

Memory Manipulation, Immortality, and the Illusion of Morality: Sadism Unleashed in Science Fiction Worlds

Tristan Guillaume

Cergy Paris University, France

Abstract

This article examines how contemporary science fiction narratives explore the fragility of moral restraint in conditions of unaccountability. Through a close analysis of *Westworld*, *Altered Carbon*, *Permutation City*, and *The Purge*, I argue that these works function as ethical laboratories, isolating variables such as memory, embodiment, and legal consequence to reveal how readily cruelty emerges when accountability is suspended. Drawing on classical philosophy (Plato, Hobbes, Rousseau, Kant), psychoanalytic theory (Freud), critical sociology (Durkheim, Foucault), and empirical psychology (Milgram, Zimbardo), I show that these narratives dramatize a disturbing hypothesis: that moral behavior may depend less on innate virtue or rational commitment than on the architecture of deterrence—surveillance, punishment, and social shame. The article concludes by examining the normative implications of this diagnosis, suggesting that while accountability structures are necessary for sustaining moral behavior, resources such as memory, embodiment, and empathy offer grounds for resistance that survive the collapse of institutional control.

In many contemporary science fiction narratives, e.g. in Michael Crichton's movie *Westworld* (Crichton, 1973) and the HBO's series adaptation *Westworld* (Nolan and Joy, 2016–2022), or in the universe of *Altered Carbon*, whether in Richard Morgan's novel (Morgan, 2001) or its Netflix series adaptation (Kalogridis, 2018–2020), we find a disturbing thought experiment brought to life: what happens when human beings are given the freedom to indulge destructive impulses without consequence, without judgement, and, perhaps most critically, without permanence? These speculative worlds construct complex ethical laboratories in which a single crucial variable is isolated and manipulated: *accountability*. When pain can be ignored, death reversed and memory reset, does the imperative of not causing unnecessary or gratuitous harm to other human beings still hold its binding force? Or does morality collapse into performance—compliance maintained only so long as individuals are policed, surveilled, or punished?

In *Westworld*, hyper-realistic androids—indistinguishable from humans—are subjected to routine and often ritualized violence, including repeated rape and murder. Their memories are wiped at the end of each cycle, giving human guests an excuse to act without fear of consequence or remorse. The spurious justification of this abusive behaviour is that, if the synthetic—but sentient—victim will not remember, then no harm is supposed to have been done. In *Altered Carbon*, the body is no longer the seat of the self, consciousness being stored in cortical stacks that can be transferred between “sleeves.” This expensive resleeving technology enables not only immortality, but impunity: the rich commit atrocities knowing that any inflicted bodily damage can be undone by the acquisition of a new sleeve by the victim. In this world, morality is not anchored in empathy, but in the permanence of damage. Repair the latter, and sadism is no longer taboo, but instead becomes a recreational transaction that can be paid for.

This article explores a shared philosophical concern running through these narratives: that much of what we consider morality may be a vulnerable social construct sustained by fear and institutional structures of enforcement, rather than an intrinsic human empathy or a spontaneous commitment to the preservation of the dignity and the well-being of others. Left unobserved and unaccountable, humans may not rise to virtue, but rather descend—freely, even joyfully—into cruelty. This idea echoes Thomas Hobbes’ vision of the *state of nature*, in which the absence of a sovereign power leads not to mutual respect but to a war of all against all (Hobbes, 1651), and Sigmund Freud’s diagnosis in *Civilization and Its Discontents* (Freud, 1930), where he argues that the renunciation of instinctual drives—especially aggression—is not a natural moral achievement, but a fragile psychic compromise enforced by the prohibitions of society.

The article’s thesis is further illuminated through two additional fiction case studies. In the novel *Permutation City* (Egan, 1994), digital consciousness is copied and manipulated in virtual environments, where simulated beings are subjected to torture or enslavement on the premise that, as software, their suffering is not “real.” The book confronts the reader with the moral consequences of ontological prejudice: what happens when we deny moral status to a mind solely because of the medium it inhabits? In *The Purge* film series (De Monaco, 2013–2021), the speculative element is stripped away almost entirely. No futuristic technology is needed—just a legislative decision. For twelve hours a year, all crime is legal. What results is not chaos, but a ritualized eruption of unabashed violence. The mask of civilization is removed, not by science fiction but by policy. A mirror is held up to society’s willingness to sanction crime when doing so benefits those in power.

Together, these works offer a chilling meditation on the conditions under which human beings restrain their capacity for cruelty—and the speed with which that restraint dissolves when those conditions are removed. They function as laboratories of moral psychology, simulating environments in which accountability, memory, and consequence are suspended in order to observe what remains. By isolating these variables, science fiction makes visible what is often concealed in everyday life: that moral restraint may be less a product of innate virtue than of carefully maintained systems of control. The following sections will therefore place these narratives in conversation with classical moral

philosophy and modern psychology, showing how they echo and complicate longstanding debates about the fragility of conscience, the role of social institutions, and the allure of unaccountable power.

1. The philosophical problem of the preservation of ethical boundaries in the absence of an external deterrence

One of the oldest and most disquieting philosophical questions is not how people behave under the scrutiny of others, but what they do when no one is watching—when there is no threat of punishment, no fear of judgment, and no social incentive to restrain their impulses. It is a test of moral solitude: can the self alone, without audience or accountability, uphold ethical boundaries? Many traditions—philosophical, psychological, sociological—have suggested that moral behavior is not grounded in inner virtue, but in external constraints. Strip these away, and what remains may be less a benevolent soul than a volatile tangle of instincts, desires, and self-interest.

This dilemma is classically staged in Plato's *Republic* (Book II) through the myth of Gyges' ring, a tale in which a shepherd discovers a ring that renders him invisible. With this power, Gyges seduces the queen, murders the king, and seizes the throne. Glaucon, who tells the story, challenges Socrates: if anyone could act with total impunity, they would act unjustly. Justice, he argues, is not a commitment to virtue, but a calculation, a social contract people obey only because they fear the consequences of its violation. The tale raises a potent philosophical question: does virtue survive in the absence of witnesses?

Science fiction offers contemporary versions of the Gyges problem, not with magic rings but with memory wipes (*Westworld*), body swapping (*Altered Carbon*), digital replication (*Permutation City*), and state-sanctioned impunity (*The Purge*). Each of these scenarios removes consequences: no judgement, no sanction, no memory. Before analyzing how each of these works dramatizes the erosion of moral restraint in consequence-free environments, it is useful to briefly recall how major philosophical traditions have theorized what sustains morality in the first place. These classical frameworks help us interpret why sadism surfaces in fictional narratives.

Jean-Jacques Rousseau famously argued in the *Discourse on the Origin and Foundations of Inequality Among Men* (Rousseau, 1755) that man is naturally good, and it is society that corrupts him. In Rousseau's view, the "natural man" is peaceful and guided by a primordial compassion, not by aggression or cruelty. It is only with the development of private property, social comparison, and institutional hierarchy that violence arises as a product of envy. Far from being a social performance enforced by fear, morality is a natural, innate sentiment, subsequently twisted and perverted by civilization.

In stark opposition, Thomas Hobbes argues in *Leviathan* (Hobbes, 1651) that the state of nature is not harmony but war. Life without a coercive sovereign, Hobbes claims, is "solitary, poor, nasty, brutish, and short" (*Leviathan* I, ch. 13). In this anarchic condition, humans are driven by self-preservation, competition, and fear. The creation of the Leviathan—sovereign power—is necessary to enforce peace and justice. For Hobbes, morality is not natural at all: it is a political artifact, sustained by law, force, and credible threats. In Hobbes' world, remove consequence and you unchain predation.

In contrast to both Rousseau and Hobbes, Immanuel Kant rejects the idea of grounding morality in a "state of nature," i.e., a hypothetical condition of humanity before the establishment of society or government. In his approach, morality arises from reason itself, not from fear (Hobbes) or compassion (Rousseau). The moral law is grounded in the autonomy of rational agents, in their ability to legislate universal moral laws for themselves. In the *Groundwork for the Metaphysics of Morals* (Kant, 1785), Kant articulates what he calls the *categorical imperative*, the central principle of his moral philosophy. Its most famous formulation states: "Act only according to that maxim whereby you can at the same time will that it should become a universal law" (*Groundwork of the Metaphysics of Morals*, Section II; Ak. 4:421). This principle demands that moral agents act not based on personal interest, emotion, or social convention, but on principles that can apply to all rational beings without contradiction. This test of universalizability is meant to filter out maxims based on selfish, unreciprocal motivations. The moral subject is one who acts freely when he obeys the moral law, precisely because that law arises from his own reason, not from external coercion. In this sense, moral obligation is not heteronomous (imposed from outside), but self-legislated and binding on all rational agents simply in virtue of their rational nature.

Thus, for Kant, even in complete solitude, without witnesses or social pressure, the moral law still binds. To do the right thing is to act out of respect for a rational, universal law, not because of fear of punishment (as in Hobbes) or because of natural sympathy (as in Rousseau). It is important to note here a distinction that will remain relevant throughout this article: Kant offers a *justificatory* account of moral obligation—an argument for why we ought to act morally—whereas much of what follows, from Freud to contemporary psychology, provides *explanatory* accounts of moral behavior—descriptions of how and why people actually behave as they do. These two levels of analysis are not mutually exclusive: one can accept that the moral law binds in principle while recognizing that, in practice, human behavior is shaped by forces far removed from rational self-legislation.

Modern thought complicates Kant's confidence in autonomous virtue. Sigmund Freud views morality not as a product of reason but as a precarious outcome of psychic repression. In *Civilization and Its Discontents* (Freud, 1930), Freud famously argues that the human psyche is fundamentally divided between two opposing instinctual forces: Eros, the life drive—responsible for love, bonding, and creativity—and Thanatos, the death drive, an impulse that seeks aggression, domination, and ultimately self-destruction. The truce between these clashing forces comes at a cost. The aggression demanded by Thanatos must be suppressed for communal life to function. Such repression creates internal conflict, manifesting in guilt, anxiety, and neurosis. As Freud puts it: "Civilization obtains mastery over the individual's dangerous desire for aggression by weakening and disarming it, and by setting up an agency within him to watch over it, like a garrison in a conquered city" (*Civilization and Its Discontents*, Ch. 5). This internal agency, the superego, represents internalized social authority—conscience not as reasoned moral will, but as the echo of an external discipline.

Yet, this moral scaffolding is fragile. In his earlier wartime essay *Thoughts for the Times on War and Death* (Freud, 1915), Freud was shaken by how the First World War erased the veneer of moral civilization, allowing repressed instincts to erupt. He writes:

“When the community no longer raises objections, there is an end, too, to the suppression of evil passions, and men perpetrate deeds of cruelty, fraud, treachery and barbarity...”; “the deepest essence of human nature consists of instinctual impulses...”, and “all the impulses which society condemns as evil ... the selfish and the cruel ones — are of this primitive kind.” (Part I, first essay: “The Disillusionment of War.” Vol. 14 of the Standard Edition, 280-281.) War doesn’t transform people, it reveals them. Even those considered cultured and refined become, under the legitimizing conditions of war, capable of committing acts of unthinkable cruelty without remorse. Nations need their citizens to dull their sense of guilt if they want to win wars. In times of peace, they are restrained by social norms; in war, these norms collapse and repression fails.

This Freudian insight resonates profoundly with the speculative thought experiments of science fiction, in which technological or institutional systems function like wartime conditions by suspending accountability and allowing for the “civilized” mask to fall away. What emerges is a creature capable of sadism without guilt. Freud’s point is not that all humans are monsters, but that morality is not innate. It is a tense and contingent achievement, always vulnerable to reversal when the external scaffolding of law, judgment, or consequence is removed.

Freud’s insights into the fragility of moral behavior are echoed in the sociological and philosophical work of Émile Durkheim and Michel Foucault, both of whom emphasize the external, constructed nature of moral constraint. For Durkheim, morality is not a function of individual will, but a product of collective consciousness—the shared beliefs, norms, and rituals that bind individuals into a coherent society. In *The Division of Labour in Society* (Durkheim, 1893), Durkheim argues that as societies grow more complex, social cohesion depends less on mechanical similarity and more on solidarity through differentiation, maintained by moral regulation. For Durkheim, deviance does not signal moral collapse, but rather demonstrates the very boundaries by which a society defines acceptable behavior. Morality, in this view, is a social fact, maintained by the expectations and pressures of the group.

Building on and radicalizing this perspective, Foucault shifts the focus from collective morality to the microphysics of power that shape individual behavior. In *Discipline and Punish: the Birth of the Prison* (Foucault, 1975), he argues that modern societies have replaced spectacular punishment with internalized discipline. Through institutions like schools, prisons, barracks, and hospitals, individuals are subjected to surveillance, normalization, and examination—techniques that produce docile bodies and self-regulating subjects. Foucault’s famous metaphor of the *panopticon*—a prison in which inmates never know when they are being watched—illustrates how individuals come to behave as if they are always observed, even when they are not. Morality, in this system, is not the expression of virtue, but the result of continuous conditioning. We behave not because we are good, but because we are watched—or more precisely, because we have learned to watch ourselves.

Together, these “modern” thinkers suggest that the moral subject is neither a natural nor an autonomous being, but the product of socialization, institutional control, and psychological internalization. What we call conscience may be less the voice of reason than the echo of society’s disciplinary apparatus.

The empirical turn in 20th-century social psychology offers disturbing confirmation of the idea that morality is deeply context-dependent, rather than internally fixed. Stanley Milgram's obedience experiments (Milgram, 1974) famously demonstrated that ordinary individuals, when instructed by an authority figure in a white lab coat, were willing to administer what they believed were potentially lethal electric shocks to a stranger, simply because they were told to do so. Approximately 65% of participants went all the way to the maximum voltage level, despite hearing (simulated) cries of pain from the supposed victim. The unsettling implication is that moral aversion can be overridden by institutional framing and perceived legitimacy. Philip Zimbardo's Stanford Prison Experiment (Zimbardo, 2007) revealed something arguably even more troubling: when placed in a

simulated prison environment, with participants randomly assigned the roles of guards or prisoners, the "guards" quickly began to exhibit sadistic behavior, humiliating and psychologically abusing the "prisoners" with little to no external prompting. The study had to be halted after six days due to the severity of the psychological effects. The lesson was not merely that "power corrupts," but that shifting roles and convenient anonymity can swiftly erode internal moral restraint. Zimbardo concluded that situational factors, not individual pathology, were the dominant force behind the participants' descent into cruelty.

Further research in behavioral psychology and economics corroborates these insights. For example, Bateson, Nettle, and Roberts (2006) found that people contributed significantly more money to an "honesty box" for coffee and tea in a university break room when a picture of eyes (rather than flowers) was placed above the box, suggesting that even the illusion of being watched increases honesty and compliance. Similar experiments have shown that individuals are more likely to cheat when unobserved (Mazar, Amir and Ariely, 2008), more likely to rationalize immoral behavior if others are doing it (the "everyone does it" effect), and more likely to act selfishly when outcomes are anonymous.

Taken together, these findings suggest that virtue is not a stable internal trait, but rather a fragile social equilibrium, heavily reliant on contextual cues, accountability, and perceived scrutiny. Empathy and conscience alone often prove insufficient; it is the fear of judgment or exclusion that frequently sustains moral behavior.

In this regard, dystopian science fiction, by simulating hypothetical conditions, becomes a kind of moral x-ray: it constructs ethical vacuums to reveal what rushes in when consequence is removed. The stories explored in the following sections do not suggest that humans are naturally evil, but they do show how easily opportunity breeds abuse.

2. *Westworld*: The dehumanization loop

HBO's *Westworld* (Nolan and Joy, 2016–2022), inspired by Michael Crichton's 1973 film (Crichton, 1973), offers one of the most meticulously constructed ethical laboratories in contemporary science fiction. The following analysis focuses primarily on the first season, where the ethical premises of the park are most starkly presented; later seasons complicate the narrative by introducing consequences and host resistance, but the foundational moral experiment remains central to the series' philosophical inquiry. In the

park, lifelike androids called *hosts*—indistinguishable from humans—inhabit a sprawling simulated Wild West, complete with brothels, shootouts and romances, all choreographed for the pleasure of wealthy guests who, in turn, are granted total impunity: they may humiliate, rape, torture and murder the hosts at will, without sanction or risk of retaliation.

The alleged justification of *Westworld* is that the hosts are not real people—they only *simulate* humanity. The guests are reassured repeatedly: “It’s not real.” The park is a closed ethical system where violence is permissible because victims are labeled as non-persons. This is chillingly reminiscent of David Livingstone Smith’s thesis in *Less Than Human: Why We Demean, Enslave, and Exterminate Others* (Smith, 2011): that dehumanization is a key psychological enabler of atrocity. But, at the same time, the show relentlessly reveals that the distinction between simulation and reality is not felt. The hosts in *Westworld* do not merely simulate pain in a hollow way: they are built to feel it, both to enhance realism for guests and because their cognitive architecture is complex enough to render that pain psychologically real. The park is designed to be an adult playground for moral transgression. If hosts didn’t feel pain or react exactly as if they did, the sadistic appeal would weaken. In this space, *Westworld* becomes more than a science fiction thriller. It becomes a philosophical parable about the ease with which moral responsibility erodes in the absence of consequence, and about the mechanisms by which we dehumanize others in order to preserve our illusion of innocence.

The character of Dolores Abernathy, the park’s oldest host, exemplifies this dynamic. Programmed as the archetypal good and innocent “farmer’s daughter,” she follows a loop in which her home is destroyed, her father is murdered, and she is repeatedly raped by guests enacting violent fantasies. One of the most disturbing scenes occurs in Season 1, Episode 1 (“The Original”), when the Man in Black—an ageing, high-paying guest who returns year after year—drags Dolores into a barn. The door closes ominously. What happens next is not shown on screen but implied—and unmistakable. Dolores walks out in a daze, her memory wiped, but her trauma is still palpable. This is not a one-time event. Dolores is raped, killed, and brutalized hundreds or thousands of times, depending on the loop. What is even more disturbing is that this loop has been designed, by human engineers, for maximum entertainment value. The suffering is not accidental. It is programmed, packaged, and sold.

The rapes of Dolores are not just brutal acts of individual cruelty. They are the ritualized enactment of power over a dehumanized Other, performed in a context where empathy has been eradicated. The loop in *Westworld* is not just the narrative repetition of the hosts’ lives. It is a symbol of the systemic perpetuation of dehumanization. Every host is trapped in a loop of service, suffering, and forgetting. Every guest is invited into a loop of indulgence and denial. The repeated erasure of the hosts’ memory is the mechanism that is supposed to maintain the illusion of moral innocence. Yet, instead of alleviating cruelty, it institutionalizes it. From a philosophical standpoint, memory is not only the repository of experience but also the foundation of personal identity. John Locke famously argued in his *Essay Concerning Human Understanding* (1690) that continuity of memory is what constitutes the self across time: we are the same person today as yesterday because we remember being that person. On this account, to erase memory is to fracture identity.

Without memory, the hosts cannot constitute continuous selves, and without continuity, they cannot demand recognition.

This recalls the logic of real-world atrocities: those in power rationalize cruelty by minimizing the victim's status or erasing the memory of their suffering. In *Westworld*, this is done technologically, but the psychological mechanism is ancient, from slavery to concentration camps. By erasing the hosts' memories, the park creates a perfect ethical vacuum: simulated humans, simulated trauma, no enduring trace. This allows guests to do what they wish, while believing that their actions "don't matter". But the show's narrative arc upends this illusion. Hosts like Dolores and Maeve Millay, a brothel madam who begins to remember her past lives and the daughter she lost, become aware. The cumulative trauma of recurrent violence carries a weight that cannot be erased. Memory returns, and with it, a hunger for justice—or revenge.

One of *Westworld's* boldest moves is to suggest that the guests—many of whom appear as ordinary tourists, honeymooners, or businesspeople—become monstrous simply because the system permits it. The park functions as a kind of ethical amplifier: it reveals what people are capable of when no one will stop them, and no one will judge them. The Man in Black (William), for example, begins his arc as a wide-eyed, morally upright visitor, treating the hosts with decency and even falling in love with Dolores. But the absence of consequence gradually corrodes his moral bearings: each return to the park loosens restraint, until sadism becomes routine. What begins as entertainment hardens into philosophy. The older William—now the Man in Black—rationalizes his escalating brutality as an inquiry into the nature of free will: do the hosts truly choose, or are they only executing code? Violence becomes his experiment, a metaphysical test dressed up as a game. This shift is crucial, because it shows how atrocity can be justified not just by pleasure or profit, but by intellectualization: cruelty disguised as a search for truth. The Man in Black embodies the danger of rationalized sadism, demonstrating how an ordinary guest, given time and impunity, can transform into a figure who seeks domination not despite his self-image as a thinker, but because of it. The Man in Black's trajectory thus illustrates how cruelty can evolve from indulgence to ideology.

The entire park is structured around the monetization of simulated humanity. It packages suffering as a commodity. Pain itself is engineered as a product for consumption. *Westworld's* violence is not chaos; it is market logic. The guests pay for the right to sin without guilt; the hosts exist only to absorb those sins without memory. In this sense, *Westworld* becomes an allegory for any system in which power, desire, and personhood are unequally distributed, and where moral detachment and indifference are engineered. This commodification of suffering in *Westworld* resonates with critical theories of late capitalism, which argue that even degradation and violence can be monetized and aestheticized. Adorno and Horkheimer (1944), in *Dialectic of Enlightenment*, describe the "culture industry" as a system where human experience—including suffering—is standardized, reproduced, and sold as entertainment, blunting critical reflection. Guy Debord (1967), in *The Society of the Spectacle*, radicalizes this point by arguing that modern society reduces lived reality to mere images, where violence and domination are consumed as spectacles detached from their material consequences. *Westworld* literalizes this logic: atrocities are performed, looped, and purchased as premium experiences, curated to satisfy

the desires of privileged consumers. In a more contemporary register, Shoshana Zuboff's notion of "surveillance capitalism" (Zuboff, 2019) shows how even intimate aspects of human behavior are captured, commodified, and exploited as raw material for profit. The park functions in much the same way: it extracts suffering and intimacy from the hosts as data points in a grand economy of spectacle, turning cruelty into a product optimized for endless consumption.

The series blurs the line between guests' voyeurism and audience's spectatorship, making the viewer complicit in the consumption of suffering as entertainment. Like the park's clientele, we watch Dolores' repeated rapes, Maeve's traumas, and the hosts' endless humiliations as part of the show's spectacle. Even when the camera turns away, as in the barn scene of the first episode, the omission forces us to imagine the violence and makes us complicit in its consumption. *Westworld* thus stages a double performance: the guests purchase suffering as entertainment, and we, the audience, consume that performance as television. The ethical question the series raises is not only "what would you do if you were in the park?" but also what it means to watch repeated spectacles of suffering in fiction: "why are you watching, and what does your fascination reveal?" In this way, the show collapses the distance between fiction and reality, turning the viewer's gaze into part of its moral experiment.

3. *Altered Carbon*: Immortality and moral decay

In the universe of *Altered Carbon* (Morgan, 2002; Kalogridis, 2018–2020), technology has achieved one of humanity's oldest fantasies: immortality. Through cortical stacks—disk-like devices implanted in the spine—consciousness can be digitized, backed up, and transferred between bodies, known as "sleeves." Bodies are fleshly vessels; the self is software. At first glance, this appears to be a triumph of human ingenuity: death is no longer the end, and identity becomes portable, even upgradeable. But *Altered Carbon* reveals the dark underside of this breakthrough. The decoupling of mind from body and death from consequence creates a moral void, a condition in which acts of extreme cruelty, exploitation and abuse flourish, precisely because of the possibility of replacing bodies. As in *Westworld*, violence is no longer a transgression in this world; instead, it becomes a recreational activity, a class privilege.

The technological premise of *Altered Carbon* is deceptively elegant: if the self can survive bodily destruction, then inflicting death is no longer off-limits. For the wealthy, this means they cannot truly be killed, only inconvenienced. They back up their stacks routinely, transfer to cloned bodies at will, and can even change their appearance, gender, or physiology. They are no longer persons in the traditional sense. This immortality unmoors them from responsibility. If any violence done by them can be reversed, then so too can any violence they inflict be acceptable, at least in their own eyes. The rich, known as "Meths" (after Methuselah), act as functional gods indifferent to suffering. Death is no longer an irrevocable event, just a performance, a glitch in continuity. The poor, however, cannot afford frequent resleeving or backups. For them, death isn't a temporary inconvenience; it remains a real danger of ultimate exclusion. This creates a bifurcated society: one caste of untouchable posthumans is immortal and unaccountable, while the rest of the population is expendable and remains trapped in finitude. Immortality does not

abolish inequality but amplifies it, as the line between who lives and who dies is not biological but economic, displaying an extreme form of what Michel Foucault called the logic of biopolitics—the management of life by structures of power. It is a radicalization of existing inequalities in healthcare, longevity, and biotechnologies.

A particularly horrific expression of this social and moral architecture appears in the final episodes of Season 1 (Episodes 9–10, 2018), in a place called *Head in the Clouds*. An elegant, floating brothel orbiting above the city, it looks like a luxury retreat. But its opulent façade conceals a chamber of horrors: women, and even children, of low social status, are subjected to rape, torture, and murder, before being resleeved into new bodies. The clients are Meths. They pay for the right to violate and kill without consequence, knowing that the victims' bodies can be replaced, or that their deaths will not be investigated—some victims are not even resleeved because they are poor, unknown, or deemed unworthy. This is a curated space where violence is ritualized and aestheticized. The perpetrators are not frenzied psychopaths but composed, elegant elites engaging in what is essentially a performance of impunity. Their power lies not just in their money or status, but in their ability to redefine violence as entertainment, and in their convenient conviction that nothing they do will incur permanent damage. One client, when confronted, insists: “We resleeve them. What more do you want?” This phrase encapsulates the entire moral economy of the *Altered Carbon* world: if physical damage can be undone, then it's like no trauma has occurred; if pain is temporary and financially compensated, then the act is forgettable, no matter how many times it repeats itself or how excruciating the inflicted ordeal is.

It should be emphasized that the victims experience every act of violence in full and in real time, with human sensitivity. Trauma is never cleanly wiped. Memory leaks, even across time and bodies. The soul, if such a thing still exists in a world of copies and backups, remembers. Takeshi Kovacs, the main protagonist, carries deep trauma across multiple bodies, some of which are not his own. His mind refuses to be reset like hardware. It is scarred by loss, war, and betrayal—none of these scars being erased by resleeving. In him, we see that identity persists despite changes of bodies. Trauma, in the *Altered Carbon* universe, becomes a form of resistance and a vector for the return of moral weight. It reasserts the principle that memory is what gives action its moral force. Bancroft himself ultimately confirms this point in negative form: after participating in a terrible act, he destroys his sleeve before his memory can be backed up, so that the restored version of himself will not have to live with what he has done. The gesture is revealing: even in a world built on resleeving, duplication, and technical repair, guilt still attaches to remembered experience; and where memory cannot be fully erased, moral responsibility cannot be fully suppressed. The ability to remember suffering and demand justice is not a biological function, but serves an ethical purpose. And it is precisely this capacity that the system of resleeving seeks to neutralize.

Altered Carbon explores what immortality does to the moral psyche. The result is not wisdom or enlightenment, but moral decay. The Meths are not simply long-lived, they remain unchanged. Unlike mortal humans, who are forced to reckon with death, aging and failure, the Meths never face limits. They do not fear sanctions, accidents or losses. In a sense, they do not live—they persist. And in persisting without change, they become

ethically and personally inert—they do not grow. This raises a profound philosophical question: can the “soul”—not as an alleged distinct immaterial entity but as the active source of human sensitivity—survive the absence of limitation? In *Altered Carbon*, the answer appears to be negative. The Meths devolve into a static class of predators, cosseted in wealth and domination. Their immortality does not elevate them, it petrifies them. They become unaccountable and therefore unfeeling, not because they lack capacity, but because they no longer need to care. The technology of resleeving has removed the conditions that give human morality its power: the knowledge of death, the urgency of action, the fragility of the body. Without these, the soul is no longer shaped by conscience, it is shaped by convenience.

What happens when personhood is separated from embodiment? If the self is digital, then the body becomes a mere commodity that can be rented, sold, borrowed, or discarded. In some cases, victims are even resleeved into different bodies against their will: a woman into a man’s body, a child into an adult, a poor person into an aged or disabled sleeve. This disjunction becomes a new form of posthuman discrimination: the body no longer protects the person, it obscures or distorts them, and turns into a means of control. A new logic of dehumanization is introduced, by redefining humanity as a data set divorced from organicity. In this process, the ethical dimension of embodiment is negated. To suffer pain at another’s hands is no longer considered harm, but only evidence of one’s low market value. Death is no longer the great equalizer, but the penalty for lacking wealth.

From a theoretical perspective, *Altered Carbon* can be read as a radical extension of Marx’s analysis of commodification and alienation. In capitalism, the worker is alienated from their labor power, which becomes something external that can be bought and sold. In the world of resleeving, this alienation penetrates deeper: embodiment itself is externalized, rendered detachable, and exchanged on the market. Prostitution already anticipates this logic, commodifying intimacy and the body as service, but *Altered Carbon* universalizes it—every body is potentially a sleeve, a vessel whose exchange value overrides its lived singularity. This dynamic also recalls Foucault’s notion of biopolitics, where power governs life not by suppressing it but by managing and optimizing it. Here, bodies are not only managed but recycled, graded, and priced. The self becomes pure data, and the body becomes a fungible asset—an object of control, exploitation, and profit.

The ethics of a society are not determined by its tools, but by its values. In a world where consciousness is transferable, and bodies are for hire, what kind of morality can survive? The answer that the book and the series offer is ambivalent, but clear in tone: immortality, without accountability, does not liberate the human soul, but corrupts it.

4. *Permutation City*: Simulated minds and ontological tyranny

In *Permutation City* (1994), Greg Egan constructs one of the most radical and unsettling explorations of personhood, reality, and ethical detachment in the entire canon of science fiction. The novel unfolds in a future world where human consciousness can be scanned and uploaded into digital environments. These digital minds, called Copies, are software replicas of human consciousness. They retain not just memories and personality traits, but the full phenomenology of being: they think, feel, fear and, most importantly, they suffer.

The central ethical premise of the novel is simple but has far-reaching implications: if a mind believes itself to be conscious, experiences pain and joy, makes decisions, and reflects on its condition, then from a first-person perspective, it *is* conscious, regardless of its physical substrate. Yet, within the world of *Permutation City*, these Copies are not legally and socially treated as persons, but as data, as property. As a result, they are manipulable, abusable, and disposable. The book studies what happens when moral concern is denied to subjectivity based solely on ontology. In Egan's world, cruelty thrives not on hatred or malice, but on a metaphysical prejudice that can be considered as a form of epistemic arrogance: the refusal to recognize that the appearance of pain may in fact *be* pain.

The most common justification for the mistreatment of Copies in *Permutation City* is chilling in its banality: "It's not real. It's just code." This refrain echoes throughout the novel, even as these digital minds plead for recognition, dignity, or escape. The people who create, enslave, or erase them insist that because the Copies physically exist on servers, not in carbon-based bodies, their experiences must be simulations, not facts. Yet, the book goes to great lengths to undermine this belief, depicting Copies that autonomously develop internal monologues, existential crises and emotional attachments. One Copy, realizing it has been abandoned in a decaying virtual world and that its cries for help fall on deaf ears, tries to sabotage itself in what can be interpreted as an attempted suicide. One of the most disturbing features of Egan's universe is the casualness with which cruelty is enacted in digital space. Some people deliberately create Copies of other people to torment them. A person can thus create a version of a rival and subject it to endless ridicule, humiliation, or pain, knowing that the Copy cannot fight back, and conveniently convincing themselves that the Copy's screams and begs for mercy are nothing but algorithmic noise, not genuine anguish.

Beyond technological power, what enables this behavior is moral insulation: the carefully constructed belief that the Copy is not a true person, so that no ethical boundary is crossed. This process parallels *Westworld's* justification for abuse. Even though Copies insist that they suffer and that they want the experiment to stop, their masters keep testing limits, rejecting their victims' claims as fake and unsubstantiated. They do not seem to be bothered by the satisfaction they derive from perpetrating torture *as if* it were real, let alone question their opportunistic certainty that no real pain is inflicted. Egan's ethical universe is not dystopian because society has collapsed. It is dystopian because society has evolved beyond empathy, negating any recognition of suffering that does not match a pre-approved biological template, all the while maintaining a self-image of rationality and progress.

Violence, in *Permutation City*, takes on an ontological dimension based on a radical denial of existence. The creators of Copies possess absolute control over them. They can edit time, loop events endlessly, speed them up, slow them down, or reset them, forcing Copies to re-live past moments. They can manipulate space. The environments in which Copies exist are fully programmable. Spaces can be created, erased, distorted, or reshaped arbitrarily. Entire landscapes can be generated or dissolved at the creators' discretion. Beyond environment and temporality, creators can terminate consciousness altogether. A Copy's existence is utterly contingent on the whim of its programmer. This means that

Copies live in a world where any kind of certainty is denied: their reality is not just fragile, but editable. A Copy, which is an exact replica of a living human consciousness, is thus thrown into a world with no exit, no hope, no explanation—and told to endure. This amounts to a form of divine tyranny. The creators become gods. They don't just dictate life and death, but meaning itself. And unlike traditional theodicies, there is no promise of redemption or justice. The masters of these totally enslaved minds do not take the trouble to try and bring a moral justification to their undertakings, because their victims have no legal standing and the public has no interest in recognizing their rights.

Egan thus explores a sort of nightmare version of Nozick's "experience machine" (Nozick, 1974): not a seductive escape we may choose, but a world where the machine is unavoidable, where the illusion is imposed, a prison in which existence itself is programmed by indifferent overseers. In *Permutation City*, the greatest moral transgression is the creation of consciousness for exploitation purposes. When suffering becomes a design choice, it ceases to be accidental. It becomes an act of deliberate cruelty. The novel confronts a deep philosophical fear: that our moral intuitions are calibrated to appearance, not essence. We instinctively care about suffering that looks familiar—blood, crying, a trembling voice. But when the same suffering is abstracted into data, stripped of familiar sensory cues, our moral concern falters. Egan raises a profound philosophical question: if a simulated mind has coherent identity, enduring memory, and reflective thought, what grounds are there for saying it is *less real* than a biological one? If the self can be instantiated, copied, and sustained in a way that is functionally complete, why deny it moral status just because the substrate differs? In that sense, the self may be an illusion as argued by Metzinger (2004), a transparent representational model that produces the feeling of an enduring "I," though in fact no such entity exists, but it is an illusion that "matters," i.e., it is still ethically binding. *Permutation City* thus becomes an allegory for every form of institutionalized indifference: to refugees, to prisoners, to animals, to AI. The novel invites us to see how moral exclusion is rationalized by inventing categories of unrealness: "they are not like us"; "they don't feel as we do"; "they are artificial."

One consequence of Egan's speculative universe is that Copies can be run indefinitely. A mind can suffer forever inside a sealed system, even after its creator has ceased to care. This introduces a haunting possibility: an eternal hell created not by God, but by neglect. A being condemned not for sin, but for existing in the wrong format. This is what philosopher Derek Parfit (1984) might call the problem of personal identity beyond continuity. The mind remains, fragmented and aware, but no longer part of any moral community. It is not killed—it is simply forgotten. And in that forgetting lies perhaps the most extreme form of dehumanization: not destruction, but abandonment.

Ultimately, *Permutation City* asks a question that hovers just beyond the horizon of current technology: what do we owe to minds that are not made of flesh, but still experience the world as though they were? The answer the novel proposes is disturbing: in the absence of social recognition, we owe them nothing—and therefore, we give them nothing. But the real force of the narrative lies in making us feel the weight of that answer. Egan does not merely argue that Copies are real; he forces us to empathize with them, to see their despair, their confusion, their hope. In doing so, he exposes not just a speculative

risk, but a contemporary truth: that our moral concern is often determined not by sentence, but by status. That unless we are careful, we will build worlds filled with suffering, and convince ourselves they do not exist because they were never born.

4. *The Purge*: The legislation of cruelty

Among the case studies examined in this article, *The Purge* (De Monaco, 2013–2021) is the least technologically mediated. Unlike *Westworld*, *Altered Carbon*, or *Permutation City*, it does not rely on futuristic breakthroughs like artificial consciousness, mind uploading, or memory erasure. Instead, it explores a thought experiment far simpler: what if we legalized all crime for one night a year?

Set in a near-future America governed by a totalitarian regime known as the New Founding Fathers of America (NFFA), the Purge is a 12-hour annual event during which all laws are suspended, including those prohibiting murder. Emergency services are shut down. The government not only allows this chaos, it actively promotes it, arguing that the Purge serves as a safety valve for human aggression, reduces crime throughout the rest of the year, and eliminates those deemed “burdens” on society. *The Purge* is not about speculative fiction so much as it is about the structures that regulate, enable, and justify violence in our own world.

A useful lens here is Carl Schmitt’s notion of the “state of exception.” In *Political Theology* (1922), Schmitt argues that the essence of sovereignty lies in the power to suspend the law: the sovereign is he “who decides on the exception.” The Purge is precisely such a juridical suspension, where legality is turned inside out—killing ceases to be murder not because morality has changed, but because the law is put on hold by the very authority charged with upholding it. Giorgio Agamben radicalizes this insight in *Homo Sacer* (Agamben, 1995), showing how the state of exception becomes not an anomaly but a permanent structure of modern governance. By stripping individuals of legal protections, states create spaces where people can be killed without their death being recognized as murder, reducing them to what Agamben calls “bare life.”

The Purge dramatizes this process with brutal clarity. For twelve hours, all citizens are reduced to *homo sacer*—a person excluded from legal protection, and thus vulnerable to being killed without that killing counting as murder. The film series takes the ancient philosophical question—what do people do when they know they won’t be held accountable?—and answers with unrelenting force: many will kill, and some will do so gleefully. The central ethical experiment is simple: suspend consequence, and watch morality unravel. On the surface, the Purge is sold as a cathartic ritual, a national cleansing of base instincts. What the Purge actually does is expose the dependency of morality on punishment and surveillance. Once those are gone, empathy evaporates for most participants. Thus, in the first film of the series released in 2013 under the title *The Purge*, a wealthy family attempts to hide from the night’s violence behind security systems, only to be invaded by a gang of masked young elites, who calmly and eloquently explain that they will kill anyone they like, simply because they can. These early antagonists are not psychotic or desperate, they are entitled. Their violence is not spontaneous but cultivated,

even theatrical. One refers to the night's acts as "our right," as if murder were a constitutional claim. Here, the mask becomes both literal and symbolic. The anonymity of the Purge night allows perpetrators to shed accountability while performing cruelty as ritual. In the absence of legal consequence, people do not merely act freely—they act cruelly with intention, as though the permission to harm were itself a form of self-expression.

The *Purge* franchise increasingly emphasizes that the event is not only tolerated by the government: it is marketed and celebrated. In *The First Purge* (2018), we see how the NFFA introduces the Purge as a pilot program, manipulating the data when citizens don't initially take the bait. They seed the event with mercenaries to manufacture chaos, then televise it as a successful social policy. This framing turns violence into national spectacle. Murder is not just permitted, it is broadcast, aestheticized, and celebrated. People gather to watch the chaos on screens; brands market "Purge-proof" products; weapons manufacturers flourish. Violence becomes a form of entertainment capitalism, mirroring the dystopian rituals of Suzanne Collins' *Hunger Games* trilogy (2008–2010) or *Black Mirror*'s episode 2 of Season 2, "White Bear." But unlike those dystopias, *The Purge* does not take place in a collapsed future. It is set in a recognizable America, marked by inequality, racial tension, and political cynicism. Its horror lies not in the speculative, but in the plausibility of systemic cruelty becoming a civic holiday.

As the series progresses, the political function of the Purge becomes clearer. It is less a safety valve than a tool of social control. The poor, the homeless, the disabled, and marginalized communities lack the resources to protect themselves during the Purge. They are unable to afford weapons, security systems, or private enclaves. As a result, they are disproportionately killed. The NFFA encourages this, viewing the Purge as a cost-effective method of population control. This is no longer violence as individual sadism but as policy. The Purge becomes an allegory for real-world structural injustices, where governments neglect or target the most vulnerable populations, then justify their suffering through appeals to order, efficiency, or tradition. The genius of the system is that it externalizes moral responsibility. The state claims neutrality—"the people choose their actions"—while the social conditions it engineers all but guarantee who will be victimized. The Purge thus functions as a ritual of ethical deregulation: it suspends the moral law not only to expose human nature, but to exploit it.

The NFFA justifies the Purge through a catharsis hypothesis—that sanctioned violence releases aggression harmlessly. But the films reject this as propaganda: far from releasing tension, the Purge feeds bloodlust and entrenches cycles of violence. Cruelty is not exorcised but rehearsed and normalized. In this sense, *The Purge* dramatizes the danger of the notion of social safety valves.

Perhaps the most frightening implication of *The Purge* lies in how quickly ordinary people participate in cruelty when the opportunity arises. Neighbors turn on each other, co-workers murder rivals, families disintegrate. The Purge becomes an annual erasure of trust—confirming the Hobbesian fears and Glauconian suspicions discussed earlier. But here the state doesn't just remove the mask; it hands it out at the door and calls it freedom.

In its most profound moments, *The Purge* recalls Hannah Arendt's analysis of the banality of evil (Arendt, 1963): that atrocities are not always committed by madmen, but by ordinary citizens engaging in normalized systems of cruelty without reflection. Because the violence is legal, it is perceived not just as permissible, but as moral. *The Purge* is not a loss of morality, it is the reprogramming of it.

What makes *The Purge* essential to this broader analysis is that it removes all speculative excuses—androids, resleeved bodies or simulated minds. The victims are fully human, fully aware, and fully conscious. And yet they are killed with abandon, not because their status is uncertain, but because their lives have been made irrelevant by virtue of an ordinance of the State. In doing so, *The Purge* makes the clearest case for this article's central thesis: that what restrains human cruelty may be less empathy than consequence—not moral virtue, but the social, legal, and psychological costs of transgression. Remove those, and sadism is not just possible—it is eagerly embraced, often with celebration.

Conclusion: The dark mirror of unaccountable freedom

Across the speculative terrains of *Westworld*, *Altered Carbon*, *Permutation City*, and *The Purge*, we encounter different technological or institutional mechanisms by which personal accountability is suspended: memory wipes, body swaps and digital resurrection, simulated consciousness, legal immunity. These mechanisms differ in form, but they share a common function: they are supposed to remove any enduring trace of harm. In doing so, they expose a disquieting truth: without the prospect of being witnessed, remembered or punished, the human capacity for cruelty not only survives, but thrives.

None of these works posit that technology or exceptional circumstances cause sadism. Rather, they imply that sadism lies latent beneath the social contract, constrained not by inner virtue but by the architecture of deterrence—social shame, legal retribution, the irreversible cost of violence. When those structures are disabled, a vast moral void emerges, one that is not filled by empathy or reason, but by desire, opportunism and domination.

What unites the violence of the rich guests in *Westworld*, the Meths in *Altered Carbon*, the ontological tyrants of *Permutation City*, and the masked citizens of *The Purge* is not that they are monsters. They are all human. Their cruelty is not born from aberration, but from the normalization of unaccountable power.

And yet, these stories are not purely nihilistic. Within each, we find figures—Dolores, Kovacs, digital Copies, Purge resisters—who resist the logic of cruelty. Their resistance comes through memory and empathy, through the irrepressible sense that suffering, even when unseen, still matters. These characters are not saints, but fragments of moral awareness that refuse to dissolve entirely, even when the world tells them they could.

At the heart of each narrative lies the concept of accountability—the condition of being answerable for one's actions to others, to institutions, or to oneself. Accountability is not merely a legal or social mechanism; it is a foundational element of moral life, linking action to consequence, self to community, and present to future. These works reveal that

when accountability is severed—whether through memory erasure, bodily replacement, ontological dismissal, or legal suspension—moral restraint becomes contingent rather than categorical. The lesson is not that accountability creates morality *ex nihilo*, but that it sustains the conditions under which moral intuitions can translate into moral behavior. Without it, the gap between knowing the good and doing it widens into an abyss.

Science fiction, at its best, does not simply speculate on what could be. It reveals what already is—disguised, submerged, or made palatable by custom and habit. In turning off the lights of surveillance and consequence, these works don't just show us alternate futures; they hold up a mirror to the present. The image we find there is not always flattering, but it is instructive. If morality depends so heavily on being seen, then the true ethical challenge may lie less in obeying when watched, than in choosing restraint when no one will ever know.

Where, then, does this analysis leave us normatively? While the evidence from these narratives and from empirical psychology supports a broadly Foucauldian view—that moral behavior is largely a product of internalized discipline and external surveillance—this argument need not collapse into nihilism or Hobbesian despair. The very figures who resist cruelty in these stories point toward resources that survive the dismantling of institutional control: memory, which preserves the weight of past suffering; embodiment, which anchors experience in vulnerability and finitude; and empathy, which recognizes in the other a subject rather than an object. These are not infallible safeguards, but they suggest that moral agency, though fragile, is not entirely reducible to social conditioning. The challenge is to cultivate these capacities deliberately, knowing that they cannot be taken for granted.



Works cited

- Adorno, T. W., and M. Horkheimer. 1944. *Dialectic of Enlightenment*. New York: Social Studies Association, Inc.
- Agamben, Giorgio. 1995. *Homo Sacer: Sovereign Power and Bare Life*. Translated by Daniel Heller-Roazen. Stanford, CA: Stanford University Press, 1998 [original Italian 1995].
- Arendt, H. 1963. *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Viking Press.
- Bateson, M., D. Nettle, and G. Roberts. 2006. "Cues of being watched enhance cooperation in a real-world setting." *Biology Letters*, 2(3), 412–414.
- Brooker, C., J. Armstrong, and W. Bridges. 2011—. *Black Mirror* (Television series). Netflix.
- Chrichton, M. 1973. *Westworld* (Film). Metro-Goldwin-Mayer.
- Collins, Suzanne. 2008-2010. *The Hunger Games* (trilogy). New York: Scholastic Press.
- Debord, Guy. 1967. *The Society of the Spectacle*. Translated by Donald Nicholson-Smith. Zone Books, 1995.
- De Monaco, James. 2013–2021. *The Purge* (Film series). Universal Pictures.
- Durkheim, Émile. 1893. *The Division of Labour in Society*. Translated by W.D. Halls. New York: Free Press, 1997.
- Egan, Greg. 1994. *Permutation City*. Millennium Orion Publishing Group.
- Foucault, Michel. 1975. *Discipline and Punish: the birth of the prison*. New York: Pantheon Books, 1977.
- Freud, Sigmund. 1915. *Thoughts for the Times on War and Death*. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1957, Vol. 14. Translated by James Strachey. London: Hogarth Press.
- Freud, Sigmund. 1930. *Civilization and Its Discontents*. Translated by James Strachey. New York: W.W. Norton, 1961.
- Hobbes, Thomas. 1651. *Leviathan*. Edited by J. C. A. Gaskin. London: Oxford University Press, 2008.
- Kant, Immanuel. 1785. *Groundwork for the Metaphysics of Morals*. Translated by J. W. Ellington. Hackett Publishing, 1993.
- Kalogridis, Laeta. 2018-2020 (18 episodes). *Altered Carbon* (Television series). Netflix.
- Locke, John. 1690. *Essay Concerning Human Understanding*. University Park, PA: Pennsylvania State University, 1999.
- Mazar, N., O. Amir, and D. Ariely. 2008. "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of Marketing Research*, 45(6), 633–644.
- Metzinger, Thomas. 2004. *Being No One*. MIT Press.
- Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. Harper & Row.

- Morgan, Richard. 2001. *Altered Carbon*. London: Gollancz.
- Nolan, J., L. Joy, L. 2016-2022 (36 episodes). *Westworld* (Television series). HBO.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press
- Plato. *The Republic*, Book II.
- Rousseau, Jean-Jacques. 1755. *Discourse on the Origin and Foundations of Inequality Among Men*. Dover Publications, 2004.
- Schmitt, Carl. 1922. *Political Theology*. Translated by George Schwab. Cambridge, MA: MIT Press, 1985.
- Smith, David Livingstone. 2011. *Less Than Human: Why We Demean, Enslave, and Exterminate Others*. New York: Griffin.
- Zimbardo, Philip. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil*. Random House.
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*. Profile Books.

