou  $e$ ? Se tomarmos proposições de explicação causal ao pé da letra — e porque não deveríamos fazê-lo? — a implicação em questão expressa simplesmente o conteúdo factual de uma explicação bem sucedida. Portanto, temos o

*Princípio do realismo explanatório*

RE Se  $Ex(Cc, Ee)$ , então  $c$  causou  $e$ .

O rótulo “realismo explanatório” foi retirado de Kim (1989), mas gostaríamos de enfatizar que ele expressa um realismo minimalista a propósito de explicações causais. A negação de RE só pode ser vista como significando o compromisso com uma teoria do erro sistemático ou uma abordagem ficcionalista da explicação causal — opções que não discutiremos aqui.

Obviamente, RE e XC, em conjunto, implicam XE( $c$ ). Ademais, se os eventos  $c$  e  $d$ , referidos em  $Cc$  e  $Dd$ , respectivamente, não são independentes, então  $Cc$  e  $Dd$  tampouco serão independentes. Assim, XC e RE, em conjunto, implicam XE( $a$ ). E se  $Cc$  e  $Dd$  são mutuamente dependentes, as explicações  $Ex(Cc, Ee)$  e  $Ex(Dd, Ee)$ , nas quais eles aparecem como *explanantia*, não podem ser independentes. Portanto, XE( $a$ ) implica XE( $b$ ). Assim, XC e RE, em conjunto, implicam XE em todas as três versões.

Podemos ilustrar este argumento com um caso de causação mental-física. Podemos supor que duas explicações completas são oferecidas para o mesmo efeito físico  $e$  (Pedro ergue a sua mão, por exemplo). Uma explicação,  $Ex(Mm, Ee)$ , cita uma causa mental  $m$  (a instanciação, por parte de Pedro, de alguma propriedade mental  $M$ ); a outra,  $Ex(Pp, Ee)$ , invoca uma causa física  $p$  (a instanciação, por parte de Pedro, de alguma propriedade física  $P$ ). Para que ambas as explicações sejam verdadeiras, RE requer que  $e$  seja realmente

causado por  $m$  e por  $p$ . E como ambas as explicações são completas *ex hypothesi*,  $m$  e  $p$  devem ser causas completas de  $e$ . De acordo com XC,  $m$  e  $p$  não podem ser independentes entre si; portanto, XE(c) é verdadeiro. Para tornar as coisas mais simples, podemos assumir que só há um evento causador, isto é,  $m = p = c$ . Se a propriedade mental  $M$  e a propriedade física  $P$  identificam o mesmo evento, então elas não podem ser independentes entre si. Assim, XE(a) também é verdadeiro. Ademais, se  $M$  e  $P$  estão numa relação de dependência, as explicações  $Ex(Mc, Ee)$  e  $Ex(Pc, Ee)$  não podem mais ser consideradas como independentes, o que satisfaz XE(b).

Em ponto este argumento da exclusão explanatória poderia ser bloqueado? A inferência de XC e RE a XE(c) é imediata. Mas para XE(a) e XE(b) dois princípios de inferência adicionais são pressupostos. Em primeiro lugar, ao proceder na direção do conseqüente XE(a), assumimos que, se os eventos invocados em dois *explanantia* não são independentes, os próprios *explanantia* tampouco são independentes — em resumo:

(A) Se  $c$  e  $d$  não são independentes, então  $Cc$  e  $Dd$  não são independentes.<sup>8</sup>

Em segundo lugar, ao inferir XE(b) de XE(a), assumimos que duas explicações de um evento *explanandum* e não são independentes se os respectivos *explanantia* estão numa relação de dependência:

(B) Se  $Cc$  e  $Dd$  não são independentes, então  $Ex(Cc, Ee)$  e  $Ex(Dd, Ee)$  não são independentes.

Os que não aceitam o argumento acima que leva de RE e XC a XE(a) devem agora explicitar um sentido de (in)dependência que torne (A) falso. Consideramos que tão logo se conceda (A) não é mais possível resistir a (B). Por esta razão, concentraremos nossa atenção, a partir de agora, na versão XE(a) do princípio da exclusão explanatória.

Kim apresenta o princípio da exclusão explanatória da seguinte forma:

Suponha que  $c$  e  $d$  sejam invocados como fornecendo, cada um deles, uma explicação completa de  $e$ . Considere as duas questões: (1)  $e$  teria ocorrido se  $c$  não tivesse ocorrido? e (2)  $e$  teria ocorrido se  $d$

não tivesse ocorrido? Se a resposta for um “sim” a ambas as questões, este é um caso clássico de sobre-determinação (...). Se a resposta for um “não” a pelo menos uma das questões, por exemplo, à primeira, a razão para isto é que se  $c$  não tivesse ocorrido  $d$  tampouco teria ocorrido. E isto quer dizer que  $c$  e  $d$  não são independentes e, portanto, que as duas explicações não são explicações independentes de  $e$ . (Kim 1989, 92)

Até a parte introduzida por “portanto,” a passagem apresenta um argumento para o princípio da exclusão causal. Para justificar a transição súbita para XE(a), Kim precisa assumir (A). Esta transição é necessária. Pois para mostrar que a retirada estratégica do âmbito da causação mental para o âmbito da explicação mental não resolve o problema da exclusão, não é suficiente permanecer com a versão minimalista XE(c) do princípio da exclusão explanatória — a versão mais forte XE(a) é necessária.

Pois o explicativista há de manter que, embora os eventos referidos pelos dois *explanantia* possam não ser independentes — eles podem ser o mesmo evento —, os *explanantia* podem ser distintos porque subsumem os eventos — ou o único evento — sob propriedades explanatoriamente distintas. Considere, por exemplo, a proposição  $Ex(Mc, P^*e)$  segundo a qual a instanciação de alguma propriedade mental  $M$  explica completamente porque algum evento  $e$  do tipo físico  $P^*$  ocorreu. No caso em que temos boas razões para pensar que o evento  $e$  *qua*  $P^*$  também tem uma explicação completa em virtude de algum aspecto  $P$  de  $c$ , estamos confrontados com uma explicação excludente *prima facie*, a saber:  $Ex(Pc, P^*e)$ . Evidentemente, XE(c) não oferece, neste ponto, uma informação útil para continuar o argumento. Mas de acordo com XE(a), os *explanantia* não podem ser independentes entre si: ou  $Mc$  depende de  $Pc$  ou vice-versa. Suponha, como parece ser mais plausível, que  $Mc$  depende de  $Pc$ . Neste caso o trabalho explanatório real é realizado por algum aspecto físico do evento causador e a propriedade  $M$  permanece explanatoriamente inerte. Portanto, a afirmação que o evento causador, *qua* instanciação de alguma propriedade mental  $M$ , explica o efeito revela-se como falsa. Ademais, se aceitarmos que todo evento físico tem uma explicação em virtude de alguma propriedade física de seus eventos causadores (o fechamento causal do mundo físico), então XE(a) nos

leva forçosamente à conclusão que as propriedades mentais nunca podem explicar verdadeiramente seus efeitos.

Em resumo, XE(a), não XE(c), gera o problema da exclusão causal de uma forma perfeitamente paralela ao modo como emerge o problema da exclusão causal. XE(a) pode, portanto, ser considerado o principal princípio da exclusão explanatória. Para evitar o problema da exclusão explanatória, teríamos de rejeitar XE(a) — tratando as outras pressuposições que integram o argumento como isentas de crítica. Mas dados RE e XC, XE(a) é uma conseqüência de (A). No que se segue, usaremos simplesmente o símbolo “XE” para fazer referência ao princípio da exclusão explanatória na sua versão principal, isto é, na versão (a). Este princípio, como já observado, depende crucialmente de (A), a tese segundo a qual os *explanantia* que invocam causas não-independentes não podem ser independentes entre si. Na próxima seção examinaremos (A) mais de perto.

### III

Em dois artigos recentes (1998, 2000) Ausonio Marras apresentou uma análise meticulosa do princípio da exclusão explanatória proposto por Kim e avançou um argumento sofisticado contra o princípio.

Marras distingue, como também o fizemos acima, entre o *explanandum* de uma dada explicação (o “fato *explanandum*”) e o evento ao que o *explanandum* faz referência (o “evento *explanandum*”). Os *explananda* são individuados pelas propriedades sob as quais eles subsumem o evento a ser explicado. Como um mesmo evento pode ser identificado sob diferentes tipos, ele pode ser associado a *explananda* diferentes, isto é, distintos. O mesmo vale, *mutatis mutandis*, para os eventos causadores e os correspondentes *explanantia*. Apesar disto, há, obviamente, um sentido apropriado no qual se pode dizer que uma explicação bem sucedida com um *explanans* Cc e um *explanandum* Ee é uma explicação de por que o evento e, referido em Ee, ocorreu. Dada esta distinção entre *explananda* e eventos explicados, por um lado, e entre *explanantia* e eventos causadores, por outro, parece-nos possível afirmar que há duas explicações distintas, Ex(Cc, Ee) e Ex(Dc, Fe), de um simples evento e, ao qual se referem Ee e Fe.<sup>9</sup> Se

esta possibilidade *prima facie* puder ser realmente mantida, então, de acordo com o argumento de Marras, teríamos um contra-exemplo para o princípio

- (i) "... não pode haver mais do que uma única explicação completa e independente para um único *evento*." (1998, 439)

Deve-se notar que isto deixa intacto o valor de verdade do princípio

- (ii) "... não pode haver mais do que uma única explicação completa e independente para um único *explanandum*." (1998, 449)

De acordo com Marras (1998, 449), o último princípio "pode bem ser verdadeiro." Deve-se notar também que Marras não põe em questão o princípio XC da exclusão causal. De fato, todos os exemplos do tipo que Marras considera contrário a (i) são tais que um único evento causador é identificado sob diferentes tipos. Deste modo, nos supostos contra-exemplos a (i), os dois *explanantia* *Cc* e *Dd* da ocorrência de *e* não são ditos independentes porque se referem a diferentes eventos, mas sim porque se referem diferentemente a um único evento. O princípio da exclusão causal é, deste modo, respeitado. Notemos, também, que Marras não questiona a tese de que explicações causais são factuais no sentido de RE.<sup>10</sup>

Aqui será suficiente apresentar o argumento de Marras numa forma abstrata. Sejam *M* e *M\** propriedades mentais, e *P* e *P\** propriedades físicas. Considere duas explicações completas *Ex*(*Mc*, *M\*e*) e *Ex*(*Pc*, *P\*e*). As duas explicações são explicações completas do mesmo evento *e*. Como Marras não quer violar XC ou RE, ele nos convida a assumir que os dois *explanantia* se referem ao mesmo evento *c*, embora sob diferentes tipos. A questão agora é: Serão *Mc* e *Pc* *explanantia* independentes para o mesmo evento *e*? Se a resposta for sim, XE será falso.

Eis o argumento simples de Marras para mostrar que *Mc* e *Pc* são independentes:

- (1) Se *Mc* e *Pc* se distinguem a propósito de suas respectivas relevâncias causais, então eles são *explanantia* independentes.

(2)  $Mc$  e  $Pc$  se distinguem a propósito de suas respectivas relevâncias causais.

Portanto,

(3)  $Mc$  e  $Pc$  são *explanantia* independentes.

O suporte racional de (1) é a idéia de que, se a instanciação de uma propriedade  $F$  deve explicar (causalmente) a instanciação de uma propriedade  $G$ , a primeira deve ser causalmente relevante para a segunda, isto é,

RX Se  $Ex(Fc, Ge)$ , então  $Fc$  é causalmente relevante para  $Ge$ .

Se  $Fc$  e  $F^*c$  se distinguem a propósito de suas relevâncias causais, então eles não podem ser livremente substituídos entre si em contextos de explicação. Neste sentido,  $Fc$  e  $F^*c$  seriam *explanantia* independentes. Vamos conceder que este raciocínio esteja correto para examiná-lo melhor mais tarde.

Marras (1998, 448) oferece a seguinte condição necessária para relevância causal:

RC Se  $c$  causa  $e$  e  $c$  é  $F$  e  $e$  é  $G$ , então  $Fc$  é causalmente relevante para  $Ge$  somente se o contrafactual  $\neg Fc \rightarrow \neg Ge$  for válido.

Logo voltaremos a esta condição “não-controversa” (Marras). Por que  $Mc$  e  $Pc$  se distinguem a propósito de suas relevâncias causais? Porque, segundo Marras,

(4)  $Mc$  é causalmente relevante para  $M^*e$ ;

(5)  $Mc$  não é causalmente relevante para  $P^*e$ ;

(6)  $Pc$  não é causalmente relevante para  $M^*e$ ;

(7)  $Pc$  é causalmente relevante para  $P^*e$ .

No que concerne a (4) e (7), Marras argumenta que os contrafactuais correspondentes

$$\neg Mc \rightarrow \neg M^*e \quad \text{e} \quad \neg Pc \rightarrow \neg P^*e$$

são verdadeiros. Deste modo, as correspondentes proposições de relevância causal (4) e (7) satisfazem uma condição necessária para que

possam ser verdadeiras. Podemos assumir que é de fato verdade que *c-qua-M* causa *e-qua-M\** e que *c-qua-P* causa *e-qua-P\**.

Feita esta concessão, podemos perguntar se (5) e (6) também são verdadeiros. Marras argumenta que os correspondentes contra-factuais

$$(5') \quad \neg Mc \rightarrow \neg P^*e \quad \text{e} \quad (6') \quad \neg Pc \rightarrow \neg M^*e$$

são falsos. Evidentemente, seria suficiente, para o propósito atual, mostrar que um dos dois, (5') ou (6'), é falso. Somos de opinião que Marras é bem sucedido num dos casos, mas não em ambos. Apresentemos, em primeiro lugar, o caso bem sucedido, (6').

Suponha que *c* é, ao mesmo tempo, *P* e *M* e, ademais, que, em virtude de ser *M*, *c* causou o fato de que *e* é *M\**. Considere agora o que ocorreria se *c* não fosse do tipo físico *P*. Neste caso, *e* deixaria de ser um evento do tipo *M\**? Não, pois podemos assumir que *P* não é o único tipo físico que pode realizar a propriedade mental *M*. Se *c* não tivesse sido *P*, poderia ter sido *P'* e *P'* poderia ter realizado *M*. Neste caso, *Mc* poderia continuar a causar *M\*e*. Portanto, ainda que *c* não fosse *P*, *e* poderia continuar sendo *M\**. Este raciocínio se baseia na existência de uma explicação bem sucedida que leva de *Mc* a *M\*e* e invoca o fato familiar de que propriedades mentais podem ser realizadas de múltiplas maneiras.

Consideremos agora o argumento de Marras contra (5'). Infelizmente, Marras (1998) não dá sequer uma indicação sobre o que poderia levar-nos a rejeitar (5'). Marras (2000, 147) contém a seguinte passagem:

*Em virtude do fato de e ser P\* (isto é, na medida em que e seja identificado, quanto ao tipo, como um evento físico do tipo P\*), c causou e em virtude do fato de c ser P e não em virtude do fato de c ser M; pois se c tivesse sido M sem ser P (mas, por exemplo, P', onde P' está na base de superveniência de M), c não teria causado uma instância de P\* (mas, ao invés disto, uma instância de P'', por exemplo).*

A passagem parece ter como objetivo corroborar  $\neg Pc \rightarrow \neg P^*e$  e rejeitar  $\neg Mc \rightarrow \neg P^*e$ , isto é, (5'). Grifamos a parte da passagem que se

refere ao segundo objetivo — está claro que nenhum argumento adicional é oferecido a propósito deste objetivo. A parte que vem após o ponto-e-vírgula formula um argumento para o primeiro contrafactual.

Talvez seja um argumento do seguinte tipo o que Marras tem em mente. Sabemos que  $e$  é  $P^*$  e que este fato foi causado em virtude do fato de que  $c$  é  $P$ . Ademais,  $P$  realizou a propriedade mental  $M$ . Suponha agora, contrafactualmente, que  $c$  não era  $M$ . Isto poderia ocorrer se anulássemos o nexos determinante entre  $P$  e  $M$ . Neste caso, estaríamos considerando um mundo de “zumbis” no qual  $c$  é  $P$  sem ser  $M$ . Num mundo assim, o vínculo causal no plano físico de  $Pc$  a  $P^*$  e ainda seria mantido —  $e$  continuaria a ser  $P^*$  em virtude do fato de que  $c$  é  $P$ . Portanto, não seria verdadeiro que, se  $c$  fosse  $M$ , então e não seria  $P^*$ .

Um mundo de zumbis, porém, não é somente contrafactual, mas também contrário à ordem natural: um mundo com uma estrutura nomológica diferente da nossa. A suposição contrafactual de que  $c$  não é  $M$  não deve implicar imediatamente que  $c$  poderia ser  $P$  sem ser  $M$ . Mundos que satisfazem a última descrição estão muito afastados do mundo real para que sejam considerados como parte das possibilidades abertas pela suposição de que  $c$  não é um evento do tipo  $M$ .

A falha deste argumento aponta para um argumento melhor. Considere as mesmas pressuposições razoáveis de antes e a suposição contrafactual de que  $c$  não era  $M$ . Em todos os mundos nomologicamente próximos ao nosso,  $c$  é destituído de toda e qualquer propriedade física que forma a base de superveniência de  $M$ . Em particular, em nenhum destes mundos  $c$  é  $P$ . Contudo,  $\neg Pc \rightarrow \neg P^*$  e continua a ser válido nestes mundo. Portanto,  $e$  não poderia ser  $P^*$  dado que  $c$  não é  $P$ . Isto quer dizer que o contrafactual  $\neg Mc \rightarrow \neg P^*$  e é verdadeiro, dado que  $\neg Pc \rightarrow \neg P^*$  e é verdadeiro.

Resumindo a discussão até este ponto, vimos que somente um dos dois contrafactuais (5') e (6') ligando propriedades mentais a físicas é verdadeiro. Como assinalado acima, isto é suficiente para estabelecer a premissa (2) do argumento simples de Marras. Devemos então aceitar sua conclusão? Neste caso teríamos obtido um contra-exemplo ao princípio da exclusão explanatória.

Deve-se notar que o contra-exemplo a XE(a) seria também um contra-exemplo a (A). Pois a conjunção da conclusão (3) do argumento simples de Marras com (A) implica que *c* é independente de si mesmo — o que certamente não pode ser. Assim, dados (1) e (2), (A) deve ser rejeitado.

#### IV

As credenciais da premissa (1) derivam de RX e RC. A conjunção dos dois implica

(†) Se  $Ex(Fc, Ge)$ , então  $\neg Fc \rightarrow \neg Ge$ .

Em termos mais extensos: se *Fc* explica *Ge*, então, se *Fc* não tivesse ocorrido, *Ge* tampouco teria ocorrido. Aparentemente, a tese não é muito promissora. Considere os seguintes exemplos.

Pedro está numa montanha. Repentinamente, uma avalanche de pedras avança na sua direção. Uma determinada pedra (número 13) atinge a cabeça de Pedro que morre logo em seguida. O fato de que a pedra de número 13 tenha atingido a cabeça de Pedro é uma explicação perfeita da sua morte. Contudo, se não tivesse sido atingido pela pedra de número 13, Pedro teria sido atingido por qualquer outra pedra. Na situação em que Pedro estava, ele não tinha chance — ele teria morrido mesmo que a pedra número 13 não o tivesse atingido. O contrafactual correspondente é falso e com isso (†) também é falso.

Marta é uma excelente debatedora. Ela também é a mais encantadora e a única com um passado limpo. Três boas razões para votar em Marta. Por fim, Paulo decide que sua reputação imaculada é a melhor razão para dar-lhe o seu voto. Mas ainda que seu passado não a diferenciasse dos outros candidatos, Paulo votaria em Marta. Mais uma vez, (†) se mostra falso.<sup>11</sup>

Felipe quer almoçar. Ele resolveu não gastar mais do que R\$ 10, exceto se lhe for oferecido espaguete carbonara. Neste caso, ele está disposto a gastar mais. Ocorre que espaguete carbonara é o único prato por menos de R\$ 10 no cardápio. Logo, ele pede espaguete carbonara. Como deve ser explicado o seu pedido? Parece que há

dois bons candidatos. Ele pede espaguete carbonara porque este é o único prato por menos de R\$ 10 e porque é o prato de que ele mais gosta. Porém, mesmo que custasse mais, ele ainda o teria pedido; e se não fosse espaguete carbonara, ele ainda teria pedido o único prato abaixo do limite estabelecido. Ambas as explicações não satisfazem (†).

Há muitos casos nos quais explicamos um evento ao mencionar um *explanans* *E*, reconhecendo, ao mesmo tempo, a existência de outros *explanantia* atuais ou potenciais que poderiam vir a ser dominantes se *E* não ocorresse e que explicariam o efeito em questão tão bem quanto *E*. Quando os *explanantia* alternativos são atuais (o exemplo de Paulo), podemos dizer que o *explanandum* mencionado é de fato explanatoriamente sobredeterminado; quando os *explanantia* alternativos são meramente potenciais, o *explanandum* é potencialmente sobredeterminado — os *explanantia* meramente potenciais são deslocados preemptivamente pelo *explanandum* atual. De qualquer modo, atual ou potencial, os contrafactuais correspondentes falham, ao contrário do que (†) prediz.

Uma resposta a tais aparentes contra-exemplos a (†) pode ser derivada de David Lewis (1986). De acordo com sua análise da causação, o evento *c* causa o evento *e* somente se for válido o contrafactual: e não teria ocorrido, se *c* não tivesse ocorrido. À primeira vista, isto é incompatível com as várias espécies de sobredeterminação explanatória. Lewis remove a incompatibilidade ao assinalar que sua análise se refere a eventos particulares finamente individuados. Assim, considerando o primeiro exemplo, a morte de Pedro subsequente ao choque com a pedra de número 13 é um evento diferente da sua morte provocada pela colisão com a pedra de número 17. A colisão de Pedro com a pedra de número 13 explica sua morte-causada-pelo-13. É verdade, portanto, que Pedro não teria morrido na forma particular como morreu (morte induzida pelo 13), se não tivesse sido atingido pelo 13. Isto deixa intacta a possibilidade de que ele ainda teria morrido se não fosse atingido pelo 13 — mas neste caso o evento da sua morte teria sido diferente, sem as características da morte induzida pelo 13. Portanto, se os eventos forem individuados finamente, a análise proposta por Lewis da dependência causal e da causação pode ser mantida a despeito dos aparentes exemplos de sobredeterminação

explanatória.

A sugestão a ser considerada agora é aplicar a mesma estratégia para salvaguardar (†) dos contra-exemplos apresentados acima. A idéia é individuar finamente fatos explicativos e fatos explicados, de modo que o correspondente contrafactual seja válido. Se o fato  $P$  explica causalmente o fato  $Q$ , então  $Q$  deve ser de tal forma marcado pela causa  $P$  que  $Q$  não poderia ocorrer sem que  $P$  tivesse ocorrido. Compreendido desta forma, (†) estabelece uma condição para a identificação correta dos *relata* da relação de explicação causal. É possível questionar esta condição?

Sim, é possível. Considere mais uma vez o argumento que levava à refutação de (6'),  $\neg Pc \rightarrow \neg M^*e$ . Éramos convidados a assumir que o fato mental  $Mc$  explica causalmente o fato mental  $M^*e$  e, ao mesmo tempo, a supor contrafactualmente que  $c$  não seja do tipo  $P$ . Argumentava-se então que, neste caso, algum outro fato físico  $P'c$  poderia ter realizado  $Mc$ , que continuaria assim a causar  $M^*e$ . Note que a estrutura deste apelo à realizabilidade múltipla é muito semelhante aos casos descritos de sobredeterminação explanatória potencial. Num caso, os realizadores alternativos estão prontos para assumir a parte do realizador atual na produção do efeito em circunstâncias contrafactuais; no outro, as causas alternativas estão prontas para ocupar o lugar da causa atual e assim garantir o efeito atual. No último caso, a sugestão era negar que seja ainda o mesmo efeito produzido pela causa contrafactual. No primeiro caso, não há, agora, mais razão para resistir à sugestão segundo a qual o fato mental realizado atualmente por  $Pc$  não é o mesmo fato realizado contrafactualmente por  $P'c$ . Mas com isso o argumento contra (6') falha e somos destituídos de razões para aceitar a premissa (2) do argumento simples de Marras:  $Mc$  e  $Pc$  não diferem a propósito de sua relevância causal.

Por que é tão difícil negar que o fato mental realizado por um fato físico seja diferente do fato mental realizado por outro fato físico? Lembre-se que ao aceitar (†) comprometemo-nos com uma individuação muito fina dos fatos e eventos particulares. Fatos neste sentido são tão finamente individuados que é impossível que um fato seja causado de diferentes formas. Mas se diferentes causas potenciais são suficientes para separar fatos, então diferentes realizadores potenciais

também o são. Portanto, a mesma estratégia que torna inofensivos os contra-exemplos a (†), presta o mesmo serviço — ou desserviço — ao argumento contra (6').

Resumindo, ou bem aceitamos (†) juntamente com a compreensão requerida de fatos e eventos. É neste caso a premissa (2) do argumento simples de Marras não se sustenta. Ou bem devemos considerar os casos de sobredeterminação explanatória como contra-exemplos a (†). Neste caso, o culpado óbvio é a condição de relevância causal RC. Esta opção solapa a premissa (1) do argumento de Marras. De qualquer forma a conclusão desejada não pode ser mantida e assim o princípio da exclusão explanatória permanece não-refutado.<sup>12</sup>

## Bibliografia

- Baker, L. Rudder (1995). *Explaining Attitudes*. Cambridge: Cambridge University Press.
- Burge, T. (1993) "Mind-body causation and explanatory practice," in J. Heil and A. Mele (orgs.) *Mental Causation*. Oxford: Oxford University Press.
- Fodor, J. (1989) "Making mind matter more." *Philosophical Topics* 17 51–75.
- Jackson, F e P. Pettit (1990) "Program explanation: a general perspective." *Analysis* 50: 107–17.
- Kim, J. (1989) "Mechanism, purpose, and explanatory exclusion." *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*, 77–108.
- . (1990) "Explanatory exclusion and the problem of mental causation," in E. Villanueva (org.) *Information, Semantics, and Epistemology*. Oxford: Blackwell, 36–56.
- . (1998) *Mind in a Physical World. An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, Mass.: MIT Press.
- Lewis, D. (1986) "Postscript to 'Causation'," in *Philosophical Papers*, Vol. II. Oxford: Oxford University Press, 172–213.
- Marras, A. (1998) "Kim's principle of explanatory exclusion." *Australasian Journal of Philosophy* 76: 439–51.
- . (2000) "Critical Notice of Kim (1998)." *Canadian Journal of Philosophy* 30: 137–60.

## Keywords

Causation, dependence, explanation, independence, mind-body problem, relevance; J. Kim, A. Marras

André Fuhrmann  
 FG Philosophie  
 Universität Konstanz  
 784534 Konstanz, Germany  
 andre.fuhrmann@uni-konstanz.de

Wilson P. Mendonça  
 Universidade Federal do Rio de Janeiro  
 Largo de São Francisco de Paula, 1  
 20051-070 Rio de Janeiro, RJ, Brasil  
 mendonca@ifcs.ufjf.br

## Notas

<sup>1</sup> Cf. p.ex. Baker (1995), Burge (1993) e Fodor (1989).

<sup>2</sup> Cf. Kim (1998, 61): O problema da causalção mental “é o problema de mostrar como a causalção mental é possível, não se ela é possível, embora, por certo, o que ocorre com a questão-como possa induzir-nos a reconsiderar nossa atitude a propósito da questão-se.

<sup>3</sup> Para uma versão particularmente cuidadosa desta posição, cf. Jackson e Pettit (1990).

<sup>4</sup> Cf. especialmente Kim (1990) e (1998).

<sup>5</sup> Sobre o papel da suposição de que a causalção mental e a física são “homogêneas,” v. principalmente Crane (1995).

<sup>6</sup> A formalização da explicação causal proposta aqui deve ser neutra relativamente a uma questão central em teorias da causalção. Esta é a questão de saber se (e em que medida) leis naturais fazem parte do conceito de causalção. A concepção “clássica” vincula conceitualmente causalção e leis. Mas há pelo menos uma concepção alternativa que não requer leis: a teoria contrafactual da causalção. A análise proposta neste ensaio é compatível com ambos as concepções. Os defensores do ponto de vista “clássico” gostariam talvez de incluir no *explanans* algum termo que fizesse referência direta a leis, mas isso não deve alterar os resultados obtidos aqui com a fórmula mais simples Cc.

<sup>7</sup> O termo “constituição” foi retirado por nós de Crane (1995, 212f).

<sup>8</sup> É importante notar que (A) formula apenas uma *condição de suficiência* para a dependência (lógica) entre *explanantia* em termos da dependência (metafísica) entre eventos-causas. Um critério completo de (in)dependência dos *explanantia* — que deveria especificar também condições de necessidade — não poderia deixar de levar em conta o fato de que relações entre *explanantia*, ao contrário de relações entre as causas descritas nos *explanantia*, são sensíveis às possíveis *descrições*. Felizmente, as considerações deste ensaio podem prescindir de um critério completo de (in)dependência dos *explanantia*. (Esta nota reflete as observações pertinentes de um “referee” anônimo.)

<sup>9</sup> Como *Ee* e *Fe* se referem a um único evento particular sob diferentes tipos, julgamos que Marras está certo ao afirmar que esta não é a “estratégia do *explanandum* duplo” contra a qual é eficaz o argumento de Kim (1990) em resposta a Dretske (1988). De fato, julgamos que a proposta de Marras só superficialmente exemplifica a estratégia do *explanandum* duplo — no fundo ela não exemplifica esta estratégia. Pois todos os supostos contra-exemplos a (1) têm uma estrutura “triangular,” não “retangular”: dois *explanantia* distintos apontando para um único evento cuja ocorrência deve ser explicada.

<sup>10</sup> Marras considera, porém, que Kim defende uma versão mais forte do realismo explanatório, isto é, RE reforçado num bicondicional. Isto, como pensamos, é um equívoco. De qualquer forma, como se argumentou acima, a versão mais forte do realismo explanatório não é necessária para o argumento de Kim a favor da exclusão explanatória.

<sup>11</sup> O fato de Marta tem uma reputação imaculada é, em primeiro lugar, uma *razão* para o comportamento de Paulo. A explicação psicológico-causal correspondente bem como a implicação negativa do exemplo para o princípio (†) dependem ainda da especificação adicional de um evento mental associado à razão para agir desta forma, por exemplo, a ocorrência do pensamento cujo conteúdo é “Marta tem uma reputação imaculada” na vida mental de Paulo. Esta especificação adicional não nos parece problemática, permanecendo, portando, suposta. O mesmo vale, *mutatis mutandis*, para o próximo exemplo.

<sup>12</sup> Gostaríamos de agradecer ao CNPq (Brasil) e ao DAAD (Alemanha) pelo apoio à pesquisa conjunta que resultou no presente trabalho. A pesquisa de A. F. também foi apoiada por uma bolsa Heisenberg da DFG (Alemanha). Tiramos proveito dos comentários críticos de vários colegas. Entre os que gentilmente nos ajudaram desta forma, gostaríamos de mencionar especialmente Maria Clara Dias, Michael Esfeld, Manfred Kupfer, Ulysses Pinheiro, Jacob Rosenthal, Holger Sturm e Max Urchs.