

## Should the law depend on luck?

Fiery Cushman

On a snowy Sunday afternoon Hal and Peter watch football and share beers at a local bar. Both drive away intoxicated, and both lose control of their cars on the slick roads. Hal collides with a neighbor's tree, while Peter collides with a young girl playing in the snow and kills her. In my home state of Massachusetts, Hal can expect a fine of several hundred dollars, along with a temporary suspension of his license. Peter's punishment is dramatically different: For vehicular manslaughter he faces between 2.5 and 15 years in prison.

Cases like this have long vexed philosophers and legal scholars, who refer to the problem as "moral luck". Should a chance outcome exert such a powerful influence on our moral judgments? Cases like these are troubling because we feel caught between two irreconcilable perspectives. On the one hand, it seems wrong to punish Hal and Peter differently when they engaged in absolutely identical behavior. On the other hand, it doesn't seem right to send Hal to prison for years on a first-offense DUI charge; neither does it seem right to let Peter off with a ticket for killing a girl.

A growing body of psychological research shows that we feel conflicted about cases like Hal and Peter's because we are, quite literally, of two minds. One mental mechanism assigns punishment in proportion to the harm that a person causes, and therefore judges Hal and Peter very differently. Another mental mechanism judges actions morally wrong based on the intent to harm, or the risk of harm, and therefore judges Hal and Peter identically. When the divergent

outputs of these mental mechanisms are laid side-by-side, the result is a difficult moral dilemma. These insights into the psychological origins of "moral luck" add a new dimension to the philosophical problem: How will advances in the science of moral judgment change the way we think about the law?

•

In the 1930s a young Swiss psychologist name Jean Piaget began to ask young children simple questions about right and wrong, and thus gave birth to the field of moral psychology. By the time of his death in 1980, Piaget had become the most influential developmental psychologist in history, contributing countless groundbreaking findings about the thought processes of young children. Piaget's great gift was to find cases where children make judgments that are strikingly different from those of adults, and his study of moral judgment is no exception.

In the most famous test of moral judgment, Piaget told children stories about two young boys who break teacups. One boy was trying to set the table to help his mother, but dropped a tray of teacups and broke 15. The other boy was trying to steal cookies while his mother wasn't looking, and knocked a single teacup to the floor. Piaget asked children which of these boys was more naughty. To an adult, the answer is perfectly obvious: The boy trying to steal cookies was more naughty because breaking the teacup was a result of his bad intention to steal, whereas the boy who broke 15

teacups acted with a good intentions. But Piaget found that the opposite answer was just as obvious to children under 5 or 6 years old: The child who broke 15 teacups was more naughty, because 15 teacups is a lot more than one.

Piaget's experiment reveals a basic tension between two factors that shape our moral judgment: the harm a person causes, and the harm a person intends. Over the past eighty years his experiment has been replicated and extended dozens of times, with highly consistent results. Young children focus relatively more on causal responsibility, while older children and adults focus relatively more on intent.

A few of these replications hint at an interesting asymmetry in the contest between causation and intention. When a person intends to do harm but doesn't cause any (for instance, throwing a ball at somebody's face, but missing), even very young children recognize that they were naughty. In this sense, young children are perfectly able to see that intentions matter. What appears gives children trouble is the opposite situation, when a person causes harm that they did not intend (for instance, throwing a ball at a target but accidentally hitting somebody's face). To follow up on these suggestive findings, I sent a team of undergraduate research assistants to Boston's Museum of Science to interview hundreds of preschoolers. Our results confirmed this asymmetry. Young children easily condemn people with bad intentions even when they don't cause harm, but they have trouble using good intentions to excuse somebody who does cause harm. In short, kids are tough on accidents.

Our study added another twist to Piaget's by asking children to make two different kinds of moral judgments: naughtiness, but also punishment. Here, again, we found an interesting asymmetry. By seven years old, nearly all children acted like adults and refused to say that somebody was a "bad, naughty boy"

if the harm they caused was accidental. But, not so with punishment: A substantial fraction still claimed that accidental harm-doers should be punished. If kids are tough on accidents, they are really tough on punishing accidents. In that way, their judgments match our laws. (Remember the extra years in prison that Peter got for causing the death of a girl, despite the fact that his intentions were no different from Hal's.) Is it possible that even adults would punish an accident?

To put the question to the test, I conducted an online survey of over one thousand adults. They read a number of hypothetical scenarios, making a moral judgment after each one. Some were asked to make judgments about "moral wrongness", while others were asked to judge the "punishment deserved". The pattern of adult judgments was strikingly similar to seven-year-olds'. Judgments of moral wrongness depended almost exclusively on intent: An attempt to harm was judged very wrong whether or not it succeeded, while accidents were fully excused. But judgments of punishment were strongly influenced by causal responsibility: Attempts to harm were punished more when they succeeded than when they failed, and accidental harms were not fully excused.

This asymmetry between judgments of wrongness and judgments of punishment can explain why moral luck presents a dilemma. From the perspective of wrongness, Hal and Peter look the same: Both intentionally drove under the influence of alcohol. But from the perspective of punishment, Hal and Peter caused very different outcomes: Hal caused negligible harm, while Peter ended a young girl's life. Either way we try to settle the case of Peter and Hal, half our brain is dissatisfied.

Over the years, philosophers and legal scholars have devised many creative, complex explanations for moral luck. But, the psychology of the phenomenon

seems to be in place among preschool children. When a harmful outcome occurs, it triggers a strong moral judgment. Very young children find it especially hard to override this negative judgment, even if the harm was accidental. By seven years old, they consistently say that accidental harms are not moral wrong. But even in adulthood, the tendency to *punish* harmful accidents persists. To be sure, the different sentences that our laws hand down to Peter and Hal probably depend in part on sophisticated philosophical principles and policy considerations. But, they probably also depend in part on a basic psychological impulse present in the youngest minds: The impulse to find and punish those who cause harm.

•

When people say that somebody should be punished for an accident, do they really mean it? Perhaps it's just cheap talk, trying to sound "tough on crime" — or maybe they don't believe that the harms described in my scenarios were *truly* accidental at all. Our next study of punishment addressed these concerns by designing a game that created real, actual accidents in the lab. We made sure that the accidents were perfectly unintentional, and then we gave people the chance to reward or punish those accidents with real money.

At the heart of the game was a simple division of money: Player A was given \$10 to divide between herself and Player B. She could keep all \$10 (stingy), divide it \$5/\$5 (fair), or give all \$10 to Player B (generous). In order to introduce accidents, we built in a catch: Player A had to make this division by choosing to roll one of three different dice. Die 1 was stingy if it came up 1,2, 3 or 4 — but a 5 was fair, and a 6 was generous. So, if you wanted to be stingy, you would roll Die 1. But, accidentally, you might be fair or

generous (by rolling a 5 or a 6). Die 2 and 3 worked similarly: you could try to be fair (Die 2) or you could try to be generous (Die 3), but each die had a small chance of an accidental outcome.

By paying attention to the die Player A selected, Player B could know exactly what A's *intentions* were. (Just to be sure, we asked several questions to check that Player B understood.) But then, depending on the roll of a die, accidents could happen. Player A could choose Die 1, the "stingy die", but end up with a generous outcome. Conversely, she could choose Die 3, the "generous die", but end up with a stingy outcome. We gave Player B the chance to reward or punish Player A by adding or subtracting money from Player A's payoff. When accidents like these occurred, how would Player B choose to reward or punish Player A? By focusing on her intentions (the die she chose), or by focusing on the outcome (the way the die came up)?

The results indicated a strong role for accidental outcomes. When Player A chose the stingy die but it came up generous, on average Player B responded by rewarding. And, when Player A chose the generous die but it came up stingy, on average Player B responded by punishing. Statistical analysis showed that Player B paid attention both to Player A's intentions, and also to the outcome of the roll — but outcomes mattered just slightly more. It's behavior like this that put the "luck" in moral luck. Our impulse to punish somebody who causes harm to us sometimes depends on nothing more than a roll of the dice.

This automatic search for casual responsibility in the face of harm can lead to unpredictable, even bizarre patterns of punishment. One example comes from a study I conducted on people's judgments of attempted murder. First, I asked one group of participants to say how much a competitive runner should be punished for trying to kill his rival by poisoning him with poppy seeds on his salad. The runner misunderstood that his

rival was fatally allergic to poppy seeds — fortunately, the allergy was actually to hazelnuts, so the rival was just fine. Most people thought that the runner should be punished for the attempted murder, assigning an average of about 20 years in prison.

Next, I asked another group of participants to judge a slightly different case. Once again, the runner sprinkled poppy seeds on his rival's salad. And, once again, the rival was actually allergic to hazelnuts. But, this time, the rival's salad happened to have been made with hazelnuts by the chef. Consequently — and completely coincidentally — the rival dies. From a rational perspective, this coincidence shouldn't affect the runner's punishment at all: He still performed exactly the same attempted murder. Yet, people were nearly twice as likely to *let the runner off the hook* when the rival died coincidentally. I've now tested dozens of cases like this in hundreds of people, and the results are strikingly consistent: Attempted crimes are punished *less* when a harm occurs coincidentally, compared to when the harm simply doesn't happen.

The explanation depends on our basic impulse to assign causal responsibility for harm. When the runner dies from eating the salad, who is causally responsible? The chef, not the runner. People get focused on causal responsibility, and assign less importance to the runner's attempted murder. If the runner isn't harmed at all, however, no assessment of causal responsibility is required, and this allows people to fully consider the runner's murderous intent.

All of these studies point to the same conclusion about moral luck. When it comes to punishment, people care a lot about causal responsibility. When you cause harm, you tend to get punished even if it was accidental (think of the dice game). And when you don't cause harm, you might be let off the hook even if you were attempting to murder (think of the runner). But *why*? And, what should we do about it?

•

It is never possible to say with certainty why our brain is designed the way it is. The forces of evolution shaped our brain over many millions of years, and we can't turn back the clock. Moreover, many of the thoughts we think weren't designed by evolution at all: For instance, our thoughts about quantum physics or global warming. But for many patterns of human thought, we can make a good guess about evolutionary design. For instance, sexual intercourse between siblings is forbidden by most human cultures throughout the world and across history. Most animals avoid incest, and even many plants have structures designed to avoid self-pollination. Incest has very clear genetic costs, and that means loss of fitness in the evolutionary struggle for survival. It is a good bet that evolution designed our brains to avoid sex with siblings.

It is also likely that evolution has a hand in our instinct to punish. Biologists and economists have used mathematical models and creative experiments to show that punishment can play an important role in supporting good social behavior. This conclusion is not very surprising: We punish in order to teach people how to behave. But, how can we explain the tendency to punish accidents? Is there some advantage to punishing a person who caused harm unintentionally?

In the end, the answer to this question depends on how people *learn* from punishment. To see this point more clearly, let's return to Piaget's example of the boy who breaks 15 teacups when he tries to help his mother set the table. How should his mother react? From one perspective she should punish him, teaching: "don't break any more cups!". But from another perspective she should reward him, teaching: "thanks for trying to help!". Which is the best strategy for this Mom? It all depends on the way her son learns. If he assumes that

punishments and rewards relate to the outcomes he causes, then punishment will teach him not to break cups. But, if he assumes that punishments and rewards relate to the outcomes he intends, then rewards will teach him to help set tables. Mom's best move depends on the design of her son's brain.

In order to answer this basic question about human learning, we set up another accident-prone game in the laboratory. This one involved darts, rather than dice; we called it "Two Player Darts". Player A throws darts at a board with four colors on it — but, depending on what she hits, she actually wins and loses money for Player B. Here's the catch: Player A doesn't know which colors are good for Player B (earning him money) and which colors are bad (costing him money). The only way that Player B can teach Player A is to reward or punish her after each throw. If Player B uses his punishments and rewards correctly, he can teach Player A to aim for the most valuable colors over the course of 12 throws.

Critically, Player A had to call her shots, announcing which color she was aiming for. (The experimenter rewarded her for accurate throws, thus ensuring her honesty). Because our subjects weren't perfect at darts, about half the time they hit the wrong color. This sets up a choice for Player B: should he reward or punish for the color Player A aimed at, or instead should he reward or punish the color that Player A hit? In order to find out, we secretly forced Player B to reward and punish based on *hits* for half our participants, and instead on *aims* for the other half. Then, we measured which strategy did the better job of teaching A which colors Player B preferred.

The results were clear: Player A learned which colors were good for Player B about twice as effectively when rewarded and punished based on what she hit. Punishing accidental outcomes was the best strategy in this game because it turns out that people *learn* from

accidental outcomes. Although we cannot know for sure, this suggests a possible evolutionary explanation for why we have a basic instinct to punish bad outcomes. Two Player Darts recreates a time in our evolutionary past when the only way to say to "do this!" was to reward, and the only way to say, "don't do that!" was to punish. Given those circumstances, the results suggest that punishing accidents is a good idea. Maybe you stepped on my tail by accident — but when I bite you, you'll learn to watch out next time!

Yet, as with many of our evolutionary instincts, it's fair to question whether this one is still advantageous. For instance, we evolved an insatiable appetite for fat and sugar when they were scarce nutrients, but today overconsumption leads to heart disease and diabetes.

Punishing accidents may have been the best way to teach good behavior in a time before we had language and laws, but in the modern world we can speak much more effectively with our tongues than our fists. The 'teacup mom' doesn't need to *punish* her son in order to say "don't break teacups" — she can just tell him. And, if I had not prohibited the players in Two Player Darts from talking to each other, surely Player B would have said to Player A, "I lose money when you hit red — so, I'm going to punish you if you even aim for it." Because we can use language and laws to communicate moral rules *before* they get violated, we may have outgrown the evolutionary necessity of punishing accidents.

By the time we're seven years old, we know that Hal and Peter acted equally wrongly when they drove drunk. For their reckless behavior, surely each deserves to be punished — but why not punish them the same amount? In our evolutionary past, extra punishment for Peter was the best way to communicate a moral rule. In our evolutionary present, maybe we can let our laws do that work.