# INTRODUCTION

Evolution built us to punish cheaters. Without that punishment instinct, we would never have been able to live in small groups, and would never have realized all the significant benefits that small-group living conferred, including mutual defense, cooperative hunting, property, divisions of labor, and economies of scale. In fact, to a large extent our notions of right and wrong, of empathy and compassion, of fairness and justice, all come from the tensions of group living, and thus indirectly owe their very existence to punishment.[1] It may sound strange that one key to civilization is our willingness to punish each other, but every parent knows it's true.

Every parent also feels the irresistible pull not to punish too much, and in fact maybe not to punish at all – to forgive – and this, too, is a remnant of evolution. Our punishment instinct is not so much a sword ready to fall as it is a finely tuned balance, sometimes susceptible to the gentlest of breezes.

But there's at least one substantial weight on this otherwise delicate balance: the idea that we generally do not punish accidents. Cheaters cheat intentionally, and the intentional wrongdoers, not the careless ones, get most of our punishment attention. We all blame the contract killer more than the inattentive driver who kills a pedestrian. We therefore care as much or more about the wrongdoer's state of mind – evil or just careless? – than we do about the actual harm his wrong has caused.

Several years ago I presided over a first-degree murder trial in which a young Czech émigré was charged with stabbing his Brazilian au pair girlfriend.[2] The crime took place in the au pair's bedroom, in the basement of her employer-family's house. The young man stabbed her seventy-four times. He confessed to the murder but denied it was

premeditated. Despite his denial, the premeditation evidence was pretty strong. He not only entered the bedroom through a window, armed with a knife and carrying some duct tape, but he also admitted to police that a few days before the killing he tried to dig a small grave in a remote field but gave up because the ground was frozen.

On the other hand, he testified that he regularly went through the window for late-night visits with her, and that he went there that night not to kill her but only to see if she would change her mind about breaking up. He claimed he had the knife and duct tape because he was moving. As for the grave, he testified that he started to dig the hole in the field to "bury her memory," and that all he intended to bury there were a few items of personal property that reminded him of her. When he went to see her that final time, and she told him she was set on leaving him, he "snapped."

But he didn't say the word "snapped." What he said was, "A darkness came across my eyes." He even said it a second time in cross-examination. It seemed oddly and rather beautifully phrased, and vaguely familiar. Neither of the lawyers asked him about it. Long after the jury convicted him of first-degree murder and I sentenced him to the mandatory life in prison without the possibility of parole, it hit me. "Darkness covered his eyes," and variations of that phrase, are used over and over by Homer to describe many of the battle deaths in The Iliad.[3]

Was this a bad literary joke played by a deranged killer, a coincidence fueled by awkwardness in a second language, an unconscious conflation from a literature class, or some deeply insightful way of expressing this terrible act? Was he describing his own loss of consciousness, a blackout of memory, a kind of disassociation of his moral bearings, or the unbearable blackness of his actions? I have no idea.

But I do know that the task facing the jury in that case – to decide not so much what happened, because that was clear, but rather what was in the defendant's mind at the time of the killing – was the same daunting task that often faces judges and juries in virtually all legal systems. This is one of the Big Secrets of the criminal law. Whether a defendant actually committed the charged act is seldom the subject of serious dispute, even in those few criminal cases (roughly 5 percent in the United States) that go to trial.[4] Sure, there are some crimes,

and even some categories of crimes – like sex assault with no forensic evidence – that are whodunits or whether-anyone-dunits. But much more commonly there is no doubt at all that the defendant killed the victim or robbed the liquor store or stole the car; the only issue for the jury to decide is what in the world was going through the defendant's mind at the time of the crime.

These judgments about whether a criminal defendant acted intentionally, unintentionally, or with some state of mind somewhere in between can have enormous consequences. A young mother leaves her infant in the bathtub to go downstairs to charge her cell phone, and the baby drowns. If a jury determines the mother was just a clueless nitwit, in my state she is guilty of a crime called negligent child abuse resulting in death and could be punished anywhere from getting released on probation to spending sixteen years in prison. But if the jury finds she "knew" she should not leave but did so anyway, this crime is tantamount to first-degree murder and carries a mandatory sentence of life in prison without the possibility of parole.

Even in civil cases, jurors must sometimes make difficult judgments about people's states of mind. In a negligence case, they may have to decide whether the defendant was simply unaware of the risky nature of his conduct, which is the essence of ordinary negligence, or whether there was something more, some kind of gross negligence that might justify some punitive damages beyond mere compensation of the victim. Even in cases of ordinary negligence, where the defendant's state of mind is not usually in play, the jury's judgment is a *community* judgment, not so much about what happened as whether the community should tolerate what happened. Some accidents are the product of such carelessness that they need to be punished – if not with incarceration then with the hammer of damages. I presided over a medical malpractice case several years ago that put these community judgment aspects of negligence in particularly sharp relief.

The plaintiff was a woman with advanced lung cancer. She claimed that several years earlier the defendant-radiologist misread her chest X-ray, and that in the intervening years what would have been a highly curable case had developed into a hopeless one.[5] Of course both sides called experts who testified that a reasonable radiologist would, or would not, have diagnosed lung cancer based on the X-ray, or at least

would, or would not, have called for further tests, and the jury, as usual, was forced to pick between the two expert narratives. So far, this case was no different from any failure-to-diagnose case.

But by the time this case reached my jury, Colorado had joined the states implementing several kinds of jury reforms, and one of those reforms was to allow jurors to submit written questions to witnesses, including expert witnesses. One of my jurors submitted this question to the plaintiff's expert: "What percentage of radiologists would have diagnosed cancer based on this X-ray, or at least called for more tests?"

Strangely, this is not a question permitted by the law in most states. Experts may opine about the standard of care to which a fictitious "reasonable" professional must adhere, but the law generally prohibits witnesses from putting any kind of statistical gloss on that fiction by telling jurors their view about whether this standard of care would be followed by 10 percent, 51 percent, or 99 percent of professionals. But there were no controlling cases in my state. So in a fit of either insight or confusion, but in either event fueled in part by the fact that this question was being asked by one of our jurors and not by a lawyer, I decided to allow the question. The plaintiff's expert responded by testifying that 95 percent of radiologists would have tested further. The jury returned a very large verdict for the plaintiff.

In post-trial motions the defendant's lawyers argued that I had erred in permitting the statistical question, and after mulling it over I decided they were right, and that my error likely infected, indeed perhaps even drove, the verdict. I granted the defendant's motion for a new trial, but the case settled before any retrial.

This rule that expert witnesses cannot describe the standard of care in terms of adherence probabilities seems crazy. How can a "standard" be standard unless at least a majority of practitioners adhere to it? Conversely, how can it not be the standard if 99 percent of practitioners adhere to it? There are many justifications for this rule, but the central one is that negligence – even negligence by a professional in a specialized area as far from ordinary experience as radiology – is the kind of community judgment we want ordinary citizen-jurors to be able to make, not the professionals themselves by plebiscite.[6]

Even in civil cases as seemingly cut-and-dried as contract cases, these kinds of squishy community judgments, often dependent on

guesses about what was in the contracting parties' minds, are not uncommon. Whether there was a contract at all depends on a determination of whether the parties had a "meeting of the minds." If they did form a contract by having a meeting of their minds, whether one party breached that contract can also depend on an assessment of mental states. For example, a buyer's obligation to pay for a shipment of rotten avocados may well depend on whether the shipper used his best efforts to deliver them before they spoiled. "Best efforts" is hardly a straightforward test, and might not only require an inquiry into what the shipper did and did not do to contribute to the delay but also what was in the shipper's mind as he made these contributing decisions. Modern law now implies in all contracts the duty of the contracting parties to act in good faith.[7] Trying to weasel through loopholes in contracts – once the proud domain of all commercial lawyers – is now largely blocked by this troublesome good-faith requirement. Today, two shippers whose delayed delivery is the product of identical acts may well suffer very different legal fates if one shipper's acts were motivated by bad faith and the other's were not.

These are just a few examples of the profoundly democratizing and morality-setting roles of jurors. They not only get to decide whether a defendant committed the acts he's charged with committing; they often get to draw the community's lines in deciding whether those acts are wrong enough to be punished.[8] And in deciding the "wrongness" of the acts, jurors often must make judgments about the wrongdoer's intentions.

This emphasis on the state of mind of the wrongdoer is hardly unique to the law. Imagine coming home and finding your teenage son holding a hammer, standing in front of your brand new, but now completely smashed, big-screen high-definition television set. When you ask him what happened, you would probably not be satisfied if all he told you was "I hit the TV with the hammer." The "Why," the "Was it an accident and if not what the hell were you thinking?" is as big a part of the What Happened as anything else. It is, in fact, the *moral* part of the inquiry, the part that will determine whether your son gets a hug, a month-long grounding and a bill for a replacement TV, or a class in home repairs.

And it has always been so, with virtually every aspect of human interaction. We are social animals highly attuned to the behaviors of others, constantly monitoring, evaluating, and predicting those behaviors. We instantly recognize behaviors that don't make sense, that by their very irrationality and unpredictability threaten us or our families. And we are drawn irresistibly to imagine the minds behind those behaviors.

When the behaviors of others get more and more irrational and unpredictable, and more distant from the ordinary experiences of inquiring parents or jurors – like stabbing your ex-girlfriend seventy-four times – it becomes more and more difficult for the judges to put their minds into the minds of the judged. At some point we all come to a big fork in the road of responsibility. Some say, "That was so crazy he must be insane," or, less technically, he "snapped." That is, his mind is, or at least was at the moment of the crime, so unlike ours that we cannot apply the ordinary human presumptions of intentionality to him, and thus cannot justly hold him as responsible as we would a less mentally disabled person.

But some take the other fork. "He wasn't crazy, he was evil." He didn't "snap," he was just a possessive, jealous lout who was so self-centered that when the victim told him he couldn't have her he made sure no one else ever would. All of us have faced rejection, but only evil people kill a helpless girl whose only wrong was to say she wanted out.

This apparent divide between crazy and evil faces every one of us in one form or another in some context or another. Legal academics have debated it for as long as there have been legal academics. Citizens ponder it each time they read about a young mother drowning her children or the latest shoot-up by a disgruntled employee or student. Judges and jurors deal with it in almost every serious criminal case, even when insanity is not formally an issue, the "crazy" narrative presented by the defense and the "evil" narrative by the prosecution.

Of course, the task of navigating these two narratives is made doubly difficult by the fact that the one person most likely to know for sure What Happened, in the interior sense of those words, is the accused himself, and he usually has a huge incentive to lie about his state of mind. "Jason was tossing the hammer to me and it slipped."

"I was burying her memories, not her." I suspect that when you read my condensed version of the Czech defendant's testimony, your immediate reaction was that he was lying. There are good evolutionary reasons for this strong presumption that all people know what they are doing, and for the corollary that people who claim harmful acts were accidental, or the product of diseased minds, are lying.

But in the right kind of case this strong presumption bumps up against an almost equally strong presumption that accidents should not be punished the same as intentional wrongs. We know from our own lives that accidents sometimes happen, even accidents that don't look like accidents from the outside. We also know, or think we know, that with sufficient pressure, ordinary humans like us might lose control and "snap." And yet most of us have never stabbed someone seventy-four times, or even come close to doing so, or had the legal incentive to lie about such "snapping." Whether that's because we have not suffered the same pressures as our young Czech killer, or because his brain was less able to withstand those pressures (either because he was crazy, or evil, or for some other arguably more informative reasons), are questions humans have been asking ourselves for as long as we have roamed the planet.

This book is about the evolutionary roots of this interior struggle in which we all engage when we judge the behaviors of others (and, of course, simultaneously judge ourselves by comparison). Evolution's last laugh, when it came to the human animal, was that it built our brains with two deeply conflicting predispositions. We are predisposed to cooperate with each other, because living in groups gave us a substantial long-term survival advantage. But we are also born cheaters, because cheating in the right circumstances gave us a short-term survival advantage. As these two conflicting tendencies tugged for our souls, we simultaneously evolved punishment behaviors – a way to dampen cheating by increasing its short-term costs to the cheater. But our punishment instincts are infected with the same conflict – our brains have been built to punish cheaters, but that punishment urge is intrinsically restrained, in no small part because we all know that we, too, are born cheaters.

What modern judges and juries do is essentially what our ancestors have been doing for 100,000 years when they had to decide what to

do with misbehaving individuals.[9] As clans banded with other clans to form larger groups, and groups of groups formed tribes and ultimately settled into villages and towns, these punishment decisions became more and more formalized and institutionalized. They also changed to reflect the specific challenges of each society and culture. But at their heart is the same ancient three-pronged punishment strategy natural selection has bequeathed us, and which will be the central topics of this book:

1. We punish ourselves with conscience and guilt.
2. We punish our tormentors with retaliation and revenge.
3. As a group we punish the wrongdoers we are able to detect with retribution.

Conscience helps us avoid the temptation to cheat in the first place, and guilt makes us feel bad after we've cheated, so that next time maybe we'll be less likely to cheat. When conscience and guilt are not enough to deter us, the risk that our victims will retaliate may be. And when neither conscience nor fear of retaliation is enough, the fear of detection and punishment by the group adds a third level of deterrence.

But this last level of punishment, this socialization of punishment, carried its own special evolutionary problems. With everyone primed to punish everyone else for every transgression, our groups would have dissolved into a kind of punishment free-for-all. Several things prevented that. First, our urge to punish came with some built-in thresholds; our brains are not in fact primed to punish every single kind of social transgression no matter how minor. We don't bother with small slights, or even with large ones when we are sure they were accidents. We also sometimes forgive. Finally, our urge to punish third-party wrongdoers comes with a simultaneous urge that someone else inflict the punishment.

In our early groups, dominant members likely took charge of most third-party punishment, as they took charge of many group-wide decisions. As societies became more and more complex, and divisions of labor started to take root, dominant members delegated third-party punishment to other members (judges), as they delegated many of their ancient tasks. For certain kinds of particularly risky punishments

(for example, when the wrongdoer was powerful or came from a powerful family), it made sense for the leader to delegate punishment not to a single member but to a small group of punishers (the jury). Thus began judge and jury, as communal representatives of group blame.

The decision of whether, and in what circumstances, a group of humans should punish one of its misbehaving members may not sound important enough for natural selection to have cared about. But 100,000 years ago these punishment decisions were not just important; they were existential. Our most significant predator was ourselves. Homicide was as big a survival concern as food and shelter. Quite apart from the risks of being killed by each other, our lives were hard, our life spans were short, and we were often on the brink of death. Even one cheater could be the difference between a successful hunt and an unsuccessful one, between a strong and prosperous group and an unraveled one, between life and death.

But punishing cheaters could also be disastrous. The two most serious kinds of punishment were execution and banishment. Because banishment was in effect an almost certain death sentence, both of these punishments risked deadly resistance not just by the wrongdoer himself but also by his relatives and friends. Even without any resistance, execution and banishment reduced by at least one the number of members who could hunt, guard, cook, and do all the other work of the group. And if a banished wrongdoer's entire family decided to leave with him, the group could be devastated.

Punishments short of execution or banishment might also be resisted. When we lived on the edge of survival, being forced to pay any kind of punishment price was often a serious matter. For many of us today, missing a meal might improve our health. Missing one 100,000 years ago might have tipped us into unrecoverable illness. So deciding who should be punished, how they should be punished, and for what infractions, was an incredibly important problem for our ancestors – important enough for evolution to have some say in a solution.

Natural selection has painted our natures both in broad themes and in subtle details. In many ways, this book is about bringing these two images together, through the lens of punishment. The big issues of guilt, innocence, responsibility, blameworthiness, apology, atonement,

and forgiveness impact the dirty little daily processes the law uses to make very practical decisions about how to punish. Law is applied human nature, and understanding that nature – not just the nature of the people being judged but also the nature of the people doing the judging – is critical to understanding the foundations of law. Understanding those legal foundations may in turn allow us to gauge the distance between them and some of the legal outbuildings our cultures have erected. I will spend time in later chapters discussing some legal dissonances – legal doctrines that seem to be in conflict with our evolved moral intuitions and punishment instincts – and whether those doctrines should be reformed. I will also explore the extent to which an evolutionarily informed view of blame and punishment might help improve our legal processes, and might even help make sense of our famously disparate and seemingly irreconcilable theories of punishment.

In the end, though, my efforts will not be focused on legal reform or on improving legal processes or theory. My goal is less lofty, more personal. My hope is that by the end of this book readers will be able to feel the connections between the twelve people who decided that my young Czech defendant must spend the rest of his life in prison and our ancestors 100,000 years ago who made similar life-and-death decisions.

After introductions to the problem of social cooperation and defection in Chapter 1 and to the evolution of detection and blame in Chapter 2, the next three chapters examine the three layers of punishment. Chapter 3 addresses first-party punishment (conscience and guilt), Chapter 4 second-party punishment (retaliation and revenge), and Chapter 5 third-party punishment (retribution). Chapter 6 deals with forgiveness – evolution's yin to punishment's yang – and the related topics of apology and atonement. Chapter 7 considers the delegation of third-party punishment to a single member (judge) and to multiple members (jury) and surveys the history of the Western jury. Chapter 8 analyzes the general problem of legal dissonances – rules that conflict with our evolved blaming and punishing instincts – and suggests a five-step framework to decide when a legal rule should give way to a conflicting evolved instinct. Chapter 9 addresses some tensions