Asimov’s Laws of Robotics and their Significance for the Prison

Curtis Blakely¹
Truman State University, United States of America

Michelle Blakely²
University of Wyoming, United States of America

Abstract
Using Asimov’s laws of robotics, the similarities between mechanical and social systems are considered. While Asimov’s laws were created to guide the actions of man-like machines and safeguard society from errant robots, a similar set of directives could be used to ensure that prisons, and other institutions, also do not err in their operations. After creating and hypothetically applying a slightly modified version of Asimov’s laws, it is determined that the prison is violating its prime directive.

Keywords: Asimov, Robot, System, Prison, Self-Interest, Blakely’s Operational Directives.

Introduction
Having an insatiable interest in humankind’s ability to benefit from science and technology, we have long been attracted to science fiction. In this unique form of literature, readers are introduced to inventions that have yet to be developed. Science fiction authors routinely predict these developments, assess their probable impact on society, and create laws to govern their operations. Few science fiction writers have achieved more success at these tasks than Isaac Asimov (1919–1992). Asimov was one of the first science fiction authors to predict the rise of “robotics” – creating this term in 1941. His predictions are now becoming reality. For example, within the next 5 years, 70% of Japan’s hotel jobs will likely be staffed by robots that resemble human beings (Semuels, 2018). As a professor of biochemistry at Boston University, Asimov lived during the “golden age” of science. Best known for his insights into how mechanical humanoids or “man like machines” might serve the needs of future generations, Asimov created the “laws of robotics”. Asimov introduced these laws in his 1942 short story

¹Associate Professor, Justice Systems Department, Truman State University, 100 E Normal Ave, Kirksville, MO 63501, United States of America. Email: cblakely@truman.edu
²Asst. Professor, School of Pharmacy, University of Wyoming, Dept. 3375, 1000 E. University Ave., Laramie, WY 82071, United States of America. Email: mblakel4@uwyo.edu
“Runaround”. These laws serve as the conceptual foundation for the remainder of this paper with implications being applied to both the prison and all publicly funded entities. Our objective is to approach our subject matter in an innovative, creative and progressive fashion that facilitates additional thought and discussion.

Admittedly, it is a bit unorthodox to consider the prison’s operations from the perspective of science fiction. However, the “laws of robotics” can and do provide insight into the creation and operation of social systems. For example, the robots portrayed by Asimov were a conglomeration of mechanical and electronic components. As such, each robot was a complex system that sought to achieve various tasks. In fact, for the purpose of the current undertaking, we will define a “system” as any collection of interdependent parts that, when functioning properly, works toward the accomplishment of a particular goal – in other words, all systems must have a clearly defined purpose! When dealing with a social system, this purpose is ideally defined by those who fund and are intended to benefit from its use (i.e. the taxpaying citizen). Furthermore, when we use the term “prison”, we are referring to the institution, its departments, its administrative personnel and its employees that collectively form a system of interdependent parts reminiscent of Asimov’s robots. It is because of the similarities that exist between mechanical and social systems that the application of Asimov’s laws can be applied in a broad and generalized fashion.

Asimov’s Laws of Robotics

To fully appreciate Asimov’s contributions to the fields of robotics and artificial intelligence, we must acknowledge his vision of a time when mankind would create imitations of itself through the use of mechanical and computer technologies. He foresaw the development of technologies that could supplement and/or replace the need for human labor, especially when it involved repetitive or specialized actions. For example, most of us have, at one time or another, seen video footage of robots at work within the automotive industry. These robots are used to weld body parts and to position axles and chassis during the assembly process. Others may have viewed footage of robotic apparatus assisting physicians during surgeries requiring absolute precision. As is evident from these two examples, robots are designed to specialize in the work they undertake, thereby increasing their ability to achieve a particular objective. Regardless of the task being performed, all robots are governed by a set of operational laws. The purpose of these laws, according to Asimov, is to facilitate task-completion and to safeguard society from deviant or injurious behaviors committed by a robot. Asimov’s laws are as follows:

1). A robot may not injure a human (or humanity) or, through inaction, allow a human (or humanity) to be harmed;
2). A robot must obey the orders it receives except when they conflict with the first law; and
3). A robot must protect its own existence except when doing so conflicts with the first or second law.

While these rules establish a set of operational parameters, they also create a series of functional boundaries that can be applied to all systems whether mechanical or not. For example, if we modify these laws slightly, we create a set of rules that can be applied to
the prison, and by default, to all social institutions (we refer to these as Blakely’s Operational Directives). These laws similarly serve to safeguard society from injurious behaviors committed by the prison. For example:

1. A prison may not injure a human (or humanity) or, through inaction, allow a human (or humanity) to be harmed;  
2. A prison must obey the orders it receives except when they conflict with the first law; and  
3. A prison must protect its own existence except when doing so conflicts with the first or second law.

Once Asimov’s laws have been modified, it then becomes necessary to apply them to the prison. We strongly assert that without a set of laws similar to those now proposed, the prison will, like the early robots depicted by Asimov, become increasingly confused, ineffective, and self-serving.

Applying and Interpreting Our Laws

It is necessary at this point to apply our modified laws to the prison. This will permit us to assess current operations and determine each law’s potential benefit. We will begin with our version of Asimov’s first law:

\[ A \text{ prison may not injure a human (or humanity) or, through inaction, allow a human (or humanity) to be harmed}. \]

This law requires the prison to promote the well-being of both the offender and society. Historically, penologists recognized that society’s future was intimately connected to the prison’s ability to facilitate offender reform. The rehabilitation of the offender was viewed as an investment that would produce long-term social dividends. To determine whether the prison is currently accomplishing this objective, one need only consider recidivism rates. Recidivism rates are generally considered the most comprehensive measure of the prison’s ability to break the crime cycle. These rates are now at their highest recorded levels as measured by subsequent arrests and incarcerations. For example, most sources agree that this rate is currently around seventy percent (Grattet, Petersilia & Lin, 2008; Langan & Levin, 2002). In essence, for every one-hundred offenders released from prison, seventy will be arrested for a felony or serious misdemeanor and/or return to prison straightaway. And when we compare current recidivism rates to past rates, we see a striking difference. For example, recidivism rates during the 1970s appear to have been approximately half of what they are today (Cahalan, 1986). Because of this increase, one must conclude that the modern prison is having a detrimental effect on the stability of society. Stated a bit differently, there is a dynamic at work in the contemporary prison that serves to perpetuate criminality. This “effect” stands in stark contrast to our first law prohibiting the prison from causing injury or allowing injury to occur.

To determine the source of this “effect” one must review the literature. A review of the existing literature reveals the prison’s contemporary abandonment of treatment initiatives. This abandonment was hastened when Ronald Reagan assumed the presidency. Advocates of this movement asserted that all attempts at rehabilitation should cease since
they considered them to be ineffective, unnecessary and costly. According to their rationale, treatment rewards offenders for their errant behaviors by incentivizing criminality. Not only did advocates of this movement seek to eliminate all treatment services, but they also sought to expand the use of incarceration as a criminal sanction. As this movement gained support, prison construction skyrocketed. For example, from 1984 to 2005 a new prison was being built every nine days in the United States (Herring, 2015). There are now more than 1800 federal and state prisons in operation. These facilities hold in excess of 1.5 million inmates. That number jumps to about 2.5 million when jail inmates are added to the equation (Wagner & Rabuy, 2016). Few who study the prison would deny that it is operating in direct violation of its original mandate. And while external political pressures have certainly affected its operations, the prison has nonetheless permitted itself (without any noticeable resistance) to be placed into a position whereby it operates without a concern for offender reform. By withholding treatment, the prison has failed to promote the well-being of either the offender or society.

Having briefly considered our version of Asimov’s first law, let us now turn our attention to a modified version of his second law:

A prison must obey the orders it receives except when they conflict with the first law.

This law pertains directly to the prison’s mission. In essence, we find Asimov’s use of robotic “orders” and the prison’s “mission or prime directive” to be comparable. A prime directive is an operational law(s) that is so fundamental that it must never be violated. Orders and prime directives can be visualized as internal codes that require, or prohibit, specific actions. Stated a bit differently, an institution’s prime directive establishes its overall purpose and its operational parameters. But what does history reveal about the prison’s prime directive? Has this directive changed over time, and if so, what are the consequences?

The most appropriate way to address these questions is to consider the prison’s purpose before and after nineteen-eighty (the pre and post-Reagan years). When searching for answers to these questions, it becomes evident that rehabilitation has long been considered an effective way to break the crime cycle and to promote public health. William Paley (1743 - 1805), an English philosopher and religious leader, wrote extensively on these subjects. His efforts, and those of many others, helped give rehabilitation a place of prominence within the early American prison. In fact, in 1787 a group of colonial leaders officially adopted rehabilitation as the primary mission of the prison. Then in 1870 (nearly a century later), at a meeting of the American Correctional Association, penologists again affirmed the importance of rehabilitation. Thus, for nearly 200 years, the prime directive of the prison was to facilitate inmate reform. However, as the end of the twentieth century approached, two exceptionally violent prison riots captured the nation’s attention. These riots were used by advocates of the get-tough movement to challenge the prison’s ability to facilitate rehabilitation. The first of these riots occurred in 1971 at the Attica Correctional Facility in New York State. Media accounts of this event portrayed inmates as brutal, inhumane, animalistic, and completely unworthy of educational, vocation and therapeutic provisions. The second riot occurred at the Penitentiary of New Mexico located in Santa Fe (1980). The timing of this riot helped solidify the anti-treatment movement that had gained significant momentum following Attica.
Collectively, these two riots were used by proponents of punishment to create the perception that America’s prisons were on the brink of anarchy and collapse. And, in fact, prisons were becoming increasingly violent. However, few considered the possibility that this increase was the result of changes in our political and penal ideologies. In essence, as the get-tough movement gained momentum, so too did inmate resistance. Increased violence prompted renowned criminologists Robert Martinson and James Q. Wilson to suggest that the prison should focus exclusively on its ability to incapacitate and control offenders. While their observations appeared to have prima facie merit, these riots involved only a small number of inmates and did not represent the entire population. Nonetheless, all inmates were portrayed as violent and predatory. These events were the impetus for the prison to rewrite its own prime directive, ignoring its unique history and traditional mission. This “rewrite” occurred without any public referendum whatsoever. Without public approval, it could be argued that this “rewrite” of the prison’s prime directive lacks legitimacy. To understand the significance of this action, imagine a robot that could simply rewrite its own orders or programs at will. This would create a situation in which it could violate its prime directive whenever it wanted to pursue other interests!

A similar dynamic is created if/when a public institution is permitted to do the same. It should be noted that the prison chose to disregard its original prime directive even when an overwhelming majority of citizens were supportive of rehabilitation. In fact, the findings of a Time magazine poll conducted fifty years ago showed that three-quarters of all citizens supported rehabilitation as the primary mission of the prison. A similar sentiment persists today (Alliance for Safety and Justice, 2016; Cullen, Fisher & Applegate, 2000). By rewriting its prime directive, the prison has placed itself into a situation whereby it justifies its own existence by downplaying the importance of rehabilitation and its role in promoting social health. The prison now seeks to protect its existence primarily through its growth and enormous size. Let’s now consider our version of Asimov’s third law:

*A prison must protect its own existence except when doing so conflicts with the first or second law.*

This law requires a determination to be made about the nature of the prison’s actions and whether they are threatening, injurious or are in violation of its original prime directive. Provided that the prison is functioning in accordance with the first two laws, it may continue operations. However, if it violates our third law, its existence must either cease or its actions must be brought into compliance. Asimov’s third law was originally intended to serve as a fail-safe mechanism to protect humanity from a robot that might seek to pursue its own interests at the expense of those it was intended to serve. This fail-safe mechanism, when applied to the prison, fulfills a similar function.

Numerous approaches can be employed to determine whether the prison is in compliance with this law. However, the most obvious is to determine whether self-interests are guiding the prison’s operations, and if so, to what end? If self-interests are at play, we would expect to see an attempt by the prison to produce an ever-increasing need for its services. As previously noted, rehabilitation is no longer an objective of the prison. Could this be an attempt by the prison to create a repeat clientele? Perhaps no better way exists for the prison to safeguard its own existence (when violating our first two laws) than
for it to create an increasing need for those services that it alone provides - and the best way to do this is to withhold treatment. This ensures that each offender released from custody has a greater likelihood of returning. And in fact, recidivism rates suggest that this is indeed the case. Thus, the contemporary prison’s abandonment of rehabilitation has allowed it to create, either intentionally or unintentionally, a group of citizens that are perpetually incarcerated. In doing so, the prison has grown significantly, creating an undeniable demand for its service. In essence, the prison has made itself indispensable not because of its ability to facilitate offender reform, but by the very nature of its size. If you doubt this observation, consider our previous statements concerning the degree to which the prison system has recently expanded and its insatiable appetite for state appropriations. Prisons are now given such a large portion of each state’s budget that the funds previously earmarked for a wide array of social services, including education, have steadily eroded (Clear, Reisig & Cole, 2016; Kelly, 2016). Arguably, there will always be a need for the prison; however, there is a difference between a need for a particular service and creating a need for that service through direct action or inaction. We have long contended that in an ideal world, the prison would seek to render itself obsolete - suggesting that if it were effective at breaking the crime-cycle, there would be a decreasing need for its services. It is through a decreasing need for the prison’s services that each of our laws is satisfied. Stated a bit differently, a decrease in the size of the prison system would indicate that it is effectively rehabilitating its inmates, and in doing so, fulfilling its social mandate. The prison does not need to protect its existence through expansion; rather it should seek its own security through effective offender reform. Any other action suggests that the prison is pursuing self-interests to the detriment of those it is intended to serve. Our third law, if applied, would ensure that the interests of the prison are forever subordinate to those of society.

Conclusion
The underlying premise of this paper is that similarities exist between mechanical and social systems. Asimov understood that robots must adhere to a set of laws to ensure that they remain harmless, loyal, and mission-driven. His laws, with slight modification, can be applied to all systems whether robotic or not. We strongly assert that without the creation and adoption of a set of laws similar to those we propose, social systems will, like the early robots appearing in Asimov’s stories, become increasingly confused, self-serving and will ultimately become ineffective.

Laws, like those we now propose, are necessary since recidivism rates are at their highest recorded levels. All indications suggest that the public overwhelmingly believes that the prison should pursue offender rehabilitation, and yet, those in a position to control the prison’s operations have ignored this sentiment. Treatment has now been replaced by a sterile approach that embraces incapacitation absent any concern for the post-release behaviors of inmates. This reveals a disjunction between the public’s expectation and the prison’s self-proscribed mission. In essence, there are competing interest at play - those of the citizen and those of the prison. The first desires rehabilitation, whereas the latter is driven by concerns about expansion, job security, and profitability.

Even though the prison was at the epicenter of political and economic changes occurring during the last two decades of the twentieth century, it should not have violated
Blakely & Blakely – Asimov’s Law of Robotics and their Significance for the Prison

(nor should it continue to violate) its original prime directive any more so than a robot facing environmental challenges. Until the prison is forced to table its own interests, the safety and well-being of both inmate and society remain at risk. If you doubt this observation, consider that there is no better way to ensure its own existence and growth than for the prison to withhold treatment. This approach has created an industry that is continually being fed an ever-increasing proportion of each state’s budget. Budgetary allocations for prison operations now exceed those that are provided to all other services including education. While we acknowledge a need for the prison, there is a difference between the need for a service and the creation of a need through action or inaction. This is especially true when a single institution creates and benefits from such a need.

One potential solution to address the many shortcomings of the contemporary prison is to have it imitate its robotic counterparts and “specialize” in a particular task. Under this approach, each individual prison would either incapacitate inmates in a safe and secure environment (free from treatment initiatives) or seek to rehabilitate them – but not both. The decision about the type of prison most appropriate for an offender would largely be based upon the inmate’s desire to either forego or undergo treatment. This would require each prison to adhere to a clearly defined and mission-driven purpose.

Since the prison is comparable to all other public institutions, we suggest that a set of prime directives, similar to those we have proposed, be created and adopted by all publicly funded entities. In doing so, the citizenry can be assured that their interests will be served. The similarities between robots and public institutions are obvious – each must be held to the highest standards possible to avoid ineffective and inefficient operations, mission confusion and constituency-harm.

References


