

Empathy, Democracy, and the Rule of Law

Author : Frank Pasquale

Date : May 8, 2019

Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, __ **J. of Crim. L. and Criminology** __ (forthcoming), available at [SSRN](#).

Are some types of robotic judging so troubling that they simply should not occur? In *Artificial Intelligence and Role-Reversible Judgment*, [Kiel Brennan-Marquez](#) and [Stephen E. Henderson](#) say yes, confronting an increasingly urgent question. They illuminate dangers inherent in the automation of judgment, rooting their analysis in a deep understanding of classic jurisprudence on the rule of law.

Automation and standardization via software and data have become a regulative ideal for many legal scholars. The more bias and arbitrariness emerge in legal systems, the more their would-be perfecters seek the pristine clarity of rules so clear and detailed that they can specify the circumstances of their own application. The end-point here would be a robotic judge, pre-programmed (and updated via machine learning) to apply the law to any situation that may emerge, calculating optimal penalties and awards via some all-commensurating logic of maximized social welfare.

Too many “algorithmic accountability” reformers, meanwhile, are in general either unaware of this grand vision of a legal singularity, or [acquiescent](#) in it. They want to use better data to inform legal automation, and to audit it for bias. The more foundational question is less often asked: Does the robo-judge not simply present problems of faulty algorithms and biased or inaccurate data, but something more fundamental—a challenge to human dignity?

Brennan-Marquez and Henderson argue that “in a liberal democracy, there must be an aspect of ‘role-reversibility’ to judgment. Those who exercise judgment should be vulnerable, reciprocally, to its processes and effects.” The problem with an avatar judge, or even some super-sophisticated robot, is that it cannot experience punishment the way that a human being would. Role-reversibility is necessary for “decision-makers to take the process seriously, respecting the gravity of decision-making from the perspective of affected parties.”

Brennan-Marquez and Henderson derive this principle from basic principles of self-governance:

In a democracy, citizens do not stand outside the process of judgment, as if responding, in awe or trepidation, to the proclamations of an oracle. Rather, we are collectively responsible for judgment. Thus, the party charged with exercising judgment—who could, after all, have been any of us—ought to be able to say:

This decision reflects constraints that we have decided to impose on ourselves, and in this case, it just so happens that another person, rather than I, must answer to them. And the judged party—who could likewise have been any of us—ought to be able to say: *This decision-making process is one that we exercise ourselves, and in this case, it just so happens that another person, rather than I, is executing it.*

Thus, for Brennan-Marquez and Henderson, “even assuming role-reversibility will not improve the accuracy of decision-making, it still has intrinsic value.”

Brennan-Marquez and Henderson are building on a long tradition of scholarship which focuses on the intrinsic value of legal and deliberative processes, rather than their instrumental value. For example, the U.S. Supreme Court’s famous [Mathews v. Eldridge](#) calculus has frequently failed to take into account the effects of abbreviated procedures on claimants’ dignity. Bureaucracies, including the judiciary, have enormous power. They owe litigants a chance to

plead their case to someone who can understand and experience, on a visceral level, the boredom and violence portended by a prison stay, the [brutal need](#) resulting from the loss of benefits, the sense of shame that liability for drunk driving or pollution can give rise to. And as the classic *Morgan v. United States* held, even in complex administrative processes, the one who hears must be the one who decides. It is not adequate for persons to play mere functionary roles in an automated judiciary, gathering data for more authoritative machines. Rather, humans must take responsibility for critical decisions made by the legal system.

This argument is consistent with other important research on the dangers of giving robots legal powers and responsibilities. For example, Joanna Bryson, Mihailis Diamantis, and Thomas D. Grant have [warned](#) that granting robots legal personality raises the disturbing possibility of corporations deploying “robots as liability shields.” A “responsible robot” may deflect blame or liability from the business that set it into the world. It cannot truly be punished, because it lacks human sensations of regret or dismay at loss of liberty or assets. It may be programmed to look as if it is remorseful upon being hauled into jail, or to frown when any [assets under its control](#) are seized. But these are simulations of human emotion, not the thing itself. Emotional response is one of many fundamental aspects of human experience that is [embodied](#).

Brennan-Marquez and Henderson are particularly insightful on how the application of law needs to be pervasively democratic in order to be legitimate. That is, of course, most obvious in the concept of the jury, but in a way that refines our common understanding of the practice. To understand “why the jury has long been celebrated as an organ of ‘folk wisdom,’” Brennan-Marquez and Henderson argue:

The idea is not that jurors have a better sense of right and wrong than institutional actors do. (Though that may also be true.) It is, more fundamentally, that jurors respond to the act of judgment as humans, not as officials, and in this respect, jury trials are a model of what role-reversibility makes possible: even when a jury trial does not lead to a different outcome than a trial before an institutional judge (or other fact-finding process), it facilitates the systemic recognition of judgment’s human toll. And even more fundamentally, it transforms the trial into a democratic act.

The common humanity of the judge (or agency director, or commissioner) and litigants is another reflection of the democratic nature of the polity that gives rise to a legal system.

It should come as little surprise that authoritarian legal systems are among the [most enthusiastic](#) for automatic, computational judgments of guilt or “trustworthiness.” Their concepts of “rule by law” place authorities above the citizenry they judge. By contrast, rule of law values, rooted in a democratic polity, require that any person dispensing justice is also eligible to be subject to the laws he or she applies.

Artificial Intelligence and Role-Reversible Judgment is a far-seeing project—one that aims to change the agenda of AI research in law, rather than merely improving its applications. Brennan-Marquez and Henderson carefully review the many objections scholars have raised to the data gathered for legal AI, and the theoretical objections to the vision of “artificial general intelligence” that seems necessary for computational legal systems to emerge. “We do not minimize any of these instrumental arguments in favor of human judgment,” they argue. “They are certainly valid today, and they may survive the next generation of AI. [But this article explores] what should happen if arguments like these do not survive.” The requirement for a human to evaluate arguments and dispense judgments in a legitimate legal system should give pause to those who are now trying to develop artificially intelligent judges. Why pursue the research program if it violates the role reversibility principle, which Brennan-Marquez and Henderson rightly characterize as a basic principle of democratic accountability?

Brennan-Marquez and Henderson’s work is a great example of how a keen phenomenology of the uncanniness and discomfort accompanying a vertiginously technified environment can deepen and extend our understanding of key normative principles. Judged by an avatar, one might wonder: “Who programmed it? What were they paid? Did they

understand the laws they were coding? What could I have done differently?" The emerging European right to an explanation is meant to give persons some answers to such queries. But Brennan-Marquez and Henderson suggest that mere propositional knowledge is not enough. The "[right to a human in the loop](#)" in legal proceedings gains new moral weight in light of their work. It should be consulted by anyone trying to advance legal automation, and those affected by it.

Cite as: Frank Pasquale, *Empathy, Democracy, and the Rule of Law*, JOTWELL (May 8, 2019) (reviewing Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, __ **J. of Crim. L. and Criminology** __ (forthcoming), available at SSRN), <https://cyber.jotwell.com/empathy-democracy-and-the-rule-of-law/>.