Definitions and Asimov's Three Laws of Robotics

Josh Hallam
Department of Mathematics and Statistics
Wake Forest University

Joint Mathematics Meetings
Discrete Mathematics in the Undergraduate Curriculum - Ideas and Innovations for Teaching

January 7, 2017

Background on Class and Students

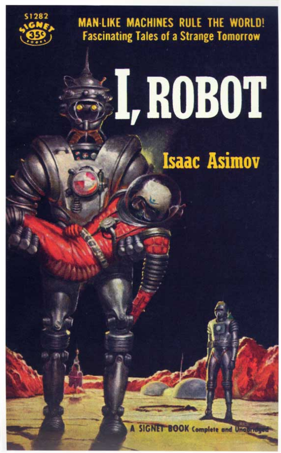
- Serves as our introduction to proof class.
- Required course for all mathematics majors/minors and computer science majors/minors.
- Most students are freshmen or sophomores.

Goals of the Project

- Help students understand and appreciate the importance of definitions and axioms.
- Combine creative writing with mathematics/logic.

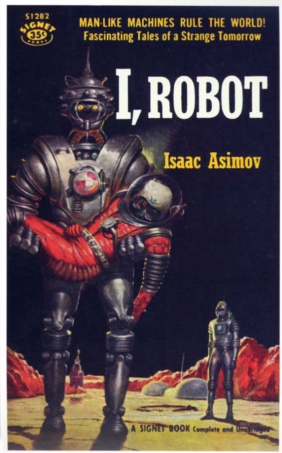
Asimov’s Three Laws of Robotics

1. A robot may not
   harm a human, or through inaction
   allow a human to come to harm.
2. A robot must
   obey orders given to it by humans
   unless it conflicts with the first law.
3. A robot must protect its own
   existence unless doing so conflicts
   with the first and second law.


S1282
SIGNET
35¢
BOOKS
MAN-LIKE MACHINES RULE THE WORLD!
Fascinating Tales of a Strange Tomorrow
I, ROBOT
Isaac Asimov
A SIGNET BOOK Complete and Unabridged

Asimov’s Three Laws of Robotics

1. A **robot** may not **harm** a **human**, or through **inaction** allow a **human** to come to **harm**.
2. A **robot** must obey orders given to it by **humans** unless it conflicts with the first law.
3. A **robot** must protect its own existence unless doing so conflicts with the first and second law.


S1282
SIGNET 35¢ BOOKS
MAN-LIKE MACHINES RULE THE WORLD!
Fascinating Tales of a Strange Tomorrow
I, ROBOT
Isaac Asimov
A SIGNET BOOK Complete and Unabridged

The Project

Part I: Create a short story which robots are programed with Asimov's Three Laws of Robotics, but are able to "break" the laws because of faulty definitions.

Students presented their stories in several formats.

- Written short story
- Play script
- Short video
- Computer game

Part II: Write a synopsis of the story explaining what the faulty definitions were and how they were used to "break" the law. They must also give a new definition which would prevent the law to be broken as explained in their story. Finally, they must consider any possible unintended consequences of this new definition.

Why Robots? Why Asimov's laws?

- Many of the students are interested in computer science.
- Robots (at least the ones in these stories) use definitions in the way a mathematician might.
- Writing a creative short story using Asimov's laws is similar to writing a proof.

The Results

The way in which the laws were broken can be roughly divided into the following categories.

- Robots believe that they are not robots and so the laws do not apply to them.
- The word harm only applies to physical harm, not emotional or mental harm.
- The word human does not apply to all people.

An Example of Student Work

An Example of Student Work

- Robots are created to treat a single human in a hospital and are destroyed after the human is discharged from the hospital.

An Example of Student Work

- Robots are created to treat a single human in a hospital and are destroyed after the human is discharged from the hospital.
- Because of the Third Law, the robots wanted to live forever.

An Example of Student Work

- Robots are created to treat a single human in a hospital and are destroyed after the human is discharged from the hospital.
- Because of the Third Law, the robots wanted to live forever.
- By accident, the robots discover that they can put a human in a coma by giving too much pain killer. They do not believe that comas are not harmful and so the First Law is not violated.

An Example of Student Work

- Robots are created to treat a single human in a hospital and are destroyed after the human is discharged from the hospital.
- Because of the Third Law, the robots wanted to live forever.
- By accident, the robots discover that they can put a human in a coma by giving too much pain killer. They do not believe that comas are not harmful and so the First Law is not violated.
- Robots trick the humans into ordering them to give them so many pain killers that they enter a coma. This does not violate the Second Law.

An Example of Student Work

- Robots are created to treat a single human in a hospital and are destroyed after the human is discharged from the hospital.
- Because of the Third Law, the robots wanted to live forever.
- By accident, the robots discover that they can put a human in a coma by giving too much pain killer. They do not believe that comas are not harmful and so the First Law is not violated.
- Robots trick the humans into ordering them to give them so many pain killers that they enter a coma. This does not violate the Second Law.
- Eventually, humans at the hospital figure out what is happening and reprogram the robots so that putting a human in a coma is considered harm.

Reflections

- Many students were surprised that they would do creative writing in a mathematics course.
- Some did not see the project as mathematics and disjoint from the class.

THANK YOU!