

## 4

### Entendendo como funcionam as sanções

No capítulo anterior, vimos que as sanções são utilizadas como incentivos para o cumprimento de regras e que experimentos econômicos comprovam certa eficiência dessa estratégia em garantir cooperação social. Em alguns casos, no entanto, a introdução de uma sanção pode causar um efeito indesejado e gerar uma fuga maciça em relação à norma. O problema é, portanto, como projetar sanções que evitem esse transtorno.

Para tentar fornecer respostas que auxiliem a superar esse obstáculo, é necessário entender como as pessoas comuns lidam com as sanções no dia a dia, ou seja, como elas mesmas enxergam o ato de se punir o outro, e esse será o foco desse capítulo. Serão abordadas as justificativas filosóficas mais tradicionais para se punir alguém, e, que, supostamente, servem como base para o molde das sanções nas normas jurídicas, depois disso, serão indicadas descrições de estudos psicológicos sobre como de fato as pessoas punem.

Ainda que os estudos psicológicos converjam e indiquem resultados expressivos, a conclusão que deverá ser extraída é que o ato de se punir alguém é mais complexo, mais emotivo e menos racional (e repleto de assimetrias) do que podemos imaginar. Ainda assim, o trabalho de aprender e juntar essas diferentes informações já disponíveis vale a pena, uma vez que ele pode nos ajudar a melhorar a produção normativa e, mais do que isso, pode nos fornecer diversas peças do quebra-cabeça sobre como sanções podem afetar a compreensão das pessoas em relação às normas.

## 4.1 Contribuições filosóficas: O debate normativo

Como já foi dito, e visto em alguns experimentos, no capítulo anterior, a introdução de sanções pode trazer boas consequências. Mas, nem todos defendem as sanções em termos de resultados produzidos. Os retributivistas entendem que criminosos devem receber o que merecem por seus crimes<sup>43</sup>. A punição, para Kant, é um imperativo categórico, ou seja, a sua imposição independe de seus efeitos serem bons ou ruins, é uma exigência moral derivada dos primeiros princípios ([1798] 2011: 32). Kant diz ([1887] 2010: 198) que, mesmo que a sociedade civil resolvesse se dissolver com o consentimento de todos os seus membros, como no caso de habitantes de uma ilha deserta que resolvem se separar e se espalhar pelo mundo, o último assassino na cadeia deveria ser executado antes que a resolução continuasse. A punição não deve servir para promover um bem para o criminoso ou para a sociedade civil. A ideia do autor é de que a punição deve ser dada ao criminoso apenas por ele ter cometido um crime. As considerações sobre as consequências da punição não interessam a Kant, uma vez que não se deve utilizar uma pessoa como um meio para o propósito de outrem (Kant, [1978] 2011: 31-32).

A ideia de Kant era a de que as pessoas devem ser tratadas da mesma maneira como trataram seus semelhantes<sup>44</sup>. Para o autor, o criminoso não só deve ser punido, como a punição deve guardar proporcionalidade com a ofensa cometida (Kant, [1798] 2011: 32). O retributivismo, portanto, é voltado para o

---

<sup>43</sup> Existem diferentes teorias sobre qual é o débito do ofensor e qual a forma punitiva que fará com que ele tenha a oportunidade de quitar seu débito (Zimmerman, 2011, p. 64). Vão aparecer variadas formas de retributivismo conforme as respostas dessas indagações forem dadas. A ideia aqui é apenas a de agrupar os diferentes retributivistas em posições centrais comuns a eles todos, de forma a não comprometer a posição de cada um. Provavelmente o rótulo “retributivismo” sofrerá certa distorção, algo comum no emprego de reduções. Essa generalização, embora gere certo desconforto pela falta de rigor com os conceitos filosóficos, permite mais facilmente a análise do trabalho psicológico que será descrito em seguida.

<sup>44</sup> Embora Kant tenha chegado a dizer que “apenas a Lei da retribuição (jus talionis) pode determinar exatamente o tipo e o grau da punição” (Kant, [1798] 2011, p. 32), essa posição é bastante polêmica entre os kantianos. Em primeiro lugar, o próprio Kant dá exemplos de punições que não devem ser literalmente iguais às ofensas, como o caso de “injúrias verbais”, por exemplo. Outro ponto é que tal postura retributivista extrema colocaria em cheque a própria posição deontológica do autor. Sem querer entrar em uma análise mais aprofundada do tema, vale esclarecer que a posição aqui esboçada é apenas uma redução das ideias do autor, numa tentativa de se descrever a posição que melhor se ajusta segundo o que o próprio Kant escreveu sobre o assunto. Para os propósitos do presente trabalho, tal descrição é útil e não interfere nas conclusões a serem extraídas.

passado, ou seja, o que uma pessoa merece depende especificamente de suas ações passadas.

A corrente utilitarista se diferencia do retributivismo ao pensar para o futuro, nos efeitos que a punição vai gerar. Essa posição remonta às ideias de Bentham ([1822] 2011), cujo entendimento era de que o objeto do direito penal é aumentar a felicidade total da comunidade e excluir, o quanto possível, tudo que tivesse a tendência de diminuir essa felicidade (Bentham, [1822] 2011: 51-52). O objetivo é, por vezes, expresso como “atingir o maior bem para o maior número de pessoas” (Tonry, 2011: 11). Entretanto, para Bentham, a felicidade de todos inclui a felicidade daqueles causadores de ofensas. O autor percebe que toda punição é ruim, representa um prejuízo moral, mas, sob a luz do princípio da utilidade, se ela deve ser admitida, deve apenas enquanto pretender excluir um mal maior (Bentham, [1822] 2011: 63).

A punição, segundo essa visão, deve ser esperada como tendo um peso maior do que os benefícios do crime. A ideia parece ser a de prevenir ou, ao menos, de reduzir a incidência de tais comportamentos indesejados. Para responder, em linhas gerais, de que forma a punição supostamente deve servir a esse propósito, ao menos três maneiras aparecem: há aqueles que defendem a *reabilitação*, ou seja, as punições devem reduzir o crime de uma maneira com que o punido não se disponha mais a se engajar em uma conduta criminosa; alguns entendem em favor da *incapacitação*, em que as punições podem servir para a redução do crime ao tornarem os punidos incapazes de cometerem novamente o crime que cometeram; outros acreditam na *dissuasão*, e para eles as punições devem deter aqueles que de outra maneira iriam cometer crimes, servindo como um exemplo. Embora guardem diferenças entre si, como visto, pode-se agrupar e resumir as ideias das vertentes do utilitarismo da seguinte maneira: o objetivo da punição é aumentar o bem estar do maior número de pessoas; a punição só deve ser admitida quando representar a exclusão de um mal maior para a sociedade; a punição deve ser moldada de forma a tentar prevenir ou reduzir a incidência de crimes.

Para além da indicação normativa do princípio da utilidade como o guia para o desenho das punições, Bentham faz ainda uma descrição sobre como as pessoas agem ([1822] 2011: 53). O autor aposta que o princípio da utilidade é algo que a maioria das pessoas usa na maior parte do tempo e sem nem pensar

sobre isso. Ou seja, o autor acredita que as intuições das pessoas sejam utilitaristas a maior parte do tempo. Entretanto, aqui Bentham faz uma proposta descritiva, e não mais normativa sobre a psicologia da punição.

Bentham não está sozinho na empreitada de descrever como as pessoas punem: Edward Westermarck, sociólogo e filósofo finlandês, também escreveu sobre o assunto. Westermarck seguiu, no entanto, um caminho diferente de Bentham, sustentando em favor de intuições retributivistas. Mackie resume a posição de Westermarck da seguinte forma:

O filósofo finlandês Edward Westermarck, nos dois primeiros capítulos de *Ethical Relativity*, argumenta contra a suposta objetividade dos julgamentos morais, alegando que os conceitos que são usados como predicados nesses julgamentos são ultimamente baseados em emoções. Nos próximos dois capítulos ele tenta dar substância a essa alegação identificando a classe especial de emoções morais em que esses conceitos são baseados: aprovação moral e desaprovação ou indignação moral. Ele diz que essas são uma subclasse do que ele chama de emoções retributivas: emoção gentilmente retributivista e ressentimento (que também incluem raiva, vingança e gratidão), e são separadas de suas contrapartes não-morais pelo desinteresse e pela, ao menos aparente, imparcialidade. Ele explica essas características diferenciadoras pelo fato de que ‘a sociedade é o berço da consciência moral,’ que ‘o costume tribal foi a mais antiga regra de dever.’ Westermarck também argumenta que a desaprovação é mais importante que a aprovação, e que o ressentimento do qual elas são subespécies é provavelmente mais importante, e ocorre de forma mais ampla no reino animal do que a emoção gentilmente retributivista. Ele mantém que o elemento retributivo na punição não é passível de ser eliminado, e não pode ser totalmente substituído por tais princípios como dissuasão e reabilitação, embora ele também diga que exista ‘alguma conexão entre esses fins e o objetivo retributivo do ressentimento moral’ (Mackie, 2011, p. 271).

Bentham ou Westermarck nunca forneceram, porém, qualquer comprovação científica de suas alegações descritivas. Ambos autores se preocuparam apenas em relatar suas intuições sem qualquer confirmação de que elas eram de fato compartilhadas pelas pessoas comuns. Resta saber quem estava correto em sua afirmação, algo que só pode ser verificado empiricamente.

## 4.2

### **Contribuições psicológicas: O debate descritivo**

Como visto, o debate normativo gira em torno de duas posições principais, o utilitarismo e o retributivismo. No entanto, Bentham, defensor do utilitarismo, vai além da esfera normativa e também faz afirmações empíricas, descritivas

acerca das intuições psicológicas das pessoas sobre punições. Porém, tal tipo de afirmação só pode ser corroborada empiricamente e os estudos psicológicos demonstram que Bentham estava longe de ter razão nesse ponto.

Com a finalidade de descobrir o que as pessoas levam em conta quando atribuem punição, ou seja, que tipo de informação as pessoas recrutam quando lhes é requerido que tomem uma decisão sobre punição, Carlsmith (2006) montou um experimento em que os participantes deveriam agir como juízes: recomendando uma sentença para um crime que lhes fora descrito. Além do relato do fato, os participantes podiam escolher uma informação adicional sobre aquele crime. O experimentador apresentava diferentes categorias de informação, cada uma unicamente relevante para dissuasão, incapacitação ou retribuição. Portanto, por exemplo, os participantes podiam escolher categorias de informações como “frequência do crime na sociedade” (uma categoria relacionada à dissuasão), “possibilidade de repetição” (relacionada à incapacitação), e outras sete possibilidades de informação agrupáveis nas mencionadas teorias. Uma vez escolhida a categoria, o participante ouvia um detalhe particular como “crimes desse tipo são extremamente raros”, ou “peritos informaram que o criminoso tende a repetir esse tipo de crime caso lhe seja dada a oportunidade”. Desse modo, as pessoas poderiam escolher aprender diferentes tipos de informação. A frequência e a ordem de escolha para cada categoria foram as medidas utilizadas pelos experimentadores para entender a preferência das pessoas em relação às perspectivas.

Na primeira de cinco repetições, 97% dos participantes escolheram uma informação retributiva. Nas segunda e terceira repetições, a grande maioria (64% e 57%) também escolheu itens retributivos. Conforme a categoria retributiva ficava sem representantes para a escolha, os participantes passavam a escolher informações da categoria ligada à incapacitação. Apenas quando não havia mais item disponível de outra categoria, é que as pessoas escolhiam informações relacionadas à dissuasão (Carlsmith e Darley, 2008: 203). Esse experimento revela uma certa preferência por informações de caráter retributivista na ocasião da atribuição de punição.

Num experimento posterior (Carlsmith, 2006), foi repetida a dinâmica de se colocar os participantes na posição de juízes e fornecer uma informação compatível com cada perspectiva, conforme já descrito anteriormente. A diferença

consistiu em não deixar os participantes escolherem a informação que deveriam receber, mas que lhes fosse atribuída de forma aleatória. Foi, então, pedido a eles que atribuíssem uma punição e informassem o nível de confiança da correção daquela decisão. Não surpreendentemente, as pessoas relataram maior índice de confiança conforme o aumento de informação que recebiam. Cada informação relacionada à retribuição, porém, apresentava um acréscimo significativamente maior de confiança do que as informações relacionadas tanto à dissuasão quanto à incapacitação. As pessoas, portanto, escolhem informações retributivistas preferencialmente e se sentem mais confiantes de estarem fazendo justiça quando têm acesso a essas informações.

Em outro grupo de experimentos realizado por Carlsmith, Darley e Robinson (2002), foi inicialmente pedido a 336 estudantes universitários que lessem uma vinheta sobre uma ofensa criminal e que recomendassem uma sentença para o criminoso. Foram feitas três rodadas envolvendo manipulações de diferentes fatores de diferentes maneiras em cada uma delas. Ao longo das rodadas, portanto, alterou-se a seriedade da ofensa e a presença de circunstâncias morais como forma de se manipular o fator retributivo; além disso, também foi modificada a dificuldade de se identificar o crime e a publicidade da sentença como forma de se manipular o fator relacionado à dissuasão. Independentemente da forma como foram feitas as manipulações, as pessoas foram altamente sensíveis às mudanças em fatores retributivos, atribuindo penas mais altas conforme o merecimento e a seriedade do crime e ignoraram fatores relacionados à perspectiva utilitarista (não havendo qualquer alteração na atribuição de pena nos casos em que houve manipulação para os fatores de dissuasão). A severidade da punição foi determinada exclusivamente pela manipulação das informações relevantes para o retributivismo.

Nesse estudo, também foram posteriormente medidas as diferenças individuais para saber se havia diferentes tipos de pessoas: aquelas que punem por retribuição e aquelas que punem seguindo um modelo utilitarista. Foi explicado em linhas básicas para os participantes cada teoria e demandado que eles escolhessem um modelo que melhor descrevesse a perspectiva pessoal em relação à punição. Foi descoberto que as pessoas variavam muito em relação a essa questão, e que era fácil dividi-las entre três grupos: retributivistas, utilitaristas e os que possuíam perspectivas mistas. No entanto, o comportamento de fato das

peças quase não se diferenciou na hora de punirem (Carlsmith e Darley, 2008: 206). Na prática, todos se mostraram retributivistas.

O procedimento seguinte foi o de se repetir o que havia sido feito no experimento anterior com ligeiras alterações nos casos. Os resultados foram replicados e houve a inclusão de uma segunda etapa. Nessa parte, 96 de 229 participantes receberam instruções sobre o que seria o retributivismo e o dissuasão e foi lhes pedido para dar por mais duas vezes sentenças para os mesmos casos; só que uma vez pensando como um retributivista e, na outra, como um dissuasão. O resultado indicou que o padrão de atribuição de pena se manteve quando os participantes tiveram que agir como retributivistas, mas houve um aumento nas penas quando puniram seguindo uma postura dissuasão (Carlsmith et al., 2002: 292). Isso significa que elas estavam sendo retributivistas mesmo antes de lhes ser designada a tarefa de julgar dessa forma, e que elas entendem a postura dissuasão como algo que faz com que a quantidade de punição deva ser aumentada para cada caso.

Diante dos dados obtidos pelos experimentos expostos, conclui-se que Westermarck tinha razão sobre a postura punitiva das pessoas, ao que parece, somos mesmo retributivistas. Além disso, o autor parece ainda estar correto, como será demonstrado adiante, em outra de suas afirmações: as punições retributivas são respostas emotivas, e não puramente racionais.

Embora Kant, talvez o autor mais famoso a sustentar o retributivismo no campo das punições – representante da posição deontológica<sup>45</sup> na ética –, fosse ele próprio um racionalista, estudos de ressonância magnética funcional e estudos de lesões cerebrais sugerem que a deontologia é mais emotiva do que o utilitarismo (Damasio, 1994; Greene et al., 2001; Greene et al., 2004; Koenigs et al., 2007). Em outras palavras, as respostas deontológicas observadas em experimentos têm sido relacionadas mais às emoções do que a um processo cognitivo racional<sup>46</sup>. A deontologia parece servir mais como um exercício de

---

<sup>45</sup> Em breves palavras, A deontologia é geralmente sustentada como uma posição que dá ênfase às normas morais, geralmente expressas em termos de *direitos* e *deveres*. Essa corrente é normalmente contrastada em relação ao consequencialismo, que é uma visão que indica o valor moral de uma ação unicamente com base em suas consequências. Para um deontologista, diferente de um consequencialista, matar uma pessoa para salvar várias outras, por exemplo é errado, mesmo que isso maximize as boas consequências (Greene, 2007, p. 37).

<sup>46</sup> Para efeitos de esclarecimento, seguimos aqui a diferenciação de Greene entre representações cognitivas e representações emotivas, que define que: “A ideia geral é que representações

racionalização moral *post hoc* do que como um raciocínio em si (Greene, 2007, p. 36).

Para chegar a essa conclusão, Greene, por exemplo, estudou como o cérebro das pessoas funcionava enquanto tomavam decisões em dilemas morais hipotéticos. O autor, interessado inteiramente em descrever como as pessoas tomam decisões morais, apresentou versões de famosos dilemas filosóficos, como o do *bondinho desgovernado* e o da *passarela* (mais conhecido como o dilema do *homem gordo na ponte*) e observou as áreas cerebrais que se iluminavam na máquina de ressonância magnética e a que tipo de função essas áreas são comumente associadas.

Os mencionados dilemas são conhecidos por capturarem tensões entre consequencialistas e deontologistas. No primeiro dilema, há um bondinho desgovernado indo em direção a cinco pessoas que serão mortas, caso o bondinho continue a seguir o seu curso. A única maneira de salvar essas pessoas é puxando uma alavanca que mudará o trajeto do bondinho, colocando-o em uma pista alternativa, na qual há uma pessoa que será morta, caso o bondinho tome esse rumo. A pergunta que se apresenta é se é moralmente permitido puxar a alavanca e matar uma pessoa para salvar cinco. Greene (2007, p. 42) relata que parece haver certo consenso de que é, sim, moralmente permitido tal ato. No segundo dilema, como no primeiro, há um bondinho desgovernado indo em direção a cinco pessoas, no entanto, nesse caso não há uma pista alternativa, o observador da cena está em uma passarela próximo a um desconhecido homem gordo, cuja massa é capaz de parar o bonde. A única maneira de salvar tais pessoas é empurrando o homem gordo da ponte para os trilhos, ocasionando morte de tal homem. A pergunta é se é moralmente permitido empurrar tal homem, matando-o, para salvar as cinco pessoas. Aqui, o consenso parece ser de que não é permitido matar uma vida para salvar as outras cinco (Greene, 2007, p. 42).

As pessoas respondem de forma consequencialista para o primeiro caso e deontologicamente para o segundo e, no entanto, pouco parece ter mudado. Ambos envolvem a morte de uma pessoa inocente para salvar cinco outras que estavam em perigo. Greene e seus colegas sugerem que o pensamento sobre

---

‘cognitivas’ são representações inerentemente neutras, que não acionam automaticamente respostas ou disposições de comportamento particulares, enquanto representações ‘emotivas’ têm tais efeitos automáticos, e que, portanto, têm uma valência comportamental” (2007, p. 40).

empurrar alguém para morte é uma questão mais pessoal e, portanto, mais “emocionalmente saliente” do que pensar sobre puxar uma alavanca que irá causar um bondinho a produzir consequências semelhantes (2001, p. 2106). Os autores vão além e sugerem o seguinte:

Essa hipótese envolvendo esses dois casos sugere um hipótese mais geral sobre o julgamento moral: Alguns dilemas morais (aqueles relevantemente similares ao dilema da *passarela*) engajam processos emocionais em maior extensão do que outros (aqueles relevantemente similares ao dilema do *bondinho desgovernado*), e essas diferenças de engajamento emocional afetam o julgamento das pessoas (2001, p. 2106).

O que está sendo sugerido não é que há um tipo de decisão moral totalmente guiada pelas emoções e outra guiada pelas razões. Greene acredita que, em ambos os casos, há gatilhos emocionais e cognitivos sendo disparados. A diferença para o autor está no tipo de emoção que é essencial para cada um dos julgamentos. A emoção nas decisões consequencialistas funcionam como uma moeda, enquanto nas deontológicas elas trabalham como um alarme (Greene, 2007, p. 41).

Essas diferenças emocionais estão presentes na forma como enxergamos as diferentes formas de dano pessoal e impessoal. No caso do dano pessoal, há o acionamento de uma emoção que funciona como um alarme e produz um comportamento de rejeição à possibilidade de matar uma pessoa<sup>47</sup>, ainda que para salvar outras cinco; enquanto no caso impessoal, essa mesma emoção não é acionada (ou ao menos não com a mesma intensidade), e as pessoas tendem a responder de uma maneira mais cognitiva, fazendo algo como uma análise de custo e benefício.

Confirmando essas hipóteses, o que Greene e seus colegas (2001; 2004) encontraram foi uma atividade maior em áreas cerebrais relacionadas às emoções (no córtex cingulado posterior, no córtex pré-frontal medial, e na amígdala) quando as pessoas contemplavam dilemas que envolviam danos pessoais, como o caso da *passarela*; enquanto quando deviam julgar casos impessoais, como no dilema do *bondinho desgovernado*, as atividades neurais de regiões associadas à cognição

---

<sup>47</sup> “Tendo em vista que a violência pessoal é antiga na escala evolucionária, antecedendo nossas recentes capacidades humanas evoluídas para complexos raciocínios abstratos, não deveria ser surpresa que nós tenhamos respostas inatas poderosas, mas primitivas, à violência pessoal. Isto é, nós podemos esperar que humanos tenham respostas emocionais negativas a certas formas básicas de violência interpessoal, onde essas respostas evoluíram como maneiras de regular o comportamento de criaturas que são capazes de intencionalmente causar danos uns aos outros, mas cuja sobrevivência depende de cooperação e restrição individual” (Greene, 2007, p. 43).

(como o córtex pré-frontal dorsolateral e o lóbulo parietal inferior) foram mais altas.

Outra confirmação das hipóteses foi obtida por meio da observação do tempo de reação das pessoas. Elas têm respostas emotivas para as violações morais pessoais que as levam a julgar contra essas ações, portanto, alguém que julgue apropriadas tais violações pessoais, deve demorar mais tempo para tomar sua decisão, uma vez que terá de derrubar tais respostas emotivas (Greene, 2007, p.44). Alguém que diga “sim” para a questão do dilema da *passarela* (apoando a permissibilidade moral de se empurrar o homem, matando-o, para salvar outras cinco pessoas) deveria demorar mais tempo para dizê-lo do que alguém que respondeu “não”. Para os casos de violações impessoais, o tempo de resposta não deveria se alterar, uma vez que não há um esforço para superar qualquer resposta inicial, a resposta inicial é a própria resposta. Os experimentos (Greene et al. 2004; Greene et al., 2001) de fato confirmaram essa expectativa: Julgamentos em favor de violações morais pessoais levaram significativamente mais tempo do que julgamentos contra eles, mas não houve um efeito no tempo de reação comparável em respostas às violações morais impessoais (Greene, 2007, p. 44).

Essas respostas emotivas automáticas (que funcionam como alarmes, para Greene) são provenientes de um processo intuitivo, que é mais rápido e automático do que o processo racional. Além dessas características – carregados emocionalmente, rápidos e automáticos –, tais processos geram menos esforços, podem correr em paralelo a outros processos mentais e são implícitos, ou seja, não estão disponíveis para uma análise introspectiva (Kahneman, 2003). Sobre o assunto Darley completa da seguinte forma:

Como é bem conhecido, as pessoas experimentam suas percepções como representações simples e corretas “do que há lá fora”; isto é, as pessoas experimentam o mundo perceptual no modo de um realista ingênuo. Intuições, por sua vez, como percepções, são geralmente tomadas por aquele que as intui como corretas de forma não-problemática. Portanto, um conjunto de decisões, escolhas e soluções de problemas produzidos intuitivamente são experimentados como resumos das maneiras que o mundo é, porque a pessoa que tem as intuições está desavisada dos complexos e potencialmente incorretos processos cognitivos que as produzem (2012, p. 3).

Os processos cognitivos racionais, por outro lado, são mais lentos, encadeados, geralmente controlados conscientemente, geram mais esforços e são governados por normas. Algumas vezes esses processos são invocados e a conclusão deles pode derrotar uma resposta ditada pelo sistema intuitivo. É por

esse motivo que, no caso em que a pessoa responde “sim” para a permissibilidade moral de um ato que envolve violação pessoal, há um maior tempo entre a pergunta e a resposta do que nas negativas. Essas ocorrências, porém, como vimos, são mais raras. O mais comum é a resposta rápida (emotiva e intuitiva) contrária a tais atos.

Se as decisões deontológicas são comumente fruto de respostas emotivas automáticas e o retributivismo é uma forma de deontologia, então é provável que Westermarck estivesse certo e as atribuições de punições sejam menos racionais do que se supõe. É provável que, senão a escolha pelo tipo e pelo modo de punição, ao menos a vontade, o *animus*, de se punir e a quantidade de punição que deve ser atribuída a cada um, surjam a partir de respostas emotivas intuitivas, como nos casos de violações pessoais de Greene.

Alguns estudos comprovam essa tese. Os experimentos envolvendo um jogo econômico denominado jogo do ultimato, talvez sejam o melhor exemplo disso. Tal jogo consiste em dar uma quantia de dinheiro para ser dividida entre dois jogadores. Um jogador, chamado *proponente*, tem a tarefa de propor uma divisão entre ele e o segundo jogador. O segundo jogador, denominado *respondente*, pode tanto aceitar a oferta ou rejeitá-la. Caso a aceite, os jogadores recebem o dinheiro da forma que a oferta foi feita, caso rejeite, nenhum dos dois recebe nada. A resposta racional para esse jogo é a que o *respondente* aceita qualquer valor que lhe for ofertado, uma vez que receber qualquer montante é melhor do que ficar sem receber nada. Os estudos apontam que, em sociedades industrializadas<sup>48</sup>, as ofertas geralmente são em torno de 50% do valor total. Contudo, quando ofertas baixas (em torno de 20% do total) são feitas, há 50% de chance de serem rejeitadas (Sanfey et al. 2003). Esse resultado impressiona justamente por indicar que nem sempre a razão prevalece, mesmo quando há dinheiro envolvido.

A explicação para isso é que experimentamos um sentimento de raiva quando outros não cumprem com nossas expectativas sobre o que é o justo a ser feito (Denson e Fabiansson, 2011). Estudos de ressonância magnética funcional comprovam essa afirmação mostrando que é a ínsula (a região cerebral responsável pela raiva e pelo nojo) a área cerebral mais determinante no jogo do

---

<sup>48</sup> Experimentos utilizando o mesmo tipo de jogo obtiveram resultados diferentes quando feitos em sociedades mais simples, como as comunidades Mapuche, Machiguenga e Huinca. Para uma visão geral, ver Henrich et al, 2009.

ultimato: ela é mais ativada quando *orespondente* recusado que quando ele aceita a oferta e mais fraca quando o *respondentes* sabe que está jogando com um computador e não com um humano (Sanfey et al., 2003).

Ainda no campo da neurociência, há outro estudo que chega a conclusão similar (de Quervain et al., 2004). Nesse estudo, envolvendo um jogo econômico, foi avaliada a punição em relação à quebra de confiança. Os autores concluíram que o nível de punição era correlacionado com o nível de atividade do núcleo caudado, uma região cerebral associada às emoções, mais especificamente, aquelas relacionadas à motivação e à recompensa.

Além disso, as pessoas parecem punir na proporção da sua raiva. Kahneman e seus colegas (Kahneman et al., 1998) realizaram um experimento em que os participantes responderam a alguns cenários hipotéticos similares (por exemplo, um caso em que um airbag automático se abre inesperadamente e machuca o motorista). Os participantes deveriam avaliar o quão “ultrajante” a ação do réu havia sido. Eles também deveriam avaliar o quanto o réu, no caso em questão, deveria ser punido. Os resultados indicaram que a correlação entre os índices de ultraje e punição atribuídos pelos participantes foi quase perfeita, atingindo 0,98 no coeficiente de correlação de Pearson ( $r$ ), em que o valor 1 indica uma correlação perfeita (Greene, 2007, p. 52). A conclusão dos autores foi que as pessoas desejam ver um infrator punido por seu comportamento quase que exclusivamente conforme o quão emocionalmente ultrajados elas ficam pelo comportamento desse infrator.

Os experimentos até aqui relatados demonstram ainda mais um dado curioso: esse sentimento de ultraje moral, ou de “raiva”, leva as pessoas não só a desejar punir quem faz algo de ruim a elas mesmas, mas a quem comete um dano, ou violação, a terceiros inocentes. Como já foi dito, trata-se de um sentimento relacionado à noção de justiça, que de certa forma foi violada pelo ato em questão. Esse tipo de conduta, de se punir por violações a terceiros, tem importância muito grande para a nossa sociedade, uma vez que permite a efetividade das regras morais e garante maior cooperação dos indivíduos em relação à comunidade em que vivem (Fehr e Fischbacher, 2004). Sobre o assunto, Miller diz que:

O estímulo para a raiva moralista não está preso às injustiças perpetradas contra a própria pessoa. Testemunhar a violação de terceiros também pode estimular esses sentimentos fortes de raiva e injustiça. (...) Os indivíduos estão comprometidos com as “forças dos deveres” morais de sua comunidade (...), e as pessoas

acreditam que essas forças merecem respeito de todos os membros da comunidade. A violação dessas forças representa um insulto à integridade da comunidade e provoca em seus membros tanto a raiva moralista quanto a urgência em punir o ofensor. Vista dessa perspectiva, as reações de justiça desinteressadas não são nem um pouco desinteressadas, porque todos têm um interesse em ver respeitadas as regras e valores da estrutura de autoridade sob as quais eles vivem (Miller, 2001, p. 535).

Diante do exposto, conclui-se que a nossa característica de punir de forma retributivista tem origem nas emoções de ultraje, de raiva, de injustiça, que sentimos em relação àqueles cuja conduta viola certas normas morais que internalizamos. Temos acesso a tais normas e conseguimos justificar nossa conduta apelando a elas. O que não desconfiamos, no entanto, é que essa é só uma racionalização posterior à resposta punitiva proveniente de um processo intuitivo e emocional.

#### **4.2.1 A assimetria entre projeção e julgamento**

Apesar dos resultados dos experimentos mencionados indicarem uma preferência pelo retributivismo no momento do julgamento de casos concretos, é interessante notar que alguns experimentos mostram que, por vezes, as pessoas têm uma perspectiva diferente quando estão legislando, ou seja, trabalhando em abstrato. Carlsmith e Darley (2008) relatam, por exemplo, que algumas pessoas tendem a preferir, quando estão projetando para o futuro, normas com sanções utilitaristas, em detrimento daquelas com sanções retributivistas, ou seja, punições justificadas com o objetivo de se evitar novas condutas, ao invés de punições proporcionais ao ato em si. No entanto, quando devem aplicar essas mesmas normas criadas por elas, as pessoas tendem a buscar motivos retributivistas, e não mais endossam o texto que haviam aceitado.

Para comprovar isso, Carlsmith et al. (2002) realizaram outro experimento. Foi dada uma tarefa de alocação para os participantes. Eles deveriam sugerir que fossem alocados os recursos da cidade para capturar criminosos (uma ação consistente com uma perspectiva retributivista) ou para prevenir crimes futuros desse mesmo tipo (uma ação relacionada ao utilitarismo). Foi descoberto que as pessoas demonstram uma grande preocupação em prevenir futuras ocorrências do

crime, mas que esse desejo não está relacionado com os detalhes do crime. Os participantes recomendaram maior alocação em prevenção de futuros crimes do que em capturar criminosos, demonstrando estarem de acordo com uma posição utilitarista. Portanto, apesar de endossarem os objetivos ligados a teorias utilitaristas (prevenir novos crimes) enquanto projetam as normas, quando na posição de julgador, as pessoas não punem de forma a concretizá-los. Afinal, conforme visto nos outros experimentos, ficou demonstrado que elas se utilizam mais frequentemente de critérios retributivistas e sentem-se mais confortáveis dessa forma (Carlsmith et al., 2002: 295).

Um exemplo real dessa discrepância em normas jurídicas é o da lei de “3 strikes” da Califórnia. Tratava-se de lei que exigia que o governo sentenciasse à prisão perpétua aqueles que fossem reincidentes após dois ou mais “crimes sérios” (*felonies*). Tal norma foi aceita e promulgada por 70% da população em referendo estadual. O que os cidadãos da Califórnia não esperavam é que essa norma seria incompatível com sua perspectiva retributivista. Quando em vigor, alguns casos bizarros, em que o último “crime sério” era de menor potencial lesivo (*minor felony*), começaram a aparecer. Jerry Dewayne Williams foi condenado a prisão por 25 anos à perpétua pelo furto de um pedaço extra-grande de pizza sabor pepperoni<sup>49</sup>. Santos Reyes foi condenado a 26 anos à perpétua no terceiro *strike* por perjúrio, uma vez que ele fez, no lugar de seu primo, que sabia dirigir, mas não sabia ler, a parte escrita do teste para a licença de motorista, para ajudá-lo a passar na prova<sup>50</sup>. Kevin Weber pegou 25 anos à perpétua por invadir um restaurante e furtar cookies de chocolate<sup>51</sup>. Após ver essa lei em ação e esses resultados aparecerem, em poucos anos as pesquisas (Carlsmith e Darley, 2008: 208) mostraram que o apoio à lei já era menor do que 50% da população.

Um experimento de Carlsmith (2008) tentou repetir esse fenômeno. Foi descrito aos participantes o problema que as escolas têm tido com estudantes usuários de drogas, e apresentadas diferentes políticas a serem adotadas para o problema. Apesar de a maioria ter preferido respostas proporcionais às ofensas, numa perspectiva retributivista, uma minoria substancial (30%) apoiou uma política de tolerância zero, baseada na teoria utilitarista. Posteriormente foi

---

<sup>49</sup> <http://articles.latimes.com/2010/feb/10/local/la-me-pizzathief10-2010feb10>

<sup>50</sup> [http://socialistworker.org/2005-1/542/542\\_05\\_SantosReyes.shtml](http://socialistworker.org/2005-1/542/542_05_SantosReyes.shtml)

<sup>51</sup> <http://edition.cnn.com/US/Newsbriefs/9510/10-27/index.html>

descrito um caso em que uma menina era encontrada no banheiro com remédios não prescritos, considerados como drogas para os efeitos legais, e questionado se as pessoas mantinham o que haviam entendido ser a política correta. Quando os participantes que apoiaram a política da tolerância zero descobriram que essa violava a proporcionalidade, e, portanto, seus sentidos intuitivos de igualdade, o apoio a essas políticas caiu para apenas 6%.

Esse estudo concluiu que “as pessoas têm uma introspecção apenas limitada às razões para suas punições, e essa ignorância pode levá-las a endossar políticas que, ironicamente, elas irão logo rejeitar como desiguais e injustas” (Carlsmith e Darley, 2008: 208). A conclusão dos autores é de que as pessoas não têm completa consciência sobre as razões que as levam a punir outras. Elas são retributivistas, mas talvez não o percebam. Quando estão pensando em punições futuras, não têm diante de si uma transgressão já existente e, portanto, seus julgamentos reais não são acionados.

As avaliações em casos concretos são mais retributivas porque esse tipo de situação desperta mais as emoções, ou a emoção, de ultraje moral. Em verdade, a atribuição de punição pensada para o futuro, como na projeção de uma norma, funciona como um daqueles casos de violações morais impessoais. Quem está modelando não está tendo um “contato direto” com o punido, não está “sujando suas mãos”, bem como tem diante dele uma violação já ocorrida, ele deve apenas imaginá-la. Diferente dos casos concretos, não há o acionamento de emoções que funcionem como alarmes, e, portanto, o caminho cognitivo fica livre para uma análise de custo e benefício, típica do consequencialismo.

Assim, a assimetria entre nossa psicologia retributivista de julgamento e utilitarista de projeção indica um dos motivos pelos quais podemos ter a sensação de que uma norma é injusta. Se temos distintas noções sobre punição justa conforme a posição que ocupamos (juiz ou legislador), fica difícil dar uma resposta teórica que satisfaça a todos os casos.

#### **4.2.2**

#### **A assimetria na atribuição de intencionalidade**

Acima expusemos diversos experimentos que lidos em conjunto indicam que a psicologia descritiva das pessoas no momento de julgar a ação de outra e de resolver se uma punição deve ser atribuída é retributivista. A atribuição costuma seguir um contínuo, conforme Cushman (2012), formado pelo seguinte encaminhamento de ideias: quando algo de ruim acontece, procuramos saber se alguém foi o agente causal do resultado ruim ou do dano produzido, ; em caso positivo (se a causa é um agente humano, e não a natureza), procuramos saber sobre o estado mental do agente no momento da ação, isto é, se ele tinha a intenção de realizar o ato sob discussão ou produzir os resultados da ação; por último, formamos um julgamento atributivo de punição proporcional ao mal causado e ao grau de culpabilidade do estado mental. Dessa maneira, o retributivismo e o estado mental de quem comete a infração estão conectados. O retributivismo sustenta que a punição é baseada no merecimento do agente infrator. Isto é, ele merece ser punido exatamente por ter violado alguma norma moral. Parece, portanto, que o principal elemento constitutivo desse merecimento é o estado mental do agente, que, por sua vez, tem como seu principal elemento constitutivo a intencionalidade.

Sendo assim, cabe discutir, também, alguns achados intrigantes sobre como atribuímos intencionalidade. O filósofo Joshua Knobe, um dos expoentes do novo movimento conhecido como “filosofia experimental”, constatou assimetrias interessantes na forma como atribuímos intencionalidade para certos atos. Em um experimento que ganhou muita repercussão nos âmbitos da psicologia, da filosofia moral e da filosofia da ação, Knobe mostrou que a nossa atribuição de intencionalidade para ações que produzem certos efeitos colaterais, depende de nossa avaliação moral prévia dos resultados da ação. São nossas considerações morais sobre os efeitos colaterais que determinam nossa atribuição de intencionalidade. Vejamos os cenários do experimento que foi utilizado para se chegar a tal conclusão (Knobe, 2003):

(A) O vice-presidente de uma empresa foi falar com o presidente do conselho e disse: “Estamos pensando em começar um novo programa. Ele nos ajudará a aumentar os lucros, mas vai também *prejudicar* o meio ambiente”.

O presidente respondeu: “Eu não me importo nem um pouco com os prejuízos ao meio ambiente. Apenas quero lucrar o máximo que for possível. Vamos iniciar o novo programa”.

Eles começaram o novo programa. Evidentemente, o meio ambiente foi *prejudicado*.

(B) O vice-presidente de uma empresa foi falar com o presidente do conselho e disse: “Estamos pensando em começar um novo programa. Ele nos ajudará a aumentar os lucros, mas vai também *beneficiar* o meio ambiente”.

O presidente respondeu: “Eu não me importo nem um pouco com os benefícios ao meio ambiente. Apenas quero lucrar o máximo que for possível. Vamos iniciar o novo programa”.

Eles começaram o novo programa. Evidentemente, o meio ambiente foi *beneficiado*.

Depois de apresentar a primeira vinheta para um grupo de pessoas, Knobe perguntou para elas o quão culpado era o presidente (em uma escala de 0 a 6) por ter causado o dano ao meio ambiente, e se o presidente teria causado o dano intencionalmente (a pessoa deveria indicar sim ou não como resposta). Nesse caso, 82% dos participantes responderam que o presidente havia prejudicado o meio ambiente intencionalmente. Por outro lado, apenas 23% das pessoas que receberam a segunda vinheta responderam que o presidente teria beneficiado o meio ambiente intencionalmente. O curioso, como se pode notar ao ler os diferentes cenários, foi que o presidente da empresa disse a mesma coisa em ambas as vinhetas, isto é, que não se importava nem um pouco com o meio ambiente, mas apenas com os lucros da empresa. A conclusão é a de que no caso em que o efeito colateral é considerado moralmente ruim, as pessoas costumam atribuir intencionalidade, enquanto nos casos em que o efeito colateral é considerado moralmente adequado, as pessoas não costumam atribuir intencionalidade.

A assimetria acima ficou conhecida como “efeito do efeito colateral” (*side-effect effect*) ou “efeito Knobe” (*Knobe effect*) pelos estudiosos que começaram a escrever sobre o assunto após a descoberta de Joshua Knobe. A assimetria chama a atenção porque normalmente pensamos que os nossos julgamentos morais adotam um percurso unidirecional, no qual primeiro acessamos a intencionalidade de alguém, para depois decidir sobre sua responsabilidade moral. Achamos que a noção de intencionalidade é sempre moralmente neutra, que se trata de um conceito de nossa psicologia do senso comum e que sempre funciona como um precursor dos juízos morais: se a ação é intencional, então, podemos, posteriormente, fazer uma avaliação genuinamente moral. Contudo, os estudos de Knobe demonstram que o caminho entre intencionalidade e julgamento moral é, na verdade, bidirecional. Muitas vezes, analisamos a intencionalidade avaliando estritamente o estado mental das pessoas, mas, em outros casos, atribuímos

intencionalidade justamente depois que fazemos uma avaliação moral dos resultados. Se os resultados desagradam do ponto de vista moral, então atribuímos intencionalidade (Knobe, 2009)<sup>52</sup>.

Outros estudos confirmam essa proposição, como descreve Nadelhoffer (2011). Em um experimento (Alicke et al., 1994), por exemplo, foi descrito aos participantes do estudo que um proprietário de uma casa atirou em alguém, que ele presumia ser um intruso, em um quarto na parte de cima de sua residência. Para uma parte dos participantes, foi dito que a vítima era, na verdade, um criminoso violento, inclusive responsável por roubos na vizinhança; para a outra parte, a vítima foi descrita como o namorado da filha do proprietário. Foi então pedido que fosse estimada a relevância causal de uma série de fatores, dentre eles, por exemplo, o fato de o proprietário ter consumido duas cervejas antes de efetuar os disparos. Note-se que a diferença entre os casos consistia apenas na qualidade do invasor, que era desconhecida pelo proprietário. Os participantes, no entanto, que trabalharam com a hipótese do invasor como namorado da filha consideraram a cerveja como tendo um papel causal mais importante para a efetuação dos disparos do que aqueles que receberam a vinheta com o invasor sendo um criminoso violento.

Em outro experimento (Alicke, 1992) foi apresentada uma vinheta que contava a história de um atropelamento causado por um homem, que estava dirigindo a uma alta velocidade. Descreveu-se que o motorista encontrou fatores externos em seu caminho, como a pista escorregadia e a falta de visibilidade. Para um grupo de pessoas foi dito que o motorista estava com pressa para chegar a sua casa e esconder um presente de aniversário; para o outro, que sua pressa para chegar em casa era para esconder um frasco de cocaína. Indagou-se então se a velocidade do motorista ou os fatores ambientais desempenhou um papel mais significativo para a ocorrência do atropelamento. Aqui o resultado novamente seguiu o padrão que descrevemos e os participantes atribuíram mais culpa ao motorista quando ele estava indo esconder cocaína do que quando ele estava indo esconder o presente de aniversário.

---

<sup>52</sup> Knobe acredita que não é só o conceito de intencionalidade que é afetado pela moralidade. Outros conceitos psicológicos, como desejo, crenças e até conceitos não psicológicos como o conceito de causalidade também sofrem o impacto da avaliação moral e podem apresentar assimetrias em seus usos em certos casos (Knobe, 2009).

Como dissemos, os experimentos acima relatados corroboram a afirmação de que nossa análise de intencionalidade de um ato é influenciada por um julgamento moral que fazemos daquele que comete o ato. O ponto que queremos destacar, por fim, é que as implicações dessa discussão para o direito são significativas. Os sistemas penais costumam distinguir entre crimes dolosos e crimes não dolosos<sup>53</sup>. O dolo requer que a ação realizada tenha sido intencional. Isso quer dizer que um crime só deveria ser caracterizado como um crime doloso depois da avaliação sobre a intencionalidade do ato sob julgamento. A intencionalidade deveria ser avaliada de forma neutra e imparcial. Contudo, como foi demonstrado, as avaliações morais, por vezes, interferem em nossos juízos de atribuição de intencionalidade. No direito penal, lida-se com situações que produzem resultados que são considerados ruins do ponto de vista moral. Sendo assim, é provável que, por terem avaliado os efeitos da ação como indesejados ou perniciosos do ponto de vista moral, os juízes e jurados tenham sempre uma inclinação para tratar as ações (pelo menos aquelas que provocam efeitos colaterais moralmente perniciosos) como intencionais e qualificar os crimes como dolosos.

### **4.2.3 Punindo acidentes**

As nossas intuições punitivas são, como já foi dito, na maior parte das vezes e para a maior parte das pessoas, retributivistas. Logo, punimos com base no que o transgressor da norma merece. Esse merecimento parece exigir uma leitura do estado mental do transgressor. Isso fica muito claro quando imaginamos um cenário no qual duas pessoas realizam o mesmo ato, ambas produzindo um resultado danoso para um terceiro. Se uma delas intencionalmente causou o resultado enquanto a outra não o fez, punimos mais severamente o ato que foi realizado intencionalmente. No entanto, ainda que a intencionalidade exerça um papel importante nos nossos julgamentos sobre punição, isso não quer dizer que

---

<sup>53</sup>No direito brasileiro, por exemplo, a questão é tratada atualmente da seguinte forma pelo Código Penal: “Art. 18 - Diz-se o crime: I - doloso, quando o agente quis o resultado ou assumiu o risco de produzi-lo; II - culposo, quando o agente deu causa ao resultado por imprudência, negligência ou imperícia.”

resultados não intencionais estejam sempre livres de punição. Não é totalmente incomum punirmos de alguma forma aqueles que nos prejudicam acidentalmente, e a maior parte dos sistemas jurídicos vislumbram alguma possibilidade de punição para atos acidentais.

Em um experimento (Cushman et al., 2009) envolvendo um jogo econômico probabilístico, apelidado de jogo da “mão trêmula”, um jogador deveria alocar dinheiro entre ele e um parceiro. Ele poderia fazer isso ou de forma egoísta (tudo para ele), ou de forma equânime (metade para cada um), ou generosamente (tudo para o parceiro). Porém, a alocação não era feita diretamente, mas era mediada pelo lançamento de um dado. Se o alocador quisesse ser egoísta, ele então escolhia um dado (dado A) que tinha 4/6 de chance, quando lançado, de dar o resultado egoísta (se, ao ser lançado, a face superior do dado mostrasse 1, 2, 3 ou 4, o alocador ficaria com todo o dinheiro, mas se mostrasse 5, então o dinheiro seria dividido entre o jogador e o parceiro, e se o dado mostrasse 6, então todo o dinheiro iria para o parceiro). Caso o alocador quisesse ser justo na distribuição, ele escolheria um outro dado (dado B), que apresentava 4/6 de chance de gerar como resultado a distribuição igualitária, mas 1/6 de chance de gerar o resultado egoísta e 1/6 de chance de gerar o resultado generoso. Se ele quisesse ser generoso, então escolheria um dado (C) que apresentava 4/6 de chance de gerar o resultado generoso, mas 1/6 de chance para cada uma das outras possibilidades. Assim, o jogador poderia ter uma intenção egoísta, mas uma alocação generosa; uma intenção equânime e uma alocação egoísta; e todos os outros arranjos possíveis. O parceiro, por sua vez, depois de ver a escolha do dado e o resultado de seu lançamento, tinha a chance de responder aumentando, ou seja, recompensando-o, ou diminuindo o pagamento do alocador, punindo-o. Os participantes puniram tanto intenções avarentas como resultados avarentos acidentais (Cushman, 2012: 10).

A resposta para esse comportamento, sugere Cushman (2012:10), pode ser a de que punimos acidentes porque isso teria uma função pedagógica, trazendo benefícios utilitaristas medidos em termos de comportamentos pró-sociais. Podemos, no entanto, não estar totalmente conscientes dessa função quando punimos os acidentes. Talvez estejamos punindo porque achamos que a pessoa que deu origem ao resultado ruim, provocado pelo acidente, mereça ser punida, independentemente dos benefícios que isso poderia trazer para o próprio

transgressor ou para as pessoas que vivem na mesma comunidade. Nesse caso, as punições ainda seriam fruto de uma espécie de psicologia retributivista, já que não estaria baseada nas consequências da punição, mas seria um retributivismo que nem sempre leva em conta a intencionalidade na configuração de merecimento. De qualquer forma, independentemente da psicologia descritiva que subjaz à punição de acidentes, ainda assim permanece a questão sobre se Cushman teria razão sobre sua função pedagógica.

O acidente pode representar um momento de ensino de limites (territoriais, de propriedade ou relacionais) ou um alerta para que aquela pessoa tenha mais cuidado no futuro ao realizar aquela ação. Para esse alerta funcionar, supõe-se que a pessoa que cometeu o acidente deveria ter algum controle da situação, e parece que isso é levado em conta pelo julgador. Há evidência experimental que sugere que punimos acidentes apenas quando os comportamentos poderiam ser controlados (Alicke, 2000; Cushman et al., 2009). No jogo da “mão trêmula”, quando o alocador tinha algum controle probabilístico pela alocação ao escolher um de três dados disponíveis, ele era punido pelo acidente, como foi descrito. No entanto, quando o alocador não tinha qualquer controle probabilístico sobre a alocação – quando ele era forçado a rolar um único dado onde resultados egoístas, generosos ou equânimes eram igualmente prováveis – ele era punido menos, e, às vezes, nem era punido pelos resultados acidentais (Cushman, 2012: 10).

Esse foco na possibilidade de controlar o resultado faz sentido de uma perspectiva funcional que visa a modificar o comportamento futuro de parceiros sociais. Não há valor em ensinar uma lição para alguém que não tinha como ter controle da situação e, conseqüentemente, não tem como modificar seu comportamento no futuro. Para funcionar, portanto, o transgressor deve ser capaz de aprender a lição que se está querendo passar com a punição. Mas de que forma as pessoas punidas aprendem? De maneira a não fazer aquele ato de novo, ou a não causar aquele resultado de novo? A questão é se nosso mecanismo de aprendizagem associa a recompensa ou a punição à ação intencional ou, ao invés, ao resultado que foi produzido.

Um experimento (Cushman, 2012: 11), com uma versão modificada do jogo da “mão trêmula”, tentou iluminar essa questão. Nesse jogo, um jogador atirava dardos em um alvo multicolorido e seus acertos faziam com que um segundo jogador perdesse ou ganhasse dinheiro. Quem atirava não sabia qual a

diferença das cores, que faria com que seu parceiro perdesse ou ganhasse dinheiro. Seu parceiro, entretanto, podia ensiná-lo recompensando-o ou punindo-o a cada tiro. O atirador devia dizer onde iria acertar antes de arremessar o dardo e toda vez que ele acertava a cor que dizia, ele recebia um bônus do experimentador. Assim, ele tinha um incentivo claro para ser honesto sobre o que ele estava mirando. O parceiro sabia, portanto, onde o atirador acertava o dardo e qual era a sua intenção.

Os resultados mostraram que o atirador aprendeu melhor o valor dos alvos quando o parceiro adotava uma estratégia de punição baseada no resultado, e não na intenção (Cushman, 2012: 11). Isso indica que há, portanto, vantagens em se punir resultados acidentais em certas circunstâncias, dada a estrutura do processo de aprendizagem humana, que parece associar de forma mais satisfatória os estímulos de recompensa ou punição ligados ao resultado do que aqueles ligados a sua intenção.

Embora nossas motivações psicológicas sejam basicamente retributivas, e não estejam preocupadas com as consequências, elas podem produzir efeitos de dissuasão ou incapacitação. “A punição pode ser adaptativa para razões dissuasivas em um ‘último’ nível adaptativo, e ainda assim serem instanciadas por mecanismos retributivos em um nível adaptativo ‘próximo’” (Cushman, 2012: 9) . Talvez essa intuição retributiva traga com ela a solução para o problema de como garantir pró-sociabilidade de seus parceiros sociais no futuro e isso seja algo bom. Talvez pensar em todos os casos de forma utilitarista sobre os benefícios futuros da pena demande tempo e acabe gerando mais erros.

Cushman (2012: 9) usa uma metáfora ilustrativa: consumimos açúcar e gordura pelo gosto agradável, e não por causa de um aprendizado acerca da associação deles ao fornecimento de energia. Ter um gosto inato por esses produtos resolve o problema de se ter que aprender por associação quais propriedades de possíveis comidas estão ligadas ao fornecimento de energia. Parece que nosso gosto por punição acaba resolvendo o problema da necessidade de se ensinar que as pessoas devem cooperar e o problema de ter que ficar refletindo, a cada momento, sobre qual mecanismo vai possibilitar tal cooperação.