Fall 12-22-2023

Role-Reversibility, AI, and Equitable Justice - Or: Why Mercy Cannot Be Automated

Stephen E. Henderson

Kiel Brennan-Marquez

0091-4169/23/1140Online-0001
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY Vol. 114 Online

Printed in U.S.A.

ROLE-REVERSIBILITY, AI, AND EQUITABLE JUSTICE— OR: WHY MERCY CANNOT BE AUTOMATED

KIEL BRENNAN-MARQUEZ & STEPHEN E. HENDERSON*

A few years ago, we developed the concept of "role-reversibility" in AI governance: the idea that it matters whether a party exercising judgment is reciprocally vulnerable to the effects of judgment. This idea, we argued, supplies a deontic reason to maintain certain spheres of human judgment even if (or when) truly intelligent machines become demonstrably superior in every utilitarian sense. While computer science remains far from that holy grail, generative AI is raging through systems as diverse as healthcare, finance, advertising, law, and academe, making it imperative to further shore up our claim. We do so by situating role-reversibility within the long arc of criminal justice philosophy, from Anaximander to Aristotle to Seneca. Simply put, role-reversibility facilitates mercy. And mercy is both (1) central to the operation of a humane legal system and (2) impossible, even in principle, to automate.

> [T]he horrible thing about all legal officials, even the best, about all judges, magistrates, barristers, detectives, and policemen, is not that they are wicked (some of them are good), not that they are stupid (several of them are quite intelligent), it is simply that they got used to it. Strictly they do not see the prisoner in the dock; all they see is the

* Kiel Brennan-Marquez is a Professor of Law, the William T. Golden Scholar, and Faculty Director of the Center on Community Safety, Policing and Inequality at the University of Connecticut; Stephen E. Henderson is the Judge Haskell A. Holloman Professor of Law at the University of Oklahoma. We are grateful to Amin Ebrahimi Afrouzi for engaging with our work, to *The Journal of Criminal Law & Criminology* for its continued interest in the same, to Guha Krishnamurthi and Kelly Sorensen for comments upon drafts, and to Shawnda Henderson and librarian Jacob Black for exceptional research assistance.

> usual man in the usual place. They do not see the awful court of judgment; they only see their own workshop.
>
> – G. K. CHESTERTON, *The Twelve Men*, in TREMENDOUS TRIFLES 85–86 (1909)
>
> The mood and temper of the public in regard to the treatment of crime and criminals is one of the most unfailing tests of the civilization of any country. A calm and dispassionate recognition of the rights of the accused against the State, and even of convicted criminals against the State, a constant heart-searching by all charged with the duty of punishment, a desire and eagerness to rehabilitate in the world of industry all those who have paid their dues in the hard coinage of punishment, tireless efforts towards the discovery of curative and regenerating processes, and an unfaltering faith that there is a treasure, if you can only find it, in the heart of every man—these are the symbols which in the treatment of crime and criminals mark and measure the stored-up strength of a nation, and are the sign and proof of the living virtue in it.
>
> – Winston Churchill, Speech to House of Commons (July 20, 1910)

INTRODUCTION

When we asked ChatGPT whether it 'knew' anything about role-reversibility, it trotted out a reasonable answer relating to inverted roles, such as that of child-to-parent as a parent ages.[1] So, we prodded a bit more directly,

---

[1] ChatGPT is of course a chatbot implementing generative AI that was released to the public in late 2022. *See Introducing ChatGPT*, OPENAI (Nov. 30, 2022), https://perma.cc/DC3A-J23Q. Our particular question, asked on March 1, 2023, was, "Do you

inquiring about a concept in the philosophy of criminal law, and were gratified that it responded with something more akin to our theory.[2]

That theory in a nutshell is this: for certain categories of decision-making—including, paradigmatically, criminal judgment—decisions are intrinsically superior if they minimize the situational differences between the judging party and the judged party.[3] When a judgment is role-reversible, but for a relatively small series of contingent events the party exercising judgment could be the one being judged, and the one being judged could be exercising the judgment. So understood, role-reversibility solemnizes and democratizes the process of judgment; when the judging party internalizes the idea that 'but for the grace of fortune there go I,' she is required, in essence, to judge herself.[4] And that is the very heart of democratic self-rule. It makes judgment fraught with uncertainty and even pain, but those qualities,

---

know anything about role reversibility?" ChatGPT's answer began like this (a transcript of the entire chat is on file with the authors):

> Yes, I am familiar with the concept of role reversibility. It refers to the idea that in any social interaction or relationship, the roles played by the individuals involved can be reversed or interchanged. For example, in a parent-child relationship, the parent is typically the caregiver and the child is the dependent. However, in certain situations, such as when the parent becomes elderly or disabled, the child may take on the role of caregiver, reversing the traditional roles.

Because ChatGPT is generative AI, its answer will change at least slightly over time even if remaining substantively the same, as would a human being's. If you are curious, you can visit the site and try the same query. Of course, then your thoughts will similarly become part of the machine. *See* Christopher Mims, *For Chat-Based AI, We Are All Once Again Tech Companies' Guinea Pigs*, WALL ST. J. (Feb. 25, 2023), https://perma.cc/E44M-FP6D.

[2] *See generally* Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, 109 J. CRIM. L. & CRIMINOLOGY 137 (2019). Not that ChatGPT quite made it to where we'd like. It focused on, for example, a juror *imagining* herself in the defendant's position (again, a chat transcript is on file with the authors), whereas our theory critically depends upon the juror being life-situated as near as possible to the defendant. *See id.* And when we requested citations to scholarship, well, things really went off the rails. ChatGPT repeatedly hung, and when it finally claimed citations, they weren't on point; when we specifically requested citations to our work, it alleged scholarship that simply does not exist. So it goes when a contemporary generative AI pulls one part of an answer from here and another from there; it doesn't *know* what a citation is in what we humans would consider a meaningful, reflective sense. And that can get folks in trouble who don't understand the technology's limitations. *See* Justin Wise, *Lawyer's AI Blunder Shows Perils of ChatGPT in 'Early Days'*, BLOOMBERG LAW, May 31, 2023, https://perma.cc/6RDF-6GW7 ("New York lawyers Steven Schwartz and Peter LoDuca face a June 8 hearing on potential sanctions after a court brief they submitted cited six nonexistent cases.").

[3] *See* Brennan-Marquez & Henderson, *supra* note 2 at 149–52.

[4] For the theist it might be John Bradford's "but for the grace of God there go [I]." THE WRITINGS OF JOHN BRADFORD, M.A. xliii (Aubrey Townsend ed., 1853), https://perma.cc/HA8J-TKM6. For the non-theist it might be Phil Ochs' "There but for fortune go . . . I." PHIL OCHS, *There But For Fortune*, *on* NEW FOLKS VOL. 2 (Vanguard 1964).

we believe, are features, not bugs.[5] In fact, we believe that role-reversibility is a strong enough normative principle, anchored in democratic equality, that it supplies a reason to retain human involvement in decision-making *even if* truly intelligent machines, meaning machines achieving the holy grail of "artificial general intelligence," become a viable option.[6]

Amin Ebrahimi Afrouzi has now grappled with our work, for which we are grateful.[7] And despite only four years having passed since publication of our article, today we are substantially more awash in daily interactions with AI,[8] including the currently-novel experience with generative large language models like that of ChatGPT.[9] For the moment, these tools mostly remain fodder for curious professors wondering whether their work is synthetically known,[10] and for students completing (and perhaps cheating on) their assignments.[11] But it is estimated that generative AI will produce 10% of all data by the year 2025,[12] and the technology is already being used in everything from the detection of financial frauds to education to computer coding to advertising to lawyering to healthcare, and to much else besides.[13]

---

[5] *See* Brennan-Marquez & Henderson, *supra* note 2 at 152–54.

[6] *See id.* at 143–45.

[7] *See* generally Amin Ebrahimi Afrouzi, *On Role-Reversible Judgments and Related Democratic Objections to AI Judges*, J. CRIM. L. & CRIMINOLOGY ONLINE (2023).

[8] *See, e.g.*, BRIAN KENNEDY, ALEC TYSON & EMILY SAKS, PUBLIC AWARENESS OF ARTIFICIAL INTELLIGENCE IN EVERYDAY ACTIVITIES (Pew Research Center 2023) (reporting on a survey of 11,004 US adults regarding awareness of, and comfort with, daily AI interactions).

[9] "Generative AI is a form of AI that learns a digital representation of artifacts from sample data and uses it to generate new, original, realistic artifacts that retain a likeness to the training data but don't repeat it." DAVID GROOMBRIDGE, TOP STRATEGIC TECHNOLOGY TRENDS FOR 2022, at 12 (Gartner 2021). *See also* McKinsey & Company, *What is Generative AI?* (Jan. 19, 2003), https://perma.cc/6EH3-GNVE.

[10] Or curious what an image generator like MidJourney (https://www.midjourney.com) will do with a request like "antonin scalia as a cute monkey during a snow storm." (Result on file with authors.) *See About*, MIDJOURNEY, https://perma.cc/Q3DS-6YZ5 (archived Oct. 30, 2023).

[11] *See* Karen Hao, *What is ChatGPT? What to Know About the AI Chatbot*, WALL ST. J. (Updated May 16, 2023), https://www.wsj.com/articles/chatgpt-ai-chatbot-app-explained-11675865177 ("Some schools have blocked access to the service on their networks to stave off cheating, while others are actively encouraging students to use the tools ethically.").

[12] *See* Groombridge, *supra* note 9, at 12; *see also* Giancarlo Frosio, *The Artificial Creatives: the Rise of Combinatorial Creativity from Dall-E to GPT-3* in HANDBOOK OF ARTIFICIAL INTELLIGENCE AT WORK: INTERCONNECTIONS AND POLICY IMPLICATIONS (Martha Garcia-Murillo, Ian MacInnes & Andrea Renda eds., Edward Elgar) (forthcoming).

[13] *See* Cem Dilmegani, *Top 70+ Generative AI Applications/Use Cases in 2023* (Updated Mar. 20, 2023), https://perma.cc/B2E5-EBMM; *see also* Daniel Martin Katz et al., *GPT-4 Passes the Bar Exam* (Mar. 15, 2023), http://dx.doi.org/10.2139/ssrn.4389233; Tammy Pettinato Oltz, *ChatGPT, Professor of Law* (Feb. 4, 2023),

In the words of Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher, "Generative artificial intelligence presents a philosophical and practical challenge on a scale not experienced since the start of the Enlightenment."[14] And all of that is happening against the backdrop of a broader debate about the use of algorithms and artificial intelligence in public life.[15]

There is no better time, then, to revisit our role-reversibility claim—and Afrouzi's thoughtful critique provides a welcome opportunity to do so. Afrouzi agrees with us (at least arguendo) that delegating certain kinds of decision-making to powerful AI would flout the role-reversibility principle. He is skeptical, however, that the principle can bear the normative weight we have assigned it. He offers two related arguments in support of this view. We

---

http://dx.doi.org/10.2139/ssrn.4347630; Daniel Schwarcz & Jonathan H. Choi, *AI Tools for Lawyers: A Practical Guide*, 108 MINN. L. REV. HEADNOTES 1 (2023), https://perma.cc/38BB-ZSYT. In one study, "researchers found that at least half of accounting tasks could be completed much faster with the technology. The same was true for mathematicians, interpreters, writers and nearly 20% of the U.S. workforce." Lauren Weber & Lindsay Ellis, *The Jobs Most Exposed to ChatGPT*, WALL. ST. J. (Mar. 28, 2023), https://perma.cc/7DB2-V6HZ. By contrast, "The jobs that will be least affected by the technology include short-order cooks, motorcycle mechanics and oil-and-gas roustabouts." *Id.* Texas legislators are already considering the legal restriction of "artificial intelligence mental health services," *see* H. B. 4695, 88th Leg., Reg. Sess. (Tex. 2023), California legislators are concerned with "algorithmic discrimination," *see* A.B. 331, 2023–24 Reg. Session (Cal. 2023), and practicing lawyers are getting in on the tech as well, *see* Steven Lerner, *Forget The Future. Attorneys Are Using Generative AI Now*, LAW360 (Jan. 30, 2023), https://perma.cc/QY23-BRV8. Other recent articles on the diverse uses of such technology include Nidhi Subbaraman, *ChatGPT Will See You Now: Doctors Using AI to Answer Patient Questions*, WALL ST. J. (April 28, 2023), https://perma.cc/VHF9-4Z4B; Sarah A. Needleman, *How AI Is Building the Next Blockbuster Videogames*, WALL ST. J. (April 18, 2023), https://perma.cc/JXR7-QZUQ; Keach Hagey et al., *Publishers Prepare for Showdown With Microsoft, Google Over AI Tools*, WALL ST. J. (Mar. 23, 2023), https://perma.cc/7KZU-UCDM; Belle Lin, *Generative AI Makes Headway in Healthcare*, WALL ST. J. (Mar. 21, 2023), https://perma.cc/88LZ-S9BR; Kim S. Nash, *ChatGPT Helped Win a Hackathon*, WALL ST. J. (Mar. 20, 2023), https://perma.cc/8P6P-JN5N; Megan Graham, *Five Things Marketers Should Know About Generative AI in Advertising*, WALL ST. J. (Mar. 16, 2023), https://perma.cc/SM2d-ECGV.

[14] Henry Kissinger, Eric Schmidt, & Daniel Huttenlocher, *ChatGPT Heralds an Intellectual Revolution*, WALL ST. J. (Feb. 24, 2023), https://perma.cc/YV88-PKH4. *See also* Ryan Tracy, *Biden Administration Weighs Possible Rules for AI Tools Like ChatGPT*, WALL ST. J. (April 11, 2023), https://perma.cc/AZ25-HM3G.

[15] *See generally, e.g.*, Cass R. Sunstein, *The Use of Algorithms in Society* (rev. Mar. 19, 2023), http://dx.doi.org/10.2139/ssrn.4310137; CHRISTOPHER SLOBOGIN, JUST ALGORITHMS: USING SCIENCE TO REDUCE INCARCERATION AND INFORM A JURISPRUDENCE OF RISK (Cambridge U. Press 2021); Peggy Noonan, *A Six-Month AI Pause? No, Longer is Needed*, WALL ST. J. (Mar. 30, 2023), https://perma.cc/AJ5Q-LQYK; Jake Rudnitsky & Rachel Metz, *Musk, Tech Leaders Urge Halt to Training Powerful AI Systems*, BLOOMBERG LAW (Mar. 29, 2023), https://perma.cc/5Y6M-VTRQ; Bruce Schneier & Nathan Sanders, *We Don't Need to Reinvent our Democracy to Save it from AI* (Feb. 9, 2023), https://perma.cc/99CZ-XWXV.

will start by (1) summarizing both components of Afrouzi's critique and (2) explaining what we are—and are not—taking up in this reply.

I. CLARIFYING THE FAULT LINES

Afrouzi's first claim is that we have conflated a *formal property* of legal decision-making—role-reversibility—with its *substantive goals*. He bifurcates those goals as follows (and in broad strokes, we agree): first, whether decision-making "promotes evenhandedness and public justifiability of the content and application of the law," and second, whether it "promotes equality of rank between those judging and judged, thereby minimizing hierarchies and relations of domination between citizens, irrespective of social roles they happen to occupy."[16] These goals, Afrouzi argues, can—and should—be decoupled conceptually from the institutional mechanisms used to vindicate them. The role-reversibility condition, in his view, is an example of the latter. In the past, that condition may have been indispensable to securing substantively sound decisions; and it may continue, as an *empirical* matter, to be indispensable in the age of powerful AI. But for Afrouzi the reason role-reversibility matters is functional, not conceptual. If it matters at all, role-reversibility matters because it advances the substantive goals of legal decision-making. It has no "intrinsic" worth.

From there, Afrouzi's second claim is that once machines become capable of perfectly mimicking human decision-making, role-reversibility will no longer be necessary, as a functional matter, to effectuate the legal system's substantive goals. Rather, machines will be able—or at any rate, there is no reason to think they will be unable—to arrive at the equivalent of human decisions, even if the machines are not "reversible" with affected parties.

We thank Afrouzi for his careful and gracious engagement with our article, and we find both lines of counterargument fruitful. Although we continue to believe that role-reversibility, qua democratic equality, has intrinsic value—and that its integrity therefore constitutes a non-utilitarian good for a liberal legal order—we leave that question for future work.[17] In what follows, we take up Afrouzi's second challenge: to explain why role-

---

[16] Afrouzi, *supra* note 7, at 25.

[17] Suffice it to say that the future work along these lines will be indebted to Afrouzi. His argument, in particular, that the formal constraint of role-reversibility proves too much, insofar as it would preclude garden-variety examples of non-fully-reversible judgment (e.g., biological males in a case of pregnancy, or sighted persons in a case of blindness), is a powerful point. Going forward, one of our tasks will be to refine the formal texture of role-reversibility such that it still has bite as a constraint—but also sits in better reflective equilibrium with our practices.

reversibility, even if reconstructed in utilitarian terms, is necessary to secure (in Afrouzi's words) "public justifiability of the content and application of the law."

Our answer focuses on mercy. We explore the idea that mercy is necessary to the operation of a humane legal system, because it infuses decision-making with a 'there but for the grace of God go I' quality that not only softens the hard edges of rules in particular cases, but also keeps the legal system, writ large, from becoming an inhuman construct. This is, in Afrouzi's terms, a "substantive goal." And its vindication requires a dynamic of relational imagination between the party exercising judgment—and deciding whether or not to exercise mercy—and the affected party. In other words, it requires role-reversibility.

To shore up this idea, we situate it within the long arc of criminal justice philosophy, looking particularly to three ancient thinkers: Anaximander, Aristotle, and Seneca. As Martha Nussbaum has demonstrated, their works can be seen to describe an increasingly rich notion of human justice,[18] and role-reversibility comfortably sits at their peak. It may be too strong to argue that role-reversible justice can thus be considered historically-developed wisdom. But such exegesis certainly supports the doctrine's worth, and that worth may prove critical as we make decisions in the coming age of advanced, even dominant, AI.

II. ANAXIMANDER

What is justice? This fundamental question has inspired philosophic and legal debate for as long as there have been humans to so engage, yielding diverse answers and correspondingly diverse systems of its criminal aspect. One possible notion looks at the 'intrusion' done by something and restores it in kind. This concept is not at all unique to criminal law, but rather might be a theory of nature itself. What heat takes from the cold, say, or wet from the dry, ought to be restored in kind. In the words of Martha Nussbaum,

> a human life . . . is a vulnerable thing, a thing that can be invaded, wounded, or violated by another's act in many ways. For this penetration, the only remedy that seems appropriate is a counterinvasion, equally deliberate, equally grave. And to right the balance truly, the retribution must be exactly, strictly proportional to the original encroachment.[19]

---

[18] *See generally* Martha C. Nussbaum, *Equity and Mercy*, 22 PHIL. & PUB. AFF. 83 (1993).

[19] *Id.* at 89.

On this view, judgment pushes back right in equal measure to every wrong. *Lex talion*'s life for a life, eye for an eye.[20]

There is much to like in a world of such 'recalibrative' or 'restorative' justice: there is no intrusion—no wrong—that simply does not matter, blithely 'forgiven' without a care. In such a world, William Faulkner's Quentin Compson could have lived at peace; every wrong is counter-invaded.[21] And that may explain why recalibration as a theory of justice has very deep roots, with one of its first proponents perhaps arguing the point some 2500 years ago.

A. ANAXIMANDER OF MILETUS

Anaximander (circa 610-546 BCE) was a Greek Presocratic philosopher.[22] We know little about his life, with the few details being found in other philosophers' doxographies, including those of Aristotle and Theophrastus.[23] According to such works, he was a follower (or perhaps a pupil) of Thales, a fellow citizen-philosopher.[24] Anaximander seems to have been involved in politics, and he may have helped establish the Miletus colony of Apollonia, located along the Black Sea.[25]

Perhaps unsurprisingly given his early day, nearly all of Anaximander's philosophy comes to us secondhand. But one gem—one verbatim

---

[20] Notice that such justice seems to permit substitution: the excessive heat of today, say, is not concerned with which breeze or snowflake made it too cold the previous winter; what is important is the rebalancing through counter-penetration. One might make historic parallels, as when a killing or other wrong was considered otherwise repaid when the killer were dead or for some other reason untouchable. (E.g., the concept of a "whipping boy." *See, e.g.*, *Whipping Boy*, WIKIPEDIA, https://perma.cc/DU2V-VDC3 (visited Sept. 22, 2023); SID FLEISCHMAN, THE WHIPPING BOY (1986).) Similarly, a recalibrative framework might explain why at times in medieval England pretrial release was triggered by a guarantor's willingness to pay the penalty upon conviction—what was important was that it be paid, not that the particular wrongdoer pay. *See* TIMOTHY R. SCHNACKE, MICHAEL R. JONES & CLAIRE M. B. BROOKER, THE HISTORY OF BAIL AND PRETRIAL RELEASE 1–2 (Pretrial Justice Institute 2010). Such substitution is also found in other philosophies, such as the Christian theory of atonement. *See, e.g.*, *Substitutionary Atonement*, https://perma.cc/358T-GWCK.

[21] *See generally* WILLIAM FAULKNER, THE SOUND AND THE FURY (1929) (character Quentin is driven to suicide over the concern that wrongs are not righted nor even sufficiently agonized over).

[22] Dirk L. Couprie, *Anaximander*, THE INTERNET ENCYCLOPEDIA OF PHILOSOPHY, https://perma.cc/U36E-CR9J (archived on Sept. 7, 2023).

[23] *Id.*

[24] James Evans, *Anaximander*, ENCYCLOPAEDIA BRITANNICA, https://perma.cc/JE9Z-N7EL (archived on Sept. 7, 2023).

[25] Couprie, *supra* note 22; 2 EARLY GREEK PHILOSOPHY: BEGINNINGS AND EARLY IONIAN THINKERS, PART 1, at 271 (André Laks & Glenn W. Most, eds. & trans., 2016), available online at https://perma.cc/EH5K-2ATJ.

fragment—survives.[26] And that fragment, the content of which we will turn to in a moment and which is found in the writings of Simplicius, is celebrated as the oldest recorded lines of Western philosophy.[27] But while we have terribly few of his words, we enjoy more of Anaximander's thought. Other philosophers relate his theories on geography, nature, cosmology, and cosmogony (concerning the origin of the universe). He posited a shape for the earth (a free-floating cylinder) and for the sun and moon (fire-filled rings).[28] He is credited with the first map of the world.[29] And, intriguingly, Anaximander believed that humans originated from fish, perhaps showing a very early glimmer of evolutionary thought.[30]

B. THE FRAGMENT

Central to Anaximander's cosmogony was what he termed the "Boundless," which he conceived as the infinite source of all existence.[31] Under his theory, the constant motion of the Boundless created the opposites (for example, hot and cold), thereby providing the basis for all substance.[32] Eventually, all of existence must be reabsorbed into the infinite, bringing about the end of the known universe.[33] And it is in the context of this broad theory that we have the famous fragment as related by Simplicius. One translation provides as follows, where the bolded words are believed to be Anaximander's own and the parentheticals are those of the translators:

> Among those who say that it (i.e. the principle) is one, in movement, and unlimited, Anaximander . . . said that the principle (*arkhê*) and element of beings is the **unlimited** (*to apeiron*); he was the first to call the principle by this term. He says that it is neither water nor any other of what are called elements, but a certain other unlimited nature from which come about all the heavens and the worlds in them. And the things out of which birth comes about for beings, into these too their destruction happens, **according to obligation: for they pay the penalty** (*dikê*) **and retribution** (*tisis*) **to each other for their injustice** (*adikia*) **according to the order of time**—this is how he says these things, with rather poetic words.[34]

Both the fragment's translation and interpretation have naturally been the topic of vigorous debate. For instance, Charles Kahn offers a rather

---

[26] Evans, *supra* note 24; Laks & Most, *supra* note 25.
[27] Couprie, *supra* note 22.
[28] *Id.*; Evans, *supra* note 24.
[29] Couprie, *supra* note 22.
[30] *Id.*; Evans, *supra* note 24.
[31] Evans, *supra* note 24.
[32] *Id.*
[33] *Id.*; Couprie, *supra* note 22.
[34] Laks & Most, *supra* note 25 (footnote omitted).

different version in which far more words are directly attributed to Anaximander:[35]

> Anaximander declared the Boundless to be principle and element of existing things, having been the first to introduce this very term of "principle"; he says that "it is neither water nor any other of the so-called elements, but some different, boundless nature, from which all the heavens arise and the κόσμοι within them; out of those things whence is the generation for existing things, into these again does their destruction take place, according to what must needs be; for they make amends and give reparation to one another for their offense, according to the ordinance of time," speaking of them thus in rather poetical terms.[36]

Yet another source—attempting to reconstruct some of the poetic elements that the author believes would have been in the original—provides the following aurally pleasing account:

> Whence things have their origin,
> Thence also their destruction happens,
> As is the order of things;
> For they execute the sentence upon one another
> —The condemnation for the crime—
> In conformity with the ordinance of time.[37]

While it seems generally agreed that the natural elements—the wet or the dry, say—are what ultimately provide the described 'reparation,' serious debate exists as to the 'receiving' entity that was previously wronged. Under one interpretation that perhaps has echoes in the modern science of matter and antimatter,[38] the elements 'wrong' the Boundless by their mere existence, and, thus, when the universe is reabsorbed, the elements are 'repaying' the Boundless, righting the 'injustice' caused at their generation.[39] Some prominent proponents of this reading include Friedrich Nietzsche and Hermann Diels.[40]

More recently, however, another interpretation has found favor,[41] and it is one of particular interest to us. Under this reading, the fragment describes

---

[35] While debate thus exists as to where the direct quote begins, it seems generally agreed that at least the final words are indeed a verbatim preservation. For one in-depth examination see CHARLES H. KAHN, ANAXIMANDER AND THE ORIGINS OF GREEK COSMOLOGY 166–78 (1960).

[36] *Id.* at 166.

[37] Couprie, *supra* note 22.

[38] *See, e.g.*, University Of California, Santa Cruz, *How Did Matter Come To Dominate The Universe?*, SCIENCEDAILY (Feb. 19, 2001), https://perma.cc/BG5R-QDQJ.

[39] *See* Couprie, *supra* note 22; KAHN, *supra* note 35, at 167–68; Gregory Vlastos, *Equality and Justice in Early Greek Cosmologies*, 42 CLASSICAL PHILOLOGY 156, 169–72 (1947).

[40] *See* KAHN, *supra* note 35, at 168; Vlastos, *supra* note 39, at 170 n. 134.

[41] This new reading stems largely from translating a previously-singular pronoun as plural. *See* KAHN, *supra* note 35, at 167–68; Vlastos, *supra* note 39, at 169–72.

the elements paying reparation *to each other*, with Gregory Vlastos translating Anaximander's final lines like this: "according to just necessity (*cheron*); for they render justice and reparation to one another for their injustices according to the ordering of time."[42] For Vlastos, Anaximander described a balanced universe in which the opposing elements (think hot and cold) hold equal power or influence.[43] Yet what requires Anaximander's universe, born from the Boundless and eventually returning thus, to be 'just' during its existence? To Vlastos, the requirement is consistent with a broader Presocratic tradition considering justice to be innate.[44] Under such conception, justice was considered the balance of opposing forces, and thus the universe—'just' by necessity—requires opposites that remain in balance.[45] And what does such a universe demand when one element becomes momentarily dominant, encroaching upon the domain of another? To restore the "injustice" done to its opposite, the element must pay "reparation" for its offense (quoting Vlastos's translation of the Anaximander fragment). An example of this is found in the seasonal cycles: the heat or cold may become dominant for a time, but what follows is a corresponding season of the opposite.[46]

Vlastos thus interprets Anaximander's universe as recalibratively just: each invasion is directly and proportionately counter-invaded, restoring the natural balance. This reading of the fragment has been echoed by subsequent scholars, including Charles Kahn[47] and—as was quoted above—Martha Nussbaum.[48]

C. RECALIBRATIVE CRIMINAL JUSTICE

Vlastos's interpretation is hardly the only possible translation and reading.[49] But whether true to Anaximander the historical person or not, it provides us a conception of justice, one which looks at the 'intrusion' done and restores it in kind. In a system of criminal justice, then, the just penalty would be that which is equal and opposite to the wrong. The victim, say,

---

[42] Vlastos, *supra* note 39, at 168.

[43] *See id.* at 168–73.

[44] *See id.* at 177.

[45] *See id.* at 172–78.

[46] Vlastos bolsters his interpretation with Presocratic Greek medical theory, which describes the cyclical supremacy of opposite powers in the human body. *See generally* Vlastos, *supra* note 39.

[47] *See* KAHN, *supra* note 35, at 166–96.

[48] *See supra* note 19 and accompanying text.

[49] For yet another example, *see* Joyce Engmann, *Cosmic Justice in Anaximander*, 36 PHRONESIS 1 (1991) (describing an alternate reading under which the elements are permitted unequal powers, resulting in an imbalanced universe more akin to aristocracy).

would have lived a long and full life, but instead was killed and so never made it beyond a young age. All that lost life, justice demands.

Yet recalibrative justice offers only a starting point in the human discussion. From this beginning, philosophies shift toward a more particularized, personalized conception: instead of focusing merely on *harm* (as the recalibrative model demands), justice becomes concerned with the *who* of the offender.

III. ARISTOTLE

Thinkers later than Anaximander proffered what has been termed an equitable theory of justice, where "equity" is meant to require looking to the particulars of both the wrong*doing* and wrong*doer*. Thus, another term for this conception would be *particularized* justice. A human wrongdoer—a killer, say—is not an entity or a binary conception like Anaximander's 'hot' or 'cold' that has preconceived amounts of any single attribute; instead, every human life is a nuanced, unique experience. Particularized justice therefore attempts to account for those nuances in determining any punishment. For example—to state a straightforward case—what was a killer's intention, her hope in acting? Some killers wish to kill; others negligently kill. Despite the same harm, most of us innately find the desert to be very different. So, if a criminal system wishes to punish according to desert, it must deviate from the demands of recalibrative justice.[50] A well-known proponent of such a theory lived and taught some 300 years after Anaximander.

A. ARISTOTLE OF MACEDONIA

While Aristotle is perhaps (at least in name) one of the best-known humans of all time, a brief introduction will be useful to highlight both why that is the case and how his thinking has helped develop our collective contemporary sense of criminal punishment. Aristotle was born in Macedonia in 384 BCE to Nicomachus, the physician to King Amyntas III.[51] At approximately age seventeen, he moved to Athens and joined Plato's

---

[50] Prominent philosophies have sometimes theorized a mechanism to achieve both recalibrative and particularized justice. The Christian atonement, for example, might be argued to have both a particularized component requiring individual repentance *and* a recalibrative component that operates both beyond and entirely independent of such particular desert.

[51] Christopher Shields, *Aristotle*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Aug. 25, 2020), https://perma.cc/Q5JA-HJ7R; *see also* Anthony J. P. Kenny, *Aristotle*, ENCYCLOPAEDIA BRITANNICA, https://perma.cc/7SR5-TN6M (archived Sept. 7, 2023); Justin Humphreys, *Aristotle*, INTERNET ENCYCLOPEDIA OF PHILOSOPHY, https://perma.cc/UA34-TKKU (archived Sept. 7, 2023); ARTISTOTLE, THE NICOMACHEAN ETHICS, ix–xxiv (H. Rackham, trans., 1926) [hereinafter, NICOMACHEAN ETHICS].

Academy, where he remained for the next twenty years and likely produced his earliest works—unfortunately, they exist today only as fragments.[52] Following Plato's death, Aristotle left Athens for what is now Turkey, where he continued his work and befriended, among others, his fellow doxographer of Anaximander, Theophrastus.[53]

Around 343 BCE, King Philip of Macedon requested that Aristotle tutor his thirteen-year-old son Alexander, who would become Alexander III of Macedon, or "Alexander the Great."[54] While we have little record of the extent, content, or influence of Aristotle's instruction—even the length of their relationship is disputed[55]—by one account Alexander held Aristotle in high esteem, admiring him as a second father.[56]

When Aristotle returned to Athens in 335, he established the Lyceum, a school dedicated to researching everything from logic to music to politics to astronomy, arguably setting the example for modern observational science.[57] It was naturally also a center of philosophy, and the school's collection of manuscripts has been considered perhaps the first great library of antiquity.[58] In a democratizing move, many of its lectures—unlike those at Plato's earlier Academy—were free and open to the public.[59] It is during this period that Aristotle is believed to have produced most of his surviving works,[60] and to have had a son Nicomachus for whom his *Nicomachean Ethics* might be named.[61] And it is within those pages that we find an articulation of "equitable" or "particularized" justice.[62]

B. ON EQUITY

In Book V of the work, Aristotle considers equity as it relates to justice, explaining that—while equity is not justice per se—the two are not distinct,

---

[52] *See* Kenny, *supra* note 51.

[53] *See id.*; Shields, *supra* note 51. Theophrastus would become leader of Aristotle's Lyceum following the latter's death. Kenny, *supra* note 51.

[54] Kenny, *supra* note 51.

[55] *See id.*

[56] *See* NICOMACHEAN ETHICS at x.

[57] *See* Shields, *supra* note 51; Kenny, *supra* note 51. One source reports Aristotle's return to Athens being one year earlier (336 BCE). *See* NICOMACHEAN ETHICS at xi. That account also describes him first spending a brief period in the city of his birth, which—although destroyed during war—had been recently restored. *Id.* at x–xi.

[58] *See* Shields, *supra* note 51.

[59] *See* Kenny, *supra* note 51.

[60] *See id.*

[61] *See* Shields, *supra* note 51. Indeed, some speculate that Nicomachus edited the treatise. *See* NICOMACHEAN ETHICS at xiii.

[62] *See* NICOMACHEAN ETHICS at 313–317.

either.[63] Rather, equity is a *type* of justice, and it is one that sometimes proves superior to what we might call *strict justice*, meaning that proscribed by any black-letter law.[64] Equity, according to Aristotle, "while superior to one sort of justice, is itself just."[65]

But what is this equity? According to Aristotle's definition, equitable judgments are those that account for the circumstances at hand.[66] In other words, they look to the *particulars* of the situation. This is a familiar enough concept; the classic child's defense that "I didn't know that" or "I didn't mean to" is evidence that we feel—intuitively and early in life—that particular circumstances, and not merely objective harm, ought to inform any judgment. This alone separates equitable justice from a purely 'Anaximanderian,' recalibrative conception. And, importantly, it does so in at least two steps. First, particularity considers not merely the harm done (recalibrative justice), but the requirements of a previously-promulgated law ("I didn't know that"). That law should embody the ideals and norms of the legislators (and thus, in a democracy, of the populace), itself possibly enacting variable conceptions of justice. For example, perhaps a society favors recalibrative justice for some crimes (stealing, say), but considers that system too harsh for others (such as murder). This is a somewhat-particularized system, as it differentiates just outcomes based on types of crime. It is a system with, to use more modern terms, a principle of legality.[67]

But such consideration of promulgated law is not alone sufficient to achieve Aristotelian particularity. Aristotle noted that *all* written law is necessarily general, able to consider only "the majority of cases."[68] Aristotle recognized this is no fault of the legislator, who cannot hope to account for all circumstances in such a variable world.[69] Hence, the promulgated rule is not unjust or any less legitimate; it is merely "defective" on particular facts.[70] (To return to our child hypothetical, perhaps that she stole only under the duress of a physical bully.) To ensure that we achieve a just result, then, we must apply an equitable solution, or one which adjusts to the circumstances at hand. Specifically, we look to the intent of the legislator, deciding the case

---

[63] *See id.* at 313; *see also* Nussbaum, *supra* note 18, at 92; Anton-Hermann Chroust, *Aristotle's Conception of Equity (Epieikeia)*, 18 NOTRE DAME L. REV. 119, 121–22 (1942).

[64] *See* Nussbaum, *supra* note 18 at 93; Chroust, *supra* note 63, at 121.

[65] NICOMACHEAN ETHICS at 315.

[66] *See id.* at 315–17; Chroust, *supra* note 63, at 122.

[67] *See Principle of Legality in Criminal Law*, WIKIPEDIA, https://perma.cc/CTZ4-73L3 (archived Sept. 22, 2023).

[68] NICOMACHEAN ETHICS at 315.

[69] *See id.*

[70] *Id.*

"as the lawgiver would himself decide if he were present on the occasion, and would have enacted if he had been cognizant of the case in question."[71]

So, in this 'second step,' equity is superior to the strict, black-letter law—at least in cases where the latter is, "because of its absoluteness . . . defective and erroneous."[72] But it cannot be superior to the *ideal* of justice itself.[73] As Aristotle put it: "[W]hile the equitable is just, and is superior to one sort of justice, it is not superior to absolute justice, but only to the error due to its absolute statement."[74] In other words, equity is error correction, or, in the words of Martha Nussbaum, it is "putting law into the condition to which it aspires in the first place."[75] It is by *deviating* from the strict law's definition of justice that we *achieve* the aims of justice, which here would be thwarted by slavish application of the general rule. In the words of Anton-Hermann Chroust, these Aristotelian deviations are thus made "for general Justice's sake."[76]

Still, while such "rectification[s]" can thus be necessary,[77] they ought to be made sparingly.[78] As Chroust describes, if we too often—or at least too cavalierly—stray from the strict law's dictates, we threaten its legitimacy, and—by extension—the state's stability.[79] After all, fairness-as-consistency is also a principle of justice. Thus, for Chroust, we ought to adjust only when we see that the written law, "due to circumstances inherent to the particular nature of the case, seems to be insufficient to achieve its aims by its own means."[80]

C. PARTICULARIZED CRIMINAL JUSTICE

Aristotle may not have been the first Western philosopher proposing equity—as Nussbaum has described, there is evidence of a larger Athenian movement toward equitable judgment.[81] But Aristotle's account in the *Nicomachean Ethics* offers a clear, operational definition, allowing us to

---

[71] *Id.* at 317.

[72] *Id.* at 315.

[73] *Id.* at 317.

[74] *Id.*

[75] Nussbaum, *supra* note 18, at 93 n. 19.

[76] Chroust, *supra* note 63, at 123.

[77] NICOMACHEAN ETHICS at 317.

[78] *See* Chroust, *supra* note 63, at 125 (arguing them a "last resort").

[79] *See id.*

[80] *Id.* at 126.

[81] *See* Nussbaum, *supra* note 18, at 95–96. Nussbaum points to speeches offered before Athenian trial juries: "These speeches show the orators relying on a concept of law and even of justice that is very much like the one that Aristotle renders explicit and systematic." *Id.* at 96.

contrast equitable—or, as we often term it, particularized—justice with other formulations. Particularization first requires we consider not merely harm done (recalibrative justice), but the requirements of previously-promulgated law. And, second, it further requires considering *every situation* as unique. We not only look to the particular law, but—in cases where the *ideals* underpinning that law so demand—we deviate therefrom.[82]

But are there times in which even that particular deviation is insufficient? Can such a particularized punishment still be unduly harsh?

IV. SENECA

One way to approach these questions is to consider not only the law—now particularized to the defendant and the surrounding circumstances—but also the party or parties sitting in judgment. Are these persons not also flawed humans, tempted by their particular worldly circumstances to do wrong, both in their lives generally and indeed in this very judgment? What effect will imposing even a particularized judgment have on *these* persons? The Stoic Seneca considered these questions some 300 years after Aristotle helped explain equity.

A. SENECA OF ROME

Lucius Annaeus Seneca was born in Spain to a noble Roman family.[83] While (like many details of his life) his birthyear is uncertain, it is commonly placed between 4 and 1 BCE.[84] His father spent much of his time in Rome, and Seneca accompanied him there at an early age.[85] After extensive

---

[82] Andrew Brien explains it like this:

> Equitable action involves fitting, or individuating, the response of the legal system to a person's individual circumstances through a consideration of the features of the case, when the law as it stands is unable to take them into account. Equity allows exceptions to the law to be admitted within the institution of the law itself and ensures that . . . justice is done, injustice is avoided, and the rule of law maintained. Equity rejects the blind application of rules in favor of the use of discretion, perception, and a consideration of the individuating features of a case.

Andrew Brien, *Mercy Within Legal Justice*, 24 SOCIAL THEORY & PRACTICE 83, 90 (1998).

[83] SENECA: ANGER, MERCY, REVENGE, vii (E. Asmis, S. Bartsch, & M. C. Nussbaum eds., R. A. Kaster & M. C. Nussbaum trans. 2010); *see also* Katja Vogt, *Seneca*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Jan. 15, 2020), https://perma.cc/5YDZ-NJ87; Robert Wagoner, *Seneca, Lucius Annaeus*, INTERNET ENCYCLOPEDIA OF PHILOSOPHY, https://perma.cc/YJ7W-679W (archived Sept. 11, 2023); 1 SENECA: MORAL ESSAYS, DE PROVIDENTIA. DE CONSTANTIA. DE IRA. DE CLEMENTIA., vii–viii (J. Henderson ed., J. W. Basore trans., 1928), available at https://perma.cc/X2AA-57ZT.

[84] *See* Wagoner, *supra* note 83.

[85] *Id.*

education in rhetoric and philosophy, Seneca embarked on a career in Roman politics.[86] That career got off to a slow start, perhaps owing to bouts of ill health that would plague his entire life,[87] but by year 41 Seneca reached the Roman Senate, at which point he encountered a more particular setback: accusations of committing adultery with former emperor Caligula's sister.[88] While the charges may have been politically motivated, Seneca was convicted by the Senate and exiled by Emperor Claudius to the island of Corsica, where he would study and write for several years.[89]

Seneca's fortunes improved in 49 CE, when Emperor Claudius' new wife, Agrippina, recalled Seneca to Rome to tutor her son Nero, whom Claudius had adopted.[90] Claudius also had a son born to a former wife, but Agrippina was eager to see her own child on the throne, and, by 54 CE, she succeeded.[91] With Nero as emperor, Seneca amassed significant wealth and power, serving as one of Nero's primary advisors and speechwriters (though most transcripts thereof have been lost).[92]

But that administrative good fortune would not last. Whatever influence Seneca may have once held over Nero, it waned, and their relationship deteriorated until issues came to a head in 65 CE, when Seneca was implicated in the "Pisonian conspiracy" to murder Nero.[93] Nero thus ordered his former tutor to kill himself, a mandate that Seneca allegedly confronted with grace.[94] According to one account, Seneca's suicide required three attempts: The first two methods—slitting his veins and ingesting hemlock—failed due to his ill health and poor circulation.[95] Only once he was placed in a hot bath to improve that blood flow did he finally suffocate on the steam.[96]

---

[86] *Id.* It seems that Seneca's father was eager for his eldest two sons (Seneca and his older brother Annaeus Novatus) to follow in his political footsteps. *See id.*

[87] *See* Wagoner, *supra* note 83; SENECA: ANGER, MERCY, REVENGE, *supra* note 83, at vii.

[88] *See* LUCIUS A. SENECA, HOW TO KEEP YOUR COOL: AN ANCIENT GUIDE TO ANGER MANAGEMENT, xiii (J. Romm, ed. & trans., 2019); Wagoner, *supra* note 83.

[89] SENECA: ANGER, MERCY, REVENGE, *supra* note 83, at viii.

[90] *Id.*

[91] *Id.*

[92] *Id*; Wagoner, *supra* note 83. It is during this period that Seneca is believed to have authored *De ira* and *De clementia*, the ethical treatises we will shortly consider. *See* SENECA: MORAL ESSAYS. DE PROVIDENTIA. DE CONSTANTIA. DE IRA. DE CELEMNTIA., *supra* note 83, at xi; SENECA, HOW TO KEEP YOUR COOL: AN ANCIENT GUIDE TO ANGER MANAGEMENT, *supra* note 88, at xiv.

[93] *See* SENECA: ANGER, MERCY, REVENGE, *supra* note 83, at viii–ix. It remains unclear whether Seneca was actually involved in the plot. *Compare id. with* Wagoner, *supra* note 83.

[94] Wagoner, *supra* note 83.

[95] *Id.*

[96] *Id.*

His wife attempted to follow him in suicide, but, according to that same telling, she was saved on Nero's orders.[97]

In the course of his tumultuous life, Seneca authored a large collection of works, including philosophical letters and treatises, a host of tragedies (including the famous *Medea* and *Agamemnon*), a quasi-scientific treatise on the natural world, and a work of political satire.[98] While he considered himself a Stoic, Seneca was not unwilling to explore matters from other perspectives, granting each its merits.[99] He has been recognized both for his rhetorical prowess and for writing in a conversational style that, though unusual at the time, was popular with his contemporary audience.[100] That style is evident in both *De ira* (*On Anger*) and *De clementia* (*On Mercy*), which—taken together—articulate a theory of *merciful* justice.

B. ON ANGER AND MERCY

Consistent with Stoic teaching, Seneca considered emotion a vice.[101] Reason alone ought to guide our acts, and in succumbing to passions, we stray from the dictates of rationality, and thus of nature.[102] Seneca found no emotion as harmful as anger,[103] which he defined as "the desire to take vengeance for a wrong."[104] Thus, in *De ira*, Seneca argues against its evils, including in the context of delivering criminal punishment. And in *De clementia*, addressed to the young Nero and possibly unfinished,[105] Seneca urges the importance of clemency in wielding the power of judgment.[106] Between these two texts, we get a comprehensive view of Seneca's justice.

Seneca's theory certainly requires that we examine a case's particulars. In *De ira*, for example, he argues that a wise person, when confronted with a wrongdoer, must consider not only the actor's intent,[107] but also her situation and age, thereby "put[ting] ourselves in the place of the person we're angry with."[108] For Seneca, this renders penalties most likely to further the threefold

---

[97] SENECA: ANGER, MERCY, REVENGE, *supra* note 83, at ix.

[98] Wagoner, *supra* note 83.

[99] *Id.*

[100] *Id.*

[101] SENECA: ANGER, MERCY, REVENGE, *supra* note 83, at 3.

[102] Wagoner, *supra* note 83.

[103] SENECA: ANGER, MERCY, REVENGE, *supra* note 83, at 14. In Seneca's (translated) words, "[N]o pestilence has been more costly to the human race." *Id.* at 16.

[104] *Id.* at 16.

[105] *See id.* at 133 (a large portion of the work is missing; it is unclear whether it was lost or the treatise was never completed).

[106] *Id.* at 133–34.

[107] *Id.* at 31–32.

[108] *Id.* at 72–73.

goals of punishment: "either to correct the person punished, or to improve everyone else by punishing him, or to allow everyone else to live more securely once the malefactors have been removed from their midst."[109] A good judge, then, has utilitarian aims, "keep[ing] his eye on the future, not the past."[110] By looking to the particulars, society metes out punishments best suited to each individual, and thus most likely to achieve a better future.

Critically for Seneca, these punishments must be delivered without anger, that nefarious passion "that finds pleasure in payback."[111] In his (translated) choice words, "And so the wrongdoer should be corrected both by admonition and by force, softly and roughly, and he must be made better for his own sake as much as for that of others, not without scolding, but without anger."[112] Because anger is itself evil, even in punishment we ought not become angry with wrongdoers.[113] Indeed, because *all* humans are flawed,[114] "the person who reproaches individuals for a vice we all share is unjust."[115] The world is a cruel place, leaving us all sinners, and "[e]ven if you've done not a jot of evil, you're capable of it."[116] Just as we would not become angry with an infant, we ought not become so with each other, as "[b]eing human is a greater excuse, and more just, than being a child."[117]

Thus, punishments given in anger harm the judge—in a world so rife with evil, one who becomes angry at all wrongdoers condemns herself to a life of misery.[118] Indeed, such a person would "not become angry but insane."[119] When, by contrast, punishment is not out of anger, and not intended to right the wrong, but instead merely to do the most good, we spare both the punished and the punisher from unnecessary harm. This, for Seneca, is *clemency*, which he defines as "the mind's inclination toward mildness in exacting punishment."[120] "[C]lemency is that which reins itself short of what

---

[109] *Id.* at 166–67.

[110] *Id.* at 32.

[111] *Id.* at 20.

[112] *Id.* at 26.

[113] Seneca stresses this point throughout both texts. *See, e.g.*, *id.* at 25–26.

[114] *Id.* at 53. "We're all inconsiderate and careless; we're all unreliable, complaining, grasping; we're all . . . wicked." *Id.* at 85.

[115] *Id.* at 84.

[116] *Id.* at 85.

[117] *Id.* at 40.

[118] *Id.* at 39, 85–86.

[119] *Id.* at 40.

[120] *Id.* at 172 (internal quotation marks omitted).

could deservedly be ordained,"[121] "mak[ing] determinations according to what is fair and good."[122]

C. MERCIFUL CRIMINAL JUSTICE

Clemency—or, alternatively termed, *mercy*—is thus a critical step beyond, or at least different from, particularity. As Andrew Brien has explained, while both particularity and mercy step us into the wrongdoer's shoes, there is a key difference in goal: particularity is employed to seek the goal of just punishment, whereas mercy's concern is the human that will receive it.[123] Indeed, Senecan mercy goes one step further still, as it also considers the human that will *deliver* it. For Seneca, such mercy as "an inclination of the mind" ought not only guide judgment in criminal cases, but ought to define our approach to the world far more generally.[124]

V. ROLE-REVERSIBILITY

Purely abstract justice, along the lines of Anaximander's "recalibration," is missing (at least) two things when used as a theory of criminal justice. One is the way particularity unavoidably informs the application of general rules; the other is the importance of attitudinal mercy in the implementation of even particular rules. In other words, sometimes general rules are too rigid, and sometimes they are too harsh—and they could, at least in principle, suffer from both defects at once.

Given all of this, how should the enterprise of judgment be structured? Is there a manner in which a system of justice, and specifically a system of criminal justice, ought to be configured in order to maximize the chances of achieving outcomes that are, at once, (1) sufficiently categorized to satisfy legality conditions, (2) sufficiently particularized to vindicate equity principles, and (3) sufficiently lenient in the face of human frailty? We believe so. Our answer comes back to role-reversibility.

Role-reversible justice means as few contingent differences as possible between the judging party and the judged party. This inclines the judging party to adopt the merciful disposition: 'Had things been only slightly different, the positions may well have been inverted, with me bearing the consequences of judgment rather than meting them out.' Since as humans we

---

[121] *Id.* at 172.

[122] *Id.* at 175.

[123] Brien, *supra* note 82 at 91–92. In Brien's words, "the object of the actions performed by an equitable actor is not the other person, but a principle [i.e., justice], while the response of the mercy-giver is to the beneficiary and her [as a human necessarily] hard lot." *Id.* at 92.

[124] *Id.*

are all deeply flawed—in principle, and often in practice—role-reversibility inclines us towards Senecan mercy, making us slow(er) to punish and quick(er) to forgive, often stopping short of the particularly-deserved (meaning the equitably-deserved) penalty.

When the judging party realizes that she could be the judged, she is required, in essence, to judge herself. And *that*, we believe, is the core of merciful justice. It makes for judgments fraught with uncertainty, and it makes the enterprise of judgment painful, pressing towards abdication and away from retribution. Indeed, it would be impossible for anyone to regularly exercise such judgment—the human brain inevitably suppresses the routinely painful—which is why laypersons, rather than professionals, tend to be the proper actors. They are more prone to feel the anguish of judgment—and to temper their decisions in response. Moreover, by collectively committing to a role-reversible framework, we may increase the likelihood that the defendant will accept a judgment—or at the least its process. And, even when a defendant does not, it prevents the judging party or parties from harming themselves by vengeance and anger.

Finally, and for the very same reasons, role-reversibility informs what we ought to do if and when truly intelligent machines arrive in our human drama. Whether or not they ought to be role-reversibly punished—a fascinating question we put aside for another day—*we* cannot be mercifully judged by them until they become role-reversible.[125] In other words, it is only when there are very few situational hops between human and machine—when, as with any other (human) member of the polity, it becomes possible to say, of a robot, 'but for the grace of fortune, there go I'—that machines might, in principle, be charged with determining human guilt.

[125] The issue of robot rights more generally is a fascinating one. *See, e.g.,* Daniel Akst, *Should Robots With Artificial Intelligence Have Moral or Legal Rights?*, WALL ST. J. (Apr. 10, 2023), https://perma.cc/E8F6-9E8S.