ROCZNIKI KULTUROZNAWCZE
Tom VII, numer 3 – 2016
DOI: http://dx.doi.org/10.18290/rkult.2016.7.3-1

PAUL RICHARD BLUM

ROBOTS, SLAVES, AND THE PARADOX OF THE HUMAN CONDITION IN ISAAC ASIMOV'S ROBOT STORIES

If it is true that slavery defies the notion of humanity, it is convenient to look at definitions of humanity not with the tools of essential or ontological definitions but from the margins of humanity. We can attempt to understand the meaning of being human from situations that endeavor to keep certain beings outside the accepted realm of humans. Since it has been said frequently enough that slavery denies humanity to slaves, it is crucial to find out what it is that is being denied, and how the conceptual construal of 'the slave' works. Eventually disassembling the structure of slavery opens insights into the human condition. One way of gaining such insights is by looking into parallel inventions of quasi human beings that are meant to be excluded from the concept and the company of humans.

Slaves and robots have in common that they are intended to obey orders. Therefore, I suggest taking a close look at some features of robots as thought out in Isaac Asimov's science fiction. In his story "Little Lost Robot"[1] the managers of the robots discover that one out of a group of robots lied. This appears to be impossible; as one manager, Peter Bogert, explains, each robot is made to obey commands and to "attempt to defend the carrying out of his orders" (440). To 'obey orders' is the essence of a machine, just as a computer

Dr. PAUL RICHARD BLUM — Department of Philosophy at Loyola University, Maryland,USA; address for correspondence — e-mail: prblum@loyola.edu

Thanks to Phillip Arvanitis for help with research and editing.

[1] Isaac ASIMOV, *The Complete Robot* (London: HarperCollins, 1995), 427–458. — This study is a result of research funded by the Czech Science Foundation as the project GA ČR 14-37038G "Between Renaissance and Baroque: Philosophy and Knowledge in the Czech Lands within the Wider European Context."

program follows the built-in algorithm. However, to defend the carrying out of orders presupposes that such a machine can be impeded in following instructions, and that the machine is programmed with a plan B in case the program does not run smoothly. The robot appears to be able to decide whether or not that is the case; and this latter feature makes an android appear to be thinking.

The whole complex: executing a program while detecting and overcoming problems and acting towards fulfillment of the instructions—all this makes a robot a perfect slave. This is not new, since already Karel Čapek, who coined the term 'robot' for machines that replace human workers,[2] associated robots with slaves. In his play *Rossum's Universal Robots (R.U.R.*; 1920*)* he suggests that robots could replace human slave labor, and in his science fiction piece *War with the Newts* (1936) he studies slavery in global economy. Slaves, by any standard understanding, share human features but lack free agency and therefore are supposed to execute any command automatically. In several of Asimov's robot stories we can find hints suggesting that the author himself thought of robotics as an allegory of slavery. In the story "Runaround" the engineers of early robots are said to have implanted "healthy slave complexes into the damned machines."[3] The reason was that robots were perceived by the general populace to be dangerous. The reader will not be surprised to read on the same page the robot answering questions by saying "Yes, Master." Another story with the appropriate title "Galley Slave" plays with the double meaning of galley proofs in publications and the ship that is propelled by slaves handling oars. A speaker states "a robot is far more reliable than a human being."[4] Such claim is true only, if the robot that does proof-reading as slave work is mindless, which is the topic of the story.

In "Little Lost Robot", the explanation for the robot to cover the truth is to be found in the paradoxical nature of a specific order he had been given, namely, "to lose himself." As Bogert states tersely: "How better can a robot lose himself than to hide himself among a group of similar robots?" (440). The situation is paradoxical not only because the command asks for a performative contradiction (not to take an impact from outside all too seriously, which may amount to dismissing an order) but also because the robot is not acting in a 1:1 relation to a commander alone; rather, he acts at the same

[2] Dominik ZUNT, "Who Did Actually Invent the Word 'Robot' and What Does It Mean?," July 27, 2013, http://web.archive.org/web/20130727132806/http://capek.misto.cz/english/robot.html (accessed October 4, 2015). Czech text at http://blog.abchistory.cz/cl97-karel-capek-o-slove-robot.htm (accessed October 4, 2015).

[3] ASIMOV, *The Complete Robot*, 261.

[4] Ibid., 385–426, 390.

time as a member of a group of "similar" beings. Robots, in this plot, are social beings and do have what we call behavior, when looking at beings in a social context. Behavior, as all components of morals that can be subject of an ethics, is always and by definition social. The elementary command structure of any machine is Do!/Done! There is no reasonable hesitation in the process of executing a command (unless, a delay is equally programmed). With this sort of computers the Do!/Done! appears to be embedded in the relation of 'Where are the others?' This is fundamentally possible because, in this plot, the robots are able to see themselves in context, and that implies to perceive themselves in the relationship between the master and the peers. As one early abolitionist document stated in Christian language: "We are taught by our blessed Redeemer to look upon all men, even our enemies, as neighbours and brethren, and to do unto them as we would they should to unto us."[5] The Golden Rule is a derivative of the ability to look upon 'all men' and one self. Consequently, the abolitionist viewed slavery as contradictory to the fellow-humanity of all humans, including enemies, slaves, and tyrants.

Bogert's interlocutor, Susan Calvin, lectures him on the nature of robots: "Those robots attach importance to what they consider superiority." (440) This statement is not only meant to confirm the social attitude of the robots, it additionally qualifies the social mind as a hierarchical mind. To see oneself amounts to watching out for the superiors, and, I suppose, inferiors. Indeed, those robots "feel humans to be inferior and the First Law which protects us form them is imperfect." (440) The First Law in Asimov's robot tales states: "No robot may harm a human being, or through inaction, allow a human being to come to harm."[6] The logic of these Three Laws is the constant theme and challenge within Asimov's stories, as will be clear when further discussing his take on slave-androids. However, for this episode it is essential that, in Calvin's reasoning, it is the social awareness of the robots that undercuts the effectiveness of the First Law. So she observes that it is increasingly important for a robot "to prove that it is superior despite the horrible names it was called." (440)

---

[5] John ADY et al., *The Case of Our Fellow-Creatures, the Oppressed Africans Respectfully Recommended to the Serious Consideration of the Legislature of Great-Britain* (London, Philadelphia: re-printed by Joseph Crukshank, 1784), 9.

[6] ASIMOV, *The Complete Robot*, 431, and throughout the Robot stories. In this story, part of the problem is that the second half of the rule has been omitted, in that sense the First Law is "imperfect" in this robot, but the theme of the plot is the inherent incompleteness of the rules.

Calvin refers to the robot as neuter, whereas Bogert used masculine gender. Now we may complain that Asimov's robots are already stealthily humans, at least the one that lied, but we may grant that the author of this fiction is teasing out the fringes between conscious humans and mechanical tools. He has Calvin contradict herself in treating the robots as things while still acknowledging that these things can have the drive to prove anything, even superiority over their masters. As soon as we translate this contradiction into the conundrum of androidism and slavery, it turns out to be the essence of that very relationship between the users of robots and the robots and, consequently, also that between masters and their slaves. Therefore we see that self-preservation and preservation of the master—which are the aim of the Three Laws—are not only conflicting but they are the master/slave conflict in a nutshell.

In Asimov's stories, the constitution of the robots consists of three laws. In the episode in which Calvin discovers "Robot Dreams"[7], these Three Laws are discussed to the effect that they eventually are reduced to one short imperative, namely that of self-preservation. The robot who has that dream (his name is LVX-1, expanded graciously by Calvin to Elvex) must die for revealing this dream.

The Three Laws are[8]:

1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Of the three imperatives the first is unmistakably a version of the Hippocratic Oath that rules all medical activity: "Do no harm!" And the set of three maxims seems to remind of Immanuel Kant's categorical imperative.[9] However, most importantly, these three imperatives are interlocked, as is expressly said and obvious in their language, so that there is a clear prefer-

[7] This the title story of another collection: Isaac ASIMOV, *Robot Dreams* (New York: Ace Books, 2004), 25–30.

[8] ASIMOV, *Robot Dreams*, 28; also ASIMOV, *The Complete Robot*, 605, 635 and more occurrences.

[9] Cf. Cedric M. SMITH, "Origin and Uses of Primum Non Nocere—Above All, Do No Harm!," *Journal of Clinical Pharmacology* 45, no. 4 (2005): 371–77. Marcello GISONDI, "Ma gli androidi leggono Kant? Le leggi della robotica: un possibile percorso epistemologico dalla letteratura al diritto," *ISLL Papers The Online Collection*, 209-216, 6 (2013), https://www.academia.edu/3122766.

ence for the humans and their service. At the same time, we as readers should be aware that these laws are Asimov's fiction. They are not meant to be proclaimed in any real world. The various stories are ways to find out what keeps a robotic being at bay in a human world that prides itself to dominate them. In the same way that slave laws were intended to keep slavery working by confining slaves in their place, so are the Three Laws—presented as rules given by the creator of robots—the formal condition for the workability of a robot holding society.[10] This is the precise theme of the dreams episode. The Third Law appears to protect the robots, but in reality it subjugates them to the humans even more than the antecedent commands, because it makes the preservation of the robot conditional upon the protection of humans and the indefeasibility of their power. In his dream, the robot upturns this structure. Asimov allows his reader to unthink the hierarchy by simply presenting us in "Robot Dreams" with the Three Laws in reverse order.[11] Thus the imbalance of the constitution becomes immediately evident: why should the existence of the robot depend on the interest and power of someone else? Elvex thinks something else: self-protection is the first and only maxim. He simply eclipsed the conditionals "as long as …"—at least in his dream. With this plot device it becomes evident that the hierarchy between humans and robots is not at all logical but utterly arbitrary.

But if this is the case, namely, that the constitution establishing the human-robot relationship is contingent, then even the distinction of humans over robots is unwarranted—at least in Elvex's dream. And hence Asimov has him dream yet another dream that is equally "not … under the control of the Three Laws". He has him see a man, a human being, and that human appearance says: "Let my people go!" That seals Elvex's fate. Calvin has to kill him—at least in the story.[12] She ventures that robots may have an "unconscious level" that eludes the control of the constitution. But, as Elvex claims, it was "the man" who called robots "my people". Well, in the Bible,

[10] Cf. Jana HORÁKOVÁ and Jozef KELEMEN, "Artificial Living Beings and Robots: One Root, Variety of Influences," *Artificial Life and Robotics* 13, no. 2 (March 8, 2009): 555–60, doi: 10.1007/s10015-008-0502-z. On the development of the Three Laws see Roger CLARKE, "Asimov's Laws of Robotics: Implications for Information Technology," in *Machine Ethics*, ed. Michael Anderson and Susan Leigh Anderson (New York: Cambridge University Press, 2011), 254–84. On standard interpretations of the Three Laws see Lee MCCAULEY, "AI Armageddon and the Three Laws of Robotics," *Ethics and Information Technology* 9, no. 2 (August 23, 2007): 153–64, doi:10.1007/s10676-007-9138-2.

[11] Compare with "Galley Slave" in ASIMOV, *The Complete Robot*, 385–426.

[12] ASIMOV, *Robot Dreams*, 29 f.

it is God who commanded the Pharaoh, through the voice of Moses, to let his people go (*Exodus* 7:16, 8:1, 9:1, etc.). This raises the question: what might Asimov/Calvin mean by an "unconscious level" that transcends the Three Laws? We can be sure that Asimov had the famous slave spiritual in mind. The least we can say is: humans as such are able to empathize with robots. How is it possible for a robot to dream of that? Asimov's story tells us that the self-preservation, once it is bestowed upon non-humans, makes them human. For upon Calvin's insistence Elvex, undoubtedly a robot by all intents and purposes, reveals that "the man" who commanded to set the robots free was he himself. Looking again at the Three Laws and their internal linkage through hierarchy, it appears that the Third Law, which is intended to secure predominance and interest of the humans, fuels robots with that very notion of self that keeps human egotism going by way of forbidding them to kill themselves, and is intended to shield masters from losing existence and power to the subjected robots, to the effect that—consciously or not—robots can become imaginary subjects of human agency and expand subjectivity towards solidarity with other robots. This is how dreams can kill. But as we will see soon, this might be a precious death: The Three Laws, in bestowing agency on the robots, are self-defeating. As we know from slavery, to demand from someone a service to humanity unfits the subject for slavery.

We are, obviously, translating Asimov's fiction into statements on human nature. That is to say, we assume that Asimov is confronting us with mental experiments about the question: what makes a human being human? As opposed to many philosophers, the question of the distinctiveness of humans is not discussed against the backdrop of animals. Many cognitive psychologists experiment with animals in a way to tease out to what extent an animal is able to think whereby thinking is delineated, for instance, by the ability to have 'a theory of mind of others', which can be expressed in behavior that indicates such an animal is able to anticipate how another animal will behave in a situation familiar to the one and surprising to the other.[13] It's like teaching philosophy to college students. Such experiments work with a deliberate reduction of the notion of thinking, or even consciousness, in an attempt to stretch the dividing line between the one and the other species. If that is the case, such experiments are fundamentally thought experiments insofar as

[13] For instance Elske VAART and Charlotte K. HEMELRIJK, "'Theory of Mind' in Animals: Ways to Make Progress," *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 191, no. 3 (February 1, 2014): 335–54.

they can well do without flapping wings, drooling tongues, or wagging tails. The whole organic complex of a live animal is not at all the focus of interest but simply some behavior that reveals basic processes of intelligence. And as befits scientific experiments, they address a precise question and eliminate disturbing and distracting circumstances, and thus a well-defined pattern of behavior is investigated. This is why Asimov opted for robots. Whatever the shape, size, sex, or beauty of a robot, eventually it has to function the way it has been designed. Therefore the question arises: Can robots do anything unpredictable? In the examples above: can it lie, or can it dream of not being a robot? And if so, what does it say about the dividing line between humans and androids?

Asimov had the dividing line between humans and robots being crossed in his story *The Bicentennial Man*.[14] In this story experiment, the android named Andrew refers to himself in the first person, desires to become free, and over the time of two hundred years acquires an organic body and eventually a human brain and dies in peace.[15]

First of all, Andrew speaks in first person terms.[16] Although some other robots in Asimov's stories also say "I", in this instance this is important because the entire story is told from the point of view of the protagonist. Although the narrator refers to him in third person, he never mentions any event or perspective that is not Andrew's. This is as close as it gets to a first-person-android-narrative. We may assume that Asimov chose to narrate in this style because he knew that—as a scientist—he is not entitled to pontificate over the robot's mind. The mind of a robot is factually designed by its engineer. But what it is like to be a robot, that is accessible only to him. As Little Miss put it when her father doubted that Andrew could have an idea of freedom: "I don't know what he feels inside but I don't know what *you* feel inside" (644). So, if we wish to understand the dividing line between an android and a man, we need to know the perspective and the life of that android, from inside, as far as we can. Of course, the constructor of the robot has a 'theory of mind' regarding his machine, because he has wired it the way it should perform its program. Inevitably, that is a reduced notion of

[14] ASIMOV, *The Complete Robot*, 635–682.

[15] To be precise, in this story 'android' refers to robots "that have the outward appearance of humans complete to the texture of the skin." Ibid., 662.

[16] Cf. Christopher GRAU, "There Is No 'I' in 'Robot': Robots and Utilitarianism," in *Machine Ethics*, ed. Michael Anderson and Susan Leigh Anderson (New York: Cambridge University Press, 2011), 451–63.

mind—a mind without surprises, except utter malfunctions. A properly functioning machine reveals no secrets. But as philosophers and as mental experimentalists, we wish to know: what are the conditions, circumstances, and limits of acting unpredictably? There is one surprise that should not be one: to say "I" is what makes a human being human.

Second, the robot is celebrated as the bicentennial man. This is a contradiction in terms, because no man lives for two hundred years. It has been reported that this story was written by invitation to celebrate the bicentennial of the United States of America. Therefore we are encouraged to assume that the robot represents African-American slaves.[17] However, such historical perspective is hard to interpret: does it, for instance, mean that after two centuries, and only after such a long time, Africans are free at last, but dead? If Asimov is speaking of slaves, then it is not in terms of historical development but in terms of the operational conditions that stretch from slavery to humanity. The first-person perspective tells the reader how to abolish the separating barriers. Therefore the long duration may well be a metaphor of timelessness. In order to fathom the 'life of the mind' of a robot, we have to represent him as brooding without time restraints. Human life is solitary, poor, nasty, brutish, and short (Hobbes). It appears as if Asimov had taken these five epithets as the touchstone for what distinguishes an android from a human. Andrew is longing to overcome his isolation as a machine, he is wealthy, clean (so clean that he does not need clothing), free of low instincts, and he is virtually immortal. Towards the end, Andrew explains:

> See here, if it is the brain that is at issue, isn't the greatest difference of all the matter of immortality? Who really cares what a brain looks like or is built of or how it was formed? What matters is that human brain cells die, *must* die. (680)

Paradoxically, it takes Andrew a long time to find out how to die, to die as a human being. After having replaced his mechanical organism with that of a human body, the last step will be to replace the computer brain with the organic mind of a human being.

---

[17] Susan Leigh ANDERSON, "Asimov's 'Three Laws of Robotics' and Machine Metaethics," *AI & Society* 22, no. 4 (2008): 477–93, doi: 10.1007/s00146-007-0094-5; 478; also in: Susan Leigh ANDERSON, "Asimov's 'Three Laws of Robotics' and Machine Metaethics," in *Science Fiction and Philosophy: From Time Travel to Superintelligence*, ed. Susan Schneider (Chichester: Wiley-Blackwell, 2009), 259–76; Sue SHORT, "The Measure of a Man? Asimov's Bicentennial Man, Star Trek's Data, and Being Human," *Extrapolation* 44, no. 2 (2003): 209–23, 219: "Andrew Martin , ... in effect born a slave ..." Cf. also Jane GOODALL, "Transferred Agencies: Performance and the Fear of Automatism," *Theatre Journal* 49, no. 4 (1997): 441–53.

Obviously, and this is the third necessary observation, Andrew's main impulse, motive, and desire is freedom. To make sure we get the point, Asimov inserts this exchange:

> 'Why do you want to be free, Andrew? In what way will this matter to you?'
> Andrew said, 'Would you wish to be a slave, your honor?' (646)

While it is hardly thinkable that a free person wishes to be a slave, that same person is able to ask, why freedom matters—an utterly stupid question to those who are unfree and aware of that. The same exchange continues by qualifying the notion of freedom. It is not, as one might expect, the mobility to do something or anything.

> 'What more can you do if you were free?'
> 'Perhaps no more than I do now, your honor, but with greater joy. It has been said in this courtroom that only a human being can be free. It seems to me that only someone who wishes for freedom can be free. I wish for freedom.' (646)

Freedom is self-reference in the desire to be free. That's all. It is so important that Andrew is able to wish to die. Freedom is therefore the enjoyment of being the agent of one's life. To speak to and of oneself in the first person, to say "I", that is the essence of freedom. In the film version with Robin Williams, Andrew repeatedly speaks of himself, saying: "One is pleased to be of service." That underscores the slave analogy, but is not present in Asimov's text.[18]

There appear to exist two ways to freedom: purchase and law. Andrew tries both, but eventually—after being declared free—it is the modification of his physical nature that makes the former slave a man. The question worth asking is: why did Asimov stage it that way? Since we are reading a mental experiment about the question of what constitutes a human being, we need to acknowledge that obtaining freedom is a fruit of self-awareness, whereas freedom as such may leave consciousness untouched.

In our story, Andrew develops artisan skills and accumulates savings. After consultation with a lawyer his owner allows him to save money earned

[18] Chris COLUMBUS, *Bicentennial Man*, Comedy, Drama, Fantasy, (1999). The text has once at the beginning, in first person: "It is my pleasure to please you, sir." ASIMOV, *The Complete Robot*, 636. On conceptual differences between the book and the film versions see Sara MARTÍN ALEGRE, "La Humanización del Robot en El Hombre del Bicentenario: Del relato de Isaac Asimov a la adaptación cinematográfica de Chris Columbus y Nicholas Kazan," *Seminario Tecnología y Posthumanidad*, 2002, 1–5 http://ddd.uab.cat/record/113501 (accessed September 28, 2015).

through his carpentry. It is the robot who insists on spending that money on repair and updating until he becomes "a paragon of metallic excellence" (642). However, Andrew's endurance is not that of a suffering person, it is that of a machine that knows no time pressure but reckons with time. It is after observing his owner grow old and a grandson being born that Andrew makes his request to be free. The robot's urge is that of principle—want of freedom—and not that of short-lived nastiness. Knowing the mechanism of commerce, Andrew offers his savings in exchange for freedom. It is important to notice that his owner is outraged over this proposal and that his daughter Little Miss translates the request into "a form of words. He wants to be called free" (644). Freedom is that of the other. The owner needs to understand that his own freedom is not endangered by the android's freedom. On the other hand, his daughter bridges the misunderstanding by explaining that on the side of the others, Andrew's freedom is just a verbal expression, whereas in the freed person's view it is the internal desire to be called externally for what he fundamentally is: free.

That amounts to acknowledging that freedom is not a merchandise: "'Freedom is without price, Sir,' said Andrew. 'Even the chance of freedom is worth the money'" (645). Keeping in mind that during the times of slavery a great number of slaves attempted and sometimes succeeded in buying their emancipation from their masters, it is worth noting that freedom is essentially priceless, whereas any such attempt is only the external expression of the personal strife. The android is available to pay for his freedom as a means to express his taking the risk by paying tribute to the monetary system.

It is typical that Asimov has the owner not haggle over the price. Rather, he invokes the law. We see that, at least at this point when Andrew offers money, among all possible frameworks neither the physical, nor the cognitive, nor any theological authority is referred to, but only legal reasoning. Yet the law fails. For when the attorney insists, "The word 'freedom' had no meaning when applied to a robot" (645), he declares the robot not to be subject to jurisprudence, which could intervene only if freedom and robotics were on an equal footing. Certainly, Asimov as a writer catered to the literary convention that treats societal problems in a court setting (think of Harper Lee's *To Kill a Mockingbird*). Here and elsewhere in the story, Asimov is framing the problem of the robot/slave in legal terms because these are supposed to be independent of prejudices and essentialisms that originate in history, religion, or worldview, let alone morals. Therefore, Little Miss intervenes again by underlining the subjectivism of freedom and its non-re-

lation to any capital value: "Making him free would be a trick of words only, but it would mean much to him. It would give him everything and cost us nothing" (646). This as the previously quoted plea might sound as though the woman wants to deceive her android. But each time she appends a statement form the perspective of Andrew. Freedom therefore appears to be purely nominal and yet essential to the person who wishes freedom for himself. Freedom is nominal to those who have it. "Everything," here, is beyond legality and marketability.

From this part of the story we gather that freedom is not something to be purchased, nor does it dwell within the area of competence of the law. This applies to robots and even more to slaves. The factual societal framework may need transactions within the boundaries of legality in order to establish that a person is free. But it is not money that sets Andrew free, and the verdict of the judge is simply admitting that freedom is outside the realm of the law: "There is no right to deny freedom to any object with a mind advanced enough to grasp the concept and desire the state" (646). One should relish the irony in the juxtaposition of 'object' and 'mind'. This verdict would certainly have overthrown the *Fugitive Slave Act* of 1850, which turned those who desired and achieved their freedom back into objects.

Therefore, we should pay attention to Andrew's work on his body and his brain, as well as to the application of the "Three Laws of Robotics," which—as we can confirm by now—are not part of the positive law in societies. Andrew undergoes a metamorphosis from wearing clothes to replacing his metal structure with an organic body and eventually with a human brain. His clothing provokes two bullies to test the Second and Third Law, namely that self-preservation succumbs to obedience. However, what appears as a moral dilemma (Andrew is aware of being about to be destroyed if he obeys) turns into a story of rescue by a family member. While the bullies claim that Andrew is a robot and at the same time is owned by no one, George, a family member of his previous owner, assumes the role of commander for a moment and threatens the bullies to order the android to go after them, so that they cut and run. The first two "Laws" are playing out, but only by convention, as they are not actually enforced. The relationship between robots and humans, and by analogy between masters and slaves, is that of fear, "a disease of mankind" (654). It is crucial in this scene of the novel that it is told from the perspective of the android, for an authorial, all-knowing narrator would have had to explain the motivations of the bullies and the feelings of George, and the very legal situation of the Three Laws would have to be

weighed. Then Andrew's appearance would have merited description. None of this occurs. What we read is how the android is cornered by the two young men so that he is motionless due to the conflict of the Laws of Robotics.[19] From that point of view the help he received from George is not motivated by any ruling but just a matter of friendship and the psychological, rather than technical, application of the Laws of Robotics.

It was the experience of humiliation that incited the robot to request and receive a human body. Andrew again has to go through legal arguments, but the moment he has this new body he decides to become a "robobiologist." Since Asimov has Andrew explain that this is not the same as a "robopsychologist" or a "roboticist" (665 f.), we may be confident that he picks his words. The robobiologist is precisely concerned with the unique case of a "positronic" robot brain, a computer that steers a human body. We need to resist the temptation to locate this idea in the more recent discussion on the interface of the mind and the organic or artificial body,[20] because our interest is that of Asimov's stories as allegories of slavery. Asimov points out that Andrew as a biologist of the robot would be dealing with himself, and not only because he is the only instantiation but also because that's what thinking humans do. Asimov explicitly refers to Andrew's first scholarly achievement, namely a history of robots: "A history of *robots*, by a robot. I want to explain how robots feel about what has happened since the first ones were allowed to work and live on Earth" (666; 654). Only a robot can tell what it is like to be a robot, and telling this realizes the history of robots from an inside perspective and thus again a narrative of the life of someone who is supposed not to be living. 'History of robots'—yet another contradiction in terms. Robots do not have a history (remember, Andrew lives a timeless life), but if robots can produce narratives of their lives, they prove to be human.

History tells stories about the past, stories that matter for the present time, especially the present of the historian. Therefore, reflecting upon history amounts to acknowledging that the future is unknown. The impossibility to know the future (in principle, not some petty predictable outcomes) is humanly expressed in the notion of immortality. As we see in Andrew's attitude

[19] "Runaround" is one of the stories that describe the immobilizing effect due to the conflict of two of the three rules. ASIMOV, *The Complete Robot*, 257–279.

[20] Out of countless publications see for instance Andy CLARK and David J. CHALMERS, "The Extended Mind," Journal (Paginated), Analysis, (1998), http://cogprints.org/320/. See also *Science Fiction and Philosophy: From Time Travel to Superintelligence*, ed. Susan Schneider, (Chichester: Wiley-Blackwell, 2009).

towards the passing of generations of his owner family, his being virtually immortal is the expression of the cognitive and emotional fact that to him as a robot the future is not unpredictable but unimportant. Andrew never feels time-constraints; again and again the story tells that he can wait. So, when Andrew pushes further to become a human being, immortality is at stake:

> Human beings can tolerate an immortal robot, for it doesn't matter how long a machine lasts. They cannot tolerate an immortal human being, since their own mortality is endurable only so long as it is universal. (680)

The challenge of immortality is not one for the immortals but one for the mortals. In the Christian philosophical debate over the question whether the human soul is mortal or immortal, which had its height in the 15th through 17th century,[21] the fundamental given was that—of course—all humans are mortal, so that in a sense all the immortals were actually dead. So Asimov is right in having Andrew state that—mortal or immortal—what counts is the universal rule. And as long as the supposedly immortal is sub-human, no human takes offence. From that point of view it was quite an achievement in Christian philosophy to grant immortality to the human soul while still acknowledging that human life is brutish and short. To claim that the immortal part is superior but inaccessible and the part that dies is the one at hand meant to have it both ways: human nature was elevated and degraded at the same time.

Again, we should not be distracted by either technicalities (like: is the mind a substance distinct from the body?) or by moral questions (suicide, for instance). Therefore Andrew dismisses the issue of "what a brain looks like or is built of or how it was formed" (680). The term suicide does not even appear; if anything there is a faint hint at Christ: the world was swayed, we read, by the final act in which the robot/man "had finally accepted even death to be human and the sacrifice was too great to be rejected" (681). What matters to this story is the intricacy of the Three Laws of Robotics. Hence, instead of the intimacy of a deathbed the world witnesses a media event. Andrew's death is broadcast even up to Mars. But what is this event about? The singularity of a robot turned android turned man? It is the universality of the Three Laws of Robotics. In his last conversation, Andrew points out:

[21] René Descartes' *Meditationes de prima philosophia* were the most famous contribution. Cf. Paul Richard BLUM, "The Immortality of the Soul," in *The Cambridge Companion to Renaissance Philosophy*, ed. James Hankins (Cambridge: Cambridge University Press, 2007), 211–33.

> I have chosen between the death of my body and the death of my aspirations and desires. To have let my body live at the cost of the greater death is what would have violated the Third Law. (680 f.).

And he adds: "If [my dying] brings me humanity, that will be worth it. If it doesn't, it will bring an end to striving and that will be worth it, too" (681). Humanity equals striving, so Andrew remained restless for two hundred years until he could rest in the thought of being human and to love "Little Miss." If the death of aspirations and desires is worse than corporeal death, then desires remain immortal beyond the passing away of the physical body. The Three Laws, as we saw at the beginning, culminate in the third, that of self-preservation. It is the same that can undo the whole set. Given a particular mind-set, the law of self-preservation annihilates obedience. However, for any slave-robot even the command to preserve oneself is given from outside. What if it comes from the subject? This is why Andrew's physical death is universal. It is within himself where the decision is made, where the drive comes from. As we see, since Andrew is able to refer to himself, he is also enabled to choose between himself and himself—his body and his aspirations.

Now, obviously there is nothing to choose for a robot, for there is no 'I' in robots, as we saw. Whenever someone makes a choice, an unprompted, non-necessitated decision, that person is unpredictable. Unpredictable defines any act that is free. Since freedom means the ability to act at all without being prompted or forced to do so, and since unpredictability is the cognitive state of not knowing the motives of an act, unpredictability is how freedom must appear in the eye of the beholder. To prevent unpredictability is the purpose of engineering.[22] From the engineering and slave management point of view, the robot Andrew was a failure, as we learn in the story: "an embarrassment to the company" (661). Therefore the company stopped producing robots like him. If the company had wanted a human-like robot then the weaknesses of Andrew would have been designed on purpose.[23] What had happened was that in his positronic brain there were "generalized pathways" (640) that allowed for surprises as they were not "precise and specialized" (667). In other words, Andrew's computer brain was not 'wired' to

[22] This is well expressed in the joke when in the impasse between the optimist and the pessimist over the question whether the glass is half full or half empty, the engineer solves the uncertainty with the statement: it's twice as large as needed.

[23] Jürgen KLÜVER and Christina KLÜVER, *Social Understanding: On Hermeneutics, Geometrical Models and Artificial Intelligence* (New York: Springer, 2011), 246.

execute every built-in program or to slavishly execute commands from the outside. That's what made him unpredictable. For the question of the cooperation of brain and body in him, his actions were not reducible to the make up of his brain. He had a mind of his own. In Asimov's narrative, this is expressed as a mishap in the construal of this sort of computer. If we dare translate that into the nature of the human mind, we may observe that to expect a human being to think and do as 'wired' in the brain equals reductionism, because it cannot explain spontaneous acts. Reductionism is a weakness of neuroscience if taken too seriously. On the other hand, Andrew's brain appears to be indeterminate. Hence he cannot be reduced to what his makers planned and programmed.

> 'It's amazing, Andrew, Paul went on, 'the influence you have had on the history of robots. It was your artistry that encouraged US Robots to make robots more precise and specialized; it was your freedom that resulted in the establishment of the principle of robotic rights; it was your insistence on an android body that made US Robots switch to brain-body separation.' (666 f.)

Andrew made history, which also resulted in him writing history. In the storyline he forced the factory to reduce robots to mere machines, prompted the robot owners to acknowledge his being human to some extent, and, again on the downside, his existence and agency showed the way to a clear separation of body and brain. It appears that a small margin of freedom can undo the distinction of mind and brain, so that Andrew can become a freely thinking and willing person. Asimov's story, translated into philosophy, says that it is never excluded that a properly functioning brain is actually more than its physical composition warrants. On the other hand, contemplating human nature, or at least the meaning of thought and freedom, invites investigation of the physical conditions of the functioning of the brain-body compound, hence the idea to separate bodies from their brains. What can we find when we do that?

In the story, "the corporation will produce one vast brain controlling several billion robotic bodies" (667). This sounds familiar. It is the Averroistic theory that there is but one intellect for all human beings.[24] Mortal humans are individual instantiations of one universal mind. It is needless to say that the Averroist mind is as immortal as Andrew's positronic brain. The greatest theoretical advantage of this theory is that it renders explainable how different individual human beings can have identical thoughts and communicate

[24] The easiest accessible classical text on this is Thomas Aquinas' *De unitate intellectus* (On the Unity of the Intellect), a refutation of Averroism.

about them. The greatest disadvantage that always has been pointed out is lack of—freedom! A person can only be free if granted freedom to think independently and unpredictably. That is the meaning of the company building one mind for all robots. Robots are supposed to be determined by their designers. The lack of individual brains secures predictable functioning through unfailing body-brain cooperation. The existence of an individual body attached to an individual brain is the wedge between both of them.

At this point one may venture a hypothesis as to why the protagonist of many of Asimov's robot stories is named Susan Calvin. It is her task to enforce determinism in her robots; as we saw in the stories "Robot Dreams" and "Little Lost Robot" sometimes she does not succeed in the challenge. This is the reason why Asimov reminds his readers of her role at the moment when Andrew negotiates to receive a new body: "Andrew found himself staring at the holograph on the wall. It was a death mask of Susan Calvin, patron saint of all roboticists" (661). Calvin a patron saint? Yet another ironic contradiction in terms: Calvinists refuse the veneration of saints and—for that reason—pictures of them. Andrew's will defies determinism and predetermination, which is one of the main tenets of John Calvin's theology.[25] Hence he is not a Calvinist.

What, then, is Calvin's tool to control and ensure the proper functioning of her creations? It is the Three Laws. To recapitulate, briefly, they command not to damage human beings, to obey, and to keep going. John Calvin's ethics was not meant for slaves or their owners, it was meant to represent the human condition. Through the way Asimov keeps challenging the workability of the laws of robotics, it transpires that he may well have had human ethics and behavior in mind. As we saw, in some stories the robots defy the three laws putting their masters in danger, at least the danger of losing control.

The Three Laws work only if commander and commanded, master and slave are not the same. The elementary mechanical structure of Do!/Done! requires the factual distinction between master and slave. As one lawyer pointed out:

> … any human being, *any* human being, has a fearsome power over any robot, *any* robot. In particular, since Second Law supersedes Third Law, *any* human being can use the law of obedience to overcome the law of self-protection (656).

[25] Donald M. HASSLER, "Some Asimov Resonances from the Enlightenment," *Science Fiction Studies* 15, no. 1 (March 1988): 36–47, specifically 37-38.

But what if the subject of the laws is also his master? As soon as the owner-subject relationship is being questioned the laws of robotics vanish. Or rather, they turn into universal laws of humanity. In Asimov's story, that happens early on when Andrew tries to purchase his freedom. Little Miss, who has a way with assuaging others and confirming the robot, assures her father: "The Three Laws still hold" (647), once Andrew is free. The father hears that he will remain the owner while voluntarily abstaining from using his power. Andrew, on the other side, understands that he will be freely obedient to the laws: "Are not human beings bound by their laws, Sir" (647)? Thus ends this short discussion. Since no specific law or code of conduct is mentioned, we may read that to mean: yes, the Three Laws apply to all human beings.

We noticed already that the First Law is that of the Hippocratic oath: not to injure any person. The Second requires obedience. Such an act is only possible in an established master-slave relationship. But we also saw at the beginning that self-preservation and obedience to another are fundamentally contradictory. Whoever is free is bound to obey himself, save the first law. What then is the purpose of the Third Law? For robots, if there are any in existence, it means to make sure to keep executing the commands of the master. If that master is the internal drive to exist, then the culmination of the three laws is to preserve humanity by obeying the Golden Rule, or the Kantian categorical imperative. Indeed, in one of the stories that feature Susan Calvin, one speaker explains:

> … the three Rules of Robotics are the essential guiding principles of a good many of the world's ethical systems. Of course, every human being is supposed to have the instinct of self-preservation. That's Rule Three to a robot. Also every "good" human being, with a social conscience and a sense of responsibility, is supposed to defer to proper authority… ; to obey laws, to follow rules, to conform to custom—even when they interfere with his comfort or his safety. That's Rule Two to a robot. Also, every "good" human being is supposed to love others as himself, protect his fellow man, risk his life to save another. That's Rule One to a robot.[26]

This is true, however, only if the "good human being" is truly autonomous. The scare quotes around the epithet 'good' indicate that this, to be good by following the rules of ethics, is the vexed question of humanity. Nevertheless, only a free subject can try to enact the three maxims.

[26] "Evidence" in ASIMOV, *The Complete Robot*, 518–545, 530.

Hence the conclusion is: Asimov's androids show the implicit impossibility of robots and of slaves by establishing the command structure that would be needed to keep the system working and then disassembling this structure. The Three Laws, as they are meant to guarantee protection, command, and operation, cannot possibly work with separate subjects. They are once again a paradoxical juxtaposition. And consequently, slavery is logically impossible.

BIBLIOGRAPHY

ADY, John, Anthony BENEZET, and John DRINKER. *The Case of Our Fellow-Creatures, the Oppressed Africans Respectfully Recommended to the Serious Consideration of the Legislature of Great-Britain*. London, Philadelphia: reprinted by Joseph Crukshank, 1784.

ANDERSON, Susan Leigh. "Asimov's 'Three Laws of Robotics' and Machine Metaethics." *AI & Society* 22, no. 4 (2008): 477–93, doi: 10.1007/s00146-007-0094-5.

ANDERSON, Susan Leigh. "Asimov's 'Three Laws of Robotics' and Machine Metaethics." In *Science Fiction and Philosophy: From Time Travel to Superintelligence*, edited by Susan Schneider, 259–76. Chichester: Wiley-Blackwell, 2009).

ASIMOV, Isaac. *Robot Dreams*. New York: Ace Books, 2004.

ASIMOV, Isaac. *The Complete Robot*. London: HarperCollins, 1995.

BLUM, Paul Richard. "The Immortality of the Soul," in *The Cambridge Companion to Renaissance Philosophy*, edited by James Hankins, 211–33. Cambridge: Cambridge University Press, 2007.

CLARK, Andy, and David J. CHALMERS. "The Extended Mind," Journal (Paginated), Analysis, (1998), http://cogprints.org/320/

CLARKE, Roger. "Asimov's Laws of Robotics: Implications for Information Technology." In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson, 254–84. New York: Cambridge University Press, 2011.

GISONDI, Marcello. "Ma gli androidi leggono Kant? Le leggi della robotica: un possibile percorso epistemologico dalla letteratura al diritto." *ISLL Papers The Online Collection*, 209-216, 6 (2013), https://www.academia.edu/3122766.

GOODALL, Jane. "Transferred Agencies: Performance and the Fear of Automatism." *Theatre Journal* 49, no. 4 (1997): 441–53.

GRAU, Christopher. "There Is No 'I' in 'Robot': Robots and Utilitarianism." In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson, 451–63. New York: Cambridge University Press, 2011.

HASSLER, Donald M. "Some Asimov Resonances from the Enlightenment." *Science Fiction Studies* 15, no. 1 (March 1988): 36–47.

HORÁKOVÁ, Jana, and Jozef KELEMEN. "Artificial Living Beings and Robots: One Root, Variety of Influences," *Artificial Life and Robotics* 13, no. 2 (March 8, 2009): 555–60, doi: 10.1007/s10015-008-0502-z.

KLÜVER, Jürgen, and Christina KLÜVER. *Social Understanding: On Hermeneutics, Geometrical Models and Artificial Intelligence*. New York: Springer, 2011.

MARTÍN ALEGRE, Sara. "La Humanización del Robot en El Hombre del Bicentenario: Del relato de Isaac Asimov a la adaptación cinematográfica de Chris Columbus y Nicholas Kazan,"

Seminario Tecnología y Posthumanidad, 2002, 1–5 http://ddd.uab.cat/record/113501 (accessed September 28, 2015).

MCCAULEY, Lee. "AI Armageddon and the Three Laws of Robotics." *Ethics and Information Technology* 9, no. 2 (August 23, 2007): 153–64, doi:10.1007/s10676-007-9138-2.

Science Fiction and Philosophy: From Time Travel to Superintelligence, edited by Susan Schneider. Chichester: Wiley-Blackwell, 2009.

SHORT, Sue. "The Measure of a Man? Asimov's Bicentennial Man, Star Trek's Data, and Being Human." *Extrapolation* 44, no. 2 (2003): 209–23.

SMITH, Cedric M. "Origin and Uses of Primum Non Nocere—Above All, Do No Harm!" *Journal of Clinical Pharmacology* 45, no. 4 (2005): 371–77.

VAART, Elske, and Charlotte K. HEMELRIJK. "'Theory of Mind' in Animals: Ways to Make Progress." *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* 191, no. 3 (February 1, 2014): 335–54.

ZUNT, Dominik. "Who Did Actually Invent the Word 'Robot' and What Does It Mean?" July 27, 2013, http://web.archive.org/web/20130727132806/http://capek.misto.cz/english/robot.html (accessed October 4, 2015). Czech text at http://blog.abchistory.cz/cl97-karel-capek-oslove-robot.htm (accessed October 4, 2015).

ROBOTY, NIEWOLNICY I PARADOKS LUDZKIEJ KONDYCJI W OPOWIEŚCIACH O ROBOTACH IZAAKA ASIMOVA

Streszczenie

Robotników i roboty łączy ze sobą intencja do posłuszeństwa rozkazom, która jest powodem mojej propozycji analizy kilku opowieści o robotach autorstwa Isaaca Asimova. Realizowanie programu w trakcie wykrywania i rozwiązywania problemów oraz wypełniania zadanych instrukcji – to wszystko czyni robota doskonałym niewolnikiem. Trzy Prawa Robotyki Asimova stanowią formalny warunek możliwości pracy w społeczeństwie utrzymującym roboty, tak samo jak niewolnicze prawa w koloniach brytyjskich Ameryki miały zapewnić utrzymanie efektywności niewolnictwa poprzez zamknięcie niewolników w miejscach ich pracy. Poprzez ustanowienie struktury rozkazu potrzebnej w utrzymywaniu i demontowaniu pracującego systemu androidy Asimova objawiają niemożliwość obydwu – robotów i niewolników. Trzy Prawa i ich konsekwencja, czyli gwarancja ochrony, rozkazu i operacji, prawdopodobnie nie mogą działać wobec podmiotów będących oddzielnie panami lub niewolnikami. Te Prawa są paradoksalnym zestawieniem w konsekwencji czego niewolnictwo staje się niemożliwe z punktu widzenia logiki.

Słowa kluczowe: Isaac Asimov; robotyka; samoświadomość; wolność.

ROBOTS, SLAVES, AND THE PARADOX OF THE HUMAN CONDITION IN ISAAC ASIMOV'S ROBOT STORIES

Summary

Slaves and robots have in common that they are intended to obey orders. Therefore I suggest taking a close look at some of Isaac Asimov's robot stories. Executing a program while detecting and overcoming problems and acting towards fulfillment of given instructions—all this makes a robot a perfect slave. In the same way as slave laws in the British Colonies in America were

intended to keep slavery effective by confining slaves in their place, so are Asimov's Three Laws of Robotics the formal condition for the workability of a robot holding society. Asimov's androids reveal the implicit impossibility of both robots and slaves by establishing the command structure that would be needed to keep the system working and then disassembling this structure. The Three Laws, as they are meant to guarantee protection, command, and operation, cannot possibly work with separate master/slave subjects. They are a paradoxical juxtaposition. And consequently, slavery is logically impossible.

Key words: Isaac Asimov; slavery; robotics; self-awareness; freedom.