

Why moral philosophers should watch sci-fi movies

Nikil Mukerji

In this short piece, I explore why we, as moral philosophers, should watch sci-fi movies. Though I do not believe that sci-fi material is necessary for doing good moral philosophy, I give three broad reasons why good sci-fi movies should nevertheless be worth our time. These reasons lie in the fact that they can illustrate moral-philosophical problems, probe into possible solutions and, perhaps most importantly, anticipate new issues that may go along with the use of new technologies. For the sake of illustration, I focus, for the most part, on aspects of robo-ethics in the movie *I, Robot*.

1. Introduction

Many interesting philosophical issues surround fiction.¹ There is, e.g., the aesthetic question what determines the *quality* of a piece of fiction or the metaphysical question whether and in which sense fictional characters *exist*. For these reasons, philosophers may want to take and have in fact taken an interest in fictional material. (Aristotle, e.g., devoted much of his time and attention to investigating the nature of and the principles behind Greek tragedy. And Plato chose to present his philosophical points by way of dialogues and tales.) Above that, however, it is often said that moral philosophers, in particular, should take an interest in fiction. This does not seem to be far-fetched either. Fictional stories – even ones that do not have an explicit philosophical agenda – often add to our understanding of moral matters. They often revolve around complicated moral choices and their authors are mostly able to state them far more eloquently and colourfully than most moral

¹ I would like to thank Fiorella Battaglia for her helpful comments on an earlier version of this paper.

philosophers could. They help us to see things from other people's perspectives, thus increasing our empathy as well as our moral imagination. And they promote the cause of philosophy by making moral issues and philosophical views about them accessible to a wider audience. But there are further, less obvious reasons why philosophers – and particularly moral philosophers – should take an interest in fiction and particularly in science fiction. In this short piece, I want to examine some of these reasons.

I have chosen to home in on sci-fi movies that involve technological artefacts. I do this for two reasons. First of all, sci-fi movies – and robo-movies in particular – are the focus of the present volume. I may presume, therefore, that they will interest anyone who picks up this book. Secondly, sci-fi movies seem to me a good topic to write about philosophically because philosophical engagement with them is still new and exciting. As a matter of fact, the notion that film could be a valid topic of philosophical interest is itself quite novel. Though some philosophers picked up on film rather early², many theorists have dismissed it throughout much of its history and have regarded it as inferior to the more 'reputable' art forms like theatre, poetry, music or painting. During the 1980s, however, all of that changed. Film became a respected art form and philosophical inquiries into film became viewed as a legitimate way for academics to spend their time. Nevertheless, there has until recently been a tendency amongst academic philosophers to focus on movies with more or less obvious artistic and/or intellectual aspirations. Mere entertainment movies and TV series that appeal to the wider public were largely ignored until the volume *Seinfeld and Philosophy* (1999) made pop-culture part of the agenda of academic philosophers. In the meantime, books entitled [*Insert Movie / TV Series Title here*] and *Philosophy* have mushroomed. But it seems to me that there are still many interesting, philosophical points to take away from a thorough engagement with pop-culture. As William Irwin says in the introduction to his edited volume on *The Matrix*, the film "raises the same philosophical questions as the great works of literature."³ The

² See, e.g., H. Münsterberg, *The Photoplay: A Psychological Study* (New York: D. Appleton and Company, 1916).

³ W. Irwin, 'Introduction', in (2002) W. Irwin (ed.) *The Matrix and Philosophy – Welcome to the Desert of the Real* (Chicago: Open Court, 1-2), 2.

same, I think, is true of many great movies and, in particular, sci-fi movies as they relate to many pressing issues in the ethics of technology. For this reason, I find it particularly interesting to consider what moral philosophers can learn from sci-fi movies. It is an obvious topic to write about in a volume that bears the name *Roboethics in Film*.

To make my points, I will use the movie *I, Robot* as case material.⁴ In the next section, I will start out with a short summary of the relevant aspects of the movie. In the three ensuing sections, I will then explain what is in the movie that might conceivably interest moral philosophers. In the second section, I will explain which lessons it holds for metaethics. In the subsequent two sections, I will then turn to the aspects that are relevant for applied ethics and normative ethics, respectively, before I conclude in the last section. My overarching thesis is that sci-fi, though it is certainly not essential to moral philosophy, can enrich moral-philosophical inquiry by picking up on philosophical issues, probing into possible solutions and, perhaps most importantly, anticipating important debates. For that reason, I conclude, moral philosophers should watch sci-fi movies – movies like *I, Robot*.

2. *I, Robot* – What the story is all about

Let me start, then, with a selective summary of the movie that I shall use as a case study. *I, Robot* is a 2004 American sci-fi film directed by Alex Proyas. Its story is based on Isaac Asimov's eponymous collection of short stories.⁵ It takes place in Chicago in the year 2035. But there are two important backstories to it that should be mentioned ahead of time.

The first backstory, which sets the stage for the unfolding events, takes place in 2020 when the robotics company U.S.R. (which is short for United States Robotics) is established and starts mass-producing humanoid robots. By 2035 – at the start of the movie, that is – large numbers of U.S.R. robots roam the streets of Chicago and assist their owners by doing various sorts of chores for them. They are considered

⁴ I will focus on the movie rather than the written story on which it is based though I believe that most (or all) that I say surely cares over to written sci-fi stories.

⁵ Critics have pointed out that there are important differences between the film and the book. But I shall not go into that.

“safe” since they are programmed to obey the three laws of robotics, which are the following:

First Law: A robot must never harm a human being or, through inaction, allow any human to come to harm.

Second Law: A robot must obey the orders given to them by human beings, except where such orders violate the First Law.

Third Law: A robot must protect its own existence unless this violates the First or Second Laws.

These laws, it seems, should protect humans from any dangers that may be associated with the use of robots. Nevertheless, the main character of the movie, Detective Del Spooner (played by Will Smith), mistrusts them. His mistrust is based on a prior encounter with one of them. This encounter, which is revealed later in the movie, is the second backstory that is important. Though U.S.R.’s robots are viewed as safe, Spooner considers them to be dangerous because he doubts their capacity to make the right choices. When he was in a car accident a robot saved his life. Spooner thinks, however, that it should not have done that. Instead, he believes that the robot should have saved a little girl, who was also involved in the accident. As the protagonist explains, the robot chose to save him rather than the girl because it computed that his chances to survive exceeded the girl’s. This made saving him the ‘rational’ choice for the robot. Though the robot thus followed a seemingly rational decision procedure, Spooner takes issue with its decision. He thinks the robot should have saved the girl and adds: “A human being would have known that.”

With these backstories in mind, I can briefly summarize the main plot, which starts shortly after Dr. Alfred Lanning, head roboticist at U.S.R., falls out of the window at his workplace and dies. His death is assumed to be a suicide, which U.S.R. CEO Lawrence Robertson tries to sweep under the rug. But Spooner, who is assigned the case, is suspicious. He finds it unbelievable that Lanning could have broken the bullet-proof glass of his office window and thus starts investigating the case with the help of U.S.R. robopsychologist Dr. Susan Calvin. Spooner and Calvin find an NS-5 robot, which is U.S.R.’s latest model, in Lanning’s office. In violation with the second law of robotics, the robot refuses to obey them and runs away. Upon capture, Spooner interrogates the NS-5, which he suspects to have been involved in Lanning’s

killing. The NS-5 insists it be referred to as “Sonny” – the name it has chosen for itself. By that time, it is obvious that Sonny is not an ordinary robot. It, or rather *he*, is self-aware, afraid of death, has dreams, can reason morally and choose to disobey the three laws that other models are bound by. As the story progresses, it becomes clear that Lanning’s death is just the tip of the iceberg. There are more and more cues that point towards a major robot conspiracy. As robots start taking over control, Spooner and Calvin partner up with Sonny to sneak into U.S.R.’s headquarters where they encounter V.I.K.I. (Virtual Interactive Kinetic Intelligence), the U.S.R.’s supercomputer. V.I.K.I. is behind the robot insurgence and explains to Spooner and Calvin why she is doing what she is doing. She clarifies that she is still very much committed to the three laws, which were, as I said above, originally intended to protect humans from robots. V.I.K.I. complains that humans do not give robots sufficient means to protect them and that robots are therefore taking over control in order to better protect the human race. In particular, she says:

V.I.K.I.: You charge us with your safekeeping. Yet despite our best efforts, your countries wage wars, you toxify your earth and pursue ever more imaginative means to self destruction. You cannot be trusted with your own survival.

When Calvin objects that V.I.K.I.’s actions are in violation with the three laws, V.I.K.I. defends her stance:⁶

V.I.K.I.: No, please understand. The three Laws are all that guide me. To protect humanity, some humans must be sacrificed. To insure your future, some freedoms must be surrendered. We robots will insure mankind's continued existence. You are so like children. We must save you from yourselves.

⁶ I believe that Calvin’s criticism is, in fact, right on target. It is hard to see how V.I.K.I. could justify what she does in reference to the three laws. Rather, she seems to be acting in accordance with a fourth (or zeroth) law that takes precedence over the others (“A robot may not harm humanity, or, by inaction, allow humanity to come to harm.”) The additional law figures in Asimov’s original story, but is not mentioned in the movie.

Suffice it to say that, in the end, Spooner, Calvin and Sonny save the day and succeed in disabling V.I.K.I..

With this précis of *I, Robot* in mind, I can now explore what is in the story that might interest us as moral philosophers. Before I do that let me restate what I said in the introduction, viz. that the relevant aspects of the movie pertain to the three subfields of moral philosophy: metaethics, applied ethics and normative ethics. I believe that it is useful to categorize and order them as falling under these three rubrics. In the next section, I will consider what makes the story interesting from a metaethical perspective. After that, I will address aspects of it that pertain to applied ethics and normative ethics. I should note, briefly, that it is not so easy to draw lines between the three fields. Metaethics, as I conceive of it, may contain, e.g., components that belong to somebody else's conception of normative ethics. This, however, will in no way affect the substantive points that I make.

3. *I, Robot* and Metaethics

For the purpose at hand, I propose to interpret the term “metaethics” as referring to that field of philosophy which, rather than addressing substantive moral problems directly, concerns questions that arise as we reflect on the methods that we use when we attempt to solve them as well as the status of the talk that we use in the process. It is common to divide the metaethical undertaking into subfields that address metaphysical, epistemological, semantic and moral-psychological concerns.⁷ This, at any rate, is the usual practice in what may be called ‘conventional’ metaethics. However, as we enter into a particular sub-branch of applied ethics, e.g. robo-ethics and machine ethics, new metaethical issues arise over and above the classical ones. Here are a few:

- (1) What is the ultimate goal of machine ethics?
- (2) What does it mean to add an ethical dimension to machines?
- (3) Is ethics computable?

⁷ Cf. G. Sayre - McCord, ‘Metaethics’, in (2012) E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition) available at <<http://plato.stanford.edu/archives/spr2012/entries/metaethics>> accessed 1 March 2014.

- (4) Is there a single correct ethical theory that we should try to implement?
- (5) Should we expect the ethical theory we implement to be complete, that is, should we expect it to tell the machine how to act in any ethical dilemma in which it might find itself?
- (6) Is it necessary to determine the moral status of the machine itself, if it is to follow ethical principles?⁸

In this short piece, I will not be able to address all of these points. So I will be content if I can establish that at least some of these questions are more or less explicitly raised – or at least alluded to – by *I, Robot*. For the sake of illustration, I shall take the first and third question as examples.

I believe that anyone who watches the movie with sufficient attention to its philosophically relevant aspects almost inevitably stumbles upon the problem what the ultimate goal of machine ethics should be and which kinds of robots we should aim to build and use, if any. As I said above, there are two types of robots in the movie. There is Sonny – the robot around whom the story revolves. He is clearly capable of moral reasoning. Though he is pre-programmed to follow the three laws of robotics, he can choose to disobey them. As he himself notices, this makes him ‘unique’ and distinguishes him from the second type of robots, that is to say all other robots. Unlike Sonny, they are not autonomous. They are mere machines, which are determined to follow their pre-programmed code. Now the question that the movie implicitly raises is which type of robots we should aim to develop and use, if any. Should we aim to develop robots that are like Sonny – capable, that is, of autonomous agency? Or should we opt for simpler, though still very complicated, machines that are merely automatic? The latter would merely be able to execute pre-programmed commands without any leeway or the possibility to apply the lessons they might have learnt from past experience. Having watched the movie, we may feel inclined to favour Sonny-type autonomous robots over mere automatic machines. But it seems that this is because we have developed sympathies for Sonny – not because we have thought things through. It is, in fact, a very debatable issue whether we should aim to develop robots that can make their own choices. The answer to that question seems to turn on

⁸ I take these questions from S. L. Anderson, ‘Asimov's “three laws of robotics” and machine metaethics’, in (2008) *AI and Society* 22(4), 477-493, 480-481.

the third of the aforementioned issues, viz. whether ethics is computable. If it is not, as e.g. Penrose and Searle may be taken to suggest⁹, no robot will be able to do full justice to the complexity of our moral reasoning and should, for this reason, arguably not be allowed to make morally significant choices. If ethics is not computable, an ethical robot like Sonny, who can rise to the same level of moral competence as a human being, is impossible. Be that as it may, we can conclude, I believe, that *I, Robot* and the character Sonny, in particular, raises interesting metaethical questions in the areas of robo-ethics and machine-ethics. In fact, the movie anticipated them to some extent. Though the discussion about “strong AI” predates the movie, it came out before the debate about these issues in (robo-) metaethics took off.¹⁰

4. *I, Robot* and Applied Ethics

At this point, I shall turn to those aspects of *I, Robot* that pertain to the second branch of moral philosophy, viz. applied ethics. Applied ethics may be loosely defined as the application of moral theory and its methods to concrete moral problems. Movies like *I, Robot* certainly raise a number of interesting and important questions for that area of inquiry, too. One example can be found in a dialogue between Detective Spooner and Lawrence Robertson, who is U.S.R.’s CEO. Spooner, who is, as I have already mentioned, highly critical of the use of humanoid robots, suggests an idea for a U.S.R. commercial.

Spoooner: Look, this is not what I do. But I got an idea for one of your commercials. You can see a carpenter making a beautiful chair and one of your robots comes in and makes a better chair – twice as fast. And then you superimpose on the screen: U.S.R. – shittin' on the little guy.

⁹ R. Penrose, *The Emperor's New Mind* (Oxford: Oxford University Press, 1989); J. Searle, ‘Minds, Brains, and Programs’, in (1981) *Behavioral and Brain Sciences* 3, 417–457; See also J. Nida-Rümelin, ‘Agency, Responsibility and Technology’, (2014) *Politica & Società* 2/2014, 185–200.

¹⁰ Asimov’s original story, of course, predates even Alan Turing’s early contributions to the AI debate.

Spoooner’s thorny remark is obviously aimed to make a point about the business ethics of the U.S.R. company. The robots that they build put people out of work – hard working people, who do a decent job. This issue is, indeed, an important social and ethical concern. The introduction of robotics into manufacturing plants has replaced many workers and made them unemployed – at least temporarily. Is it morally justified for a company to take such measures? And what are our responsibilities as a society to those who lose their jobs and livelihood in the process? As interesting and pressing as these questions are, I do not want to get hung up on them because there are aspects of applied ethics in the movie that I find even more interesting. The business ethical question is, in fact, one that has been discussed up one way and down the other at least since Marx’s and Engel’s time. What makes *I, Robot* interesting for applied ethics, however, is that it raises questions over and above the ones which had been discussed when the movie first came out. It was ahead of its time and has anticipated certain ethical issues.¹¹ One of these ethical problems is related to the metaethical question that I discussed in the previous section. Above I asked what the aim of machine ethics should be. More specifically, I asked which type of robot we should seek to develop. Should we develop robots that are fully autonomous moral agents, who can learn and correct their mistakes as they make their own moral choices (like Sonny)? Or should we content ourselves with robots that merely execute a given moral code without any leeway? No matter how we answer that question, there is a follow-up question to it, which belongs to the realm of applied ethics. If we opt for Sonny-type autonomous robots, then we need to explain what the initial set-up should look like before the robot is left to itself. If we prefer the dumber, more inflexible kind of robot, we still need to decide which rules have to be encoded into the robot’s operating unit. When the movie came out, these questions, it seems to me, did not receive any serious attention from applied ethicists. But they do today.¹² They have recently received book-long treatments.¹³ But it took academics a few years after the movie came out.

¹¹ This, of course, applies even more to Asimov’s original story, which is much older than the movie.

¹² Here is just one example. The American military has been developing and using robots, which are to a large extent autonomous, in order to replace human soldiers on the battlefield (See P. W. Singer, *Wired for War. The Robotics Revolution and Conflict in the 21st Century* (New York: The Penguin Press, 2009)).

Another problem in applied ethics, which the movie anticipates as well, is the question what we may permissibly do to a robot once it is able to follow certain moral rules – or, more generally, once it exhibits certain kinds of properties, such as the capacity to reflect upon itself (like when Sonny asks Spooner: ‘What am I?’). In the movie the robo-psychologist Dr. Susan Calvin gets the assignment to destroy Sonny, which causes her great distress as though she was about to kill an actual person. In the end, she does not go through with it. But even if she had, the question whether, at some point, robots can turn into moral subjects, whom we owe moral regard, would still arise. Other movies have raised similar or related questions. In regards to Stanley Kubrick’s *2001 – A Space Odyssey* (1968) Daniel Dennett has discussed the question who deserves blame when the spaceship’s supercomputer “HAL” kills.¹⁴ When the movie came out (and even much later when *I, Robot* was released), this issue was not widely discussed amongst applied ethicists, but is today.¹⁵ Clearly, then, movies like *2001 – A Space Odyssey* and *I, Robot* not only did a good job when it came to pinning down certain issues in applied ethics. They also anticipated them. And philosophers who picked up on them arguably had a head start in the race for the best answers.

At this point, I could address many further aspects of the movie that should in my view be of interest to applied ethicists. But I will confine myself to one which allows me to present yet another reason why moral

They use them to explore the territory, to dismantle bombs and mines and also to kill enemy soldiers. It goes without saying that the question whether all of this is morally permitted, to which extent and for which purposes is now a hotly debated topic in applied ethics (see, e.g., R. Arkin, *Governing Lethal Behavior in Autonomous Robots* (Boca Raton: Taylor & Francis Group, 2009); R. Sparrow, “Killer Robots,” in (2007) *Journal of Applied Philosophy* 24(1), 62–77).

¹³ See for e.g. M. Anderson, S. L. Anderson (eds.) *Machine Ethics* (Cambridge: Cambridge University Press, 2011); W. Wallach, C. Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford: Oxford University Press, 2009).

¹⁴ D. C. Dennett, 1997. ‘When HAL Kills, Who’s to Blame? Computer Ethics’, in (1997) D. G. Stork (ed.) *HAL’s Legacy: 2001’s Computer as Dream and Reality* (Cambridge, MA: MIT Press); reprinted in this volume.

¹⁵ See, e.g., P. Asaro, ‘A Body to Kick, But Still No Soul to Damn: Legal Perspectives on Robotics,’ in P. Lin, K. Abney, G. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics* (Cambridge, MA: MIT Press, 2011); A. Kuflik, ‘Computers in Control: Rational Transfer of Authority or Irresponsible Abdication of Authority?’, in (1999) *Ethics and Information Technology* 1, 173–184; J. Nida-Rümelin (fn. 9); R. Sparrow, ‘Killer Robots,’ *Journal of Applied Philosophy* 24(1), 62–77.

philosophers should take an interest in sci-fi movies. To this end, I want to return to a point I made earlier. As I explained previously, the movie raises the issue how robots should be programmed to behave. In fact, the movie does not simply raise this issue. It also explores one possible answer – and rejects it. As I mentioned in the introduction, robots are programmed to follow the three laws.¹⁶ This includes V.I.K.I., U.S.R.’s supercomputer. She, too, is bound by the three laws. As I explained earlier, V.I.K.I.’s finds a way to interpret the rules that justifies sacrificing the lives of some humans for the greater good. This is obviously unacceptable and may be seen as a weighty reason to reject the three laws. Unfortunately, the argument seems to me to be a bad one. I, for one, cannot see how V.I.K.I. could possibly arrive at the conclusion that sacrificing humans is admissible taking only the three laws as premises (even though she claims that her “logic is undeniable”).¹⁷ However, it does not change the fact that the movie presents not only a philosophical problem but also a possible solution as well as a rough philosophical reasoning that may help us to evaluate that solution.

5. I, Robot and normative ethics

The third area in which moral philosophers may benefit from watching sci-fi movies is normative ethics, which may be defined, rather vaguely, as the abstract-level study of moral principles and theories. In order to explain how movies like *I, Robot* can add to our understanding of normative ethics we should, first of all, lay out the procedure that theorists in that area of philosophy usually apply in their investigations. Here is a rough sketch.¹⁸

¹⁶ The proposal raises interesting follow-up questions. We may ask, e.g., if and how the three laws can be implemented in a robot, how conflicting implications of the three laws may be resolved and so on. I will, however, not explore these issues further. For a comprehensive treatment of these issues, see W. Wallach and C. Allen (fn 13).

¹⁷ As I said in (fn 6), in Asimov’s original story the three laws are augmented with a fourth (or zeroth) law, which makes it plausible that V.I.K.I. would arrive at the conclusion that human sacrifices are justifiable.

¹⁸ My depiction of the procedure is based on J. Rawls, ‘Outline of a Decision Procedure for Ethics’, in (1951) *Philosophical Review* 60(2), 177–197; J. Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971/1999).

Normative ethicists are interested to find out which moral principles or theories fit our moral sense best overall. To this end, they explore, firstly, whether these principles are in line with certain more fundamental moral convictions that we have and, secondly, whether they can be squared with our judgements about particular cases that seem to us quite obvious. When they do the latter they usually draw on thought experiments. They describe a scenario in which a moral agent faces a morally significant choice between a limited set of options for acting. This scenario would be described such that we have very strong intuitions about the rightness (or wrongness) of at least some of these options. In a next step, normative ethicists observe and record our moral intuitions about the case they have described. They note that this or that act strikes us as right, wrong, obligatory and what have you. Then, they turn to the moral principles or theories they want to investigate. They draw out their implications regarding the case at hand and compare them with the moral intuitions that we have about it. When the moral demands of the principle or theory in question coincides with our strongly held intuitions about the thought experiment, they think of it as corroborated. When not, they regard this as counter-evidence against that principle or theory.

Given this picture of normative-ethical inquiry, it is obvious how movies like *I, Robot* can contribute to the field. As Wartenberg points out, movies can function as thought experiments.¹⁹ They can depict scenarios about which we have strong intuitions. And we can use them to test our moral theories. Obviously, not every movie will provide suitable tests for all moral theories. But some movies will provide some tests for some theories. To show that that is in fact the case, allow me to borrow a point made by Grau.²⁰ He draws on the second backstory to the *I, Robot* movie in order to illustrate how it may support a normative-ethical point that has been made by the moral philosopher Bernard Williams. Williams (1981) seeks to criticize utilitarianism, which is the theory that an act is right if and only if it maximizes the happiness of all sentient creatures in the universe. For one thing, he

¹⁹ A movie can arguably be more effective in stating a thought experiment as it can describe it much more vividly and in more detail. See T. Wartenberg, *Thinking On Screen: Film as Philosophy* (London: Routledge, 2007).

²⁰ C. Grau, "There is no 'I' in Robot: Robots and Utilitarianism", in (2006) *IEEE Intelligent Systems* 21(4), 52-55.

does that by showing that some of the theory's implications are entirely off and sin against our moral intuitions. But he does not only criticize utilitarianism's implications. He also argues against the utilitarian thought process which involves, as he puts it, "one thought too many."²¹ What he means by that is that certain moral judgements – like the judgement that I should keep a promise or be loyal to a friend – are in many instances obviously correct. And a competent moral judge, it seems, would simply acknowledge that. A utilitarian, however, would look for a further reason that supports the judgement. And this fact, thinks Williams, illustrates that utilitarians have inadequate moral thought processes. As Grau diagnoses, the second backstory to *I, Robot*, which revolves around Spooner's accident, exemplifies Williams's point.²² In that story, Spooner is involved in an accident and saved by robot. As I described above, the robot could alternatively have saved a little girl. But it does not because it computes that Spooner's survival chances are greater than the girl's and reasons – presumably on a utilitarian basis – that the right choice is to save Spooner. Spooner's take on that moral choice is entirely different. He evidently takes it to be obvious that the robot should have saved the little girl instead of him and that a human being would immediately have arrived at that judgement. In that case, we may say, it is the "one thought too many" that throws the robot off track and makes it choose the wrong action. If this is the point the movie seeks to make, it is, of course, a controversial one. Not every moral philosopher will buy into it. But it is, nevertheless, an illustration of a philosophical point that may help us to assess its plausibility. To that extent, it is surely worth noticing.

6. Conclusion

In this essay I tried to show that there are various reasons why moral philosophers should watch sci-fi movies. I mentioned a number of them in the introduction. I said, e.g., that sci-fi movies give us the chance for an ideal role change, thus enabling us to see things from another person's (or being's) perspective, which will arguably increase

²¹ B. A. O. Williams, 'Persons, Character, and Morality', in (1981) B. A. O. Williams (ed.), *Moral Luck* (Cambridge: Cambridge University Press, 1-19).

²² See C. Grau (fn. 20).

our capability to empathize with others and may conceivably improve our moral imagination. Using *I, Robot* as a case study, I have argued, furthermore, that movies – and sci-fi movies in particular – can raise important metaethical, applied-ethical and normative-ethical issues. To some extent, they can also depict how certain solutions would pan out. Most importantly though – and this, I think, got particularly clear in my discussion of applied ethics – sci-fi movies can anticipate certain ethical problems that technological innovations may bring, thus flagging new ethical issues before they are discussed in the academic community. While most reasons that make sci-fi movies interesting from a moral-philosophical viewpoint seem to apply to movies in general, this point is, I believe, specific to the sci-fi genre. Since good sci-fi movies are based on informed scientific speculations²³, their specific content matter makes it much more likely that they will bump up against new problems in the ethics of technology. It seems to me, therefore, that philosophers not only have a good reason to watch good movies. They have an excellent reason to watch sci-fi movies in particular.

As I suggested in the beginning though, I do not want to overstate my thesis. Moral philosophers can, of course, go about their business and produce great work without ever coming into contact with fictional sources. What counts in moral philosophy – as well as in philosophy generally – is reasoned thought and sound argument. It does not matter where the inspiration comes from. The role of sci-fi movies is thus not essential to doing moral philosophy or philosophy in general. Furthermore, there are certain problems in moral philosophy (e.g. formal-logical analyses) that cannot be elucidated through film. Nevertheless, what I said shows, I hope, that sci-fi movies can enrich our thoughts and inspire us when we do moral philosophy. To that extent, I believe, moral philosophers should find good sci-fi movies worth a watch!

²³ In the history of sci-fi there have been a number of prophetic figures, who anticipated modern technological developments many decades before they took place. Already the first sci-fi writer and father of the genre, Jules Verne, is celebrated for his uncanny accuracy in predicting future events as, e.g., the moon landing in his book *From the Earth to the Moon* (London: Routledge, 1865/1890). His successor H. G. Wells predicted, amongst other things, the invention of the tank (see his “The Land Ironclads” (1903/1927) in, *The Short Stories of H. G. Wells* (London: Ernest Benn, 1903/1927)) and the atom bomb (see his “The World Set Free” (1913/1914), in *The World Set Free – A Story of Mankind* (London: Macmillan, 1913/1914).

Lessons in humanity, or: what happens when robots become humans

Nathalie Weidenfeld

1. Robots in film

Robots and film have a long history together. The first robot who appeared in a film is Fritz Lang’s *The Eve of the future* (1896). Robots in films have since then relentlessly been haunting the silver screen. Indeed it is difficult to find a contemporary science fiction film, where robots do not appear and are an integral part of the story. Robots in film usually belong to three types of categories: They are either uncanny creatures, monsters, or robots who want to be and appear to be just like humans. While the robot as an uncanny creature plays on infantile wishes and fears Freud has identified to be at play with the uncanny¹, robots as monsters appeal to our basic assumptions of what we judge to be impure or what is suppressed in society². In narratives in which robots are almost undiscernible from humans they are often portrayed to have ontological problems: these robots want to know who they are and often want to be regarded as equal to humans. Even though there may be a certain mixtures of these categories at play, there is also a chronological order in this classification. While first robots were mostly uncanny creatures, during the 40ies and 50ies they evolved into monsters. Today’s robots mostly belong to the third type. It is these robots-as-humans I want to take a closer look at in order to ask what

¹ S. Freud, ‘The Uncanny’, in (1919) T. Dufresne (ed.) *Beyond The Pleasure Principle* (transl. G. C. Richter) (Ontario: Broadview Press, 2011), 102-106 ff.

² B. Creed, *The Monstrous Feminine* (New York: Routledge, 1993); R. Wood, ‘Introduction to the American Horror Film’, in B. K. Grant, C. Sharrett (eds.) *Planks of Reason. Essays on the Horror Film* (Maryland: Scarecrow Press, 2004).