An edited version of this appeared in *The AISB Quarterly* Issue 133 (Spring 2012) page 14-15.

The Making of the EPSRC Principles of Robotics

Joanna J. Bryson
Artificial Models of Natural Intelligence (AmonI) Group
Department of Computer Science, University of Bath
http://www.cs.bath.ac.uk/∼jjb

December 5, 2011

In late 2010 the EPSRC unexpectedly invited me to attend a meeting in the New Forest on the topic of Ethics and Robots. I have been writing occasional articles on that topic since 1996, in response to my experience of being on Cog (a humanoid robot project) and seeing how readily and even insistently people attributed moral obligation towards a completely non-functional (in 1993) but vaguely humanoid robot. Since 1998 I've also maintained a web page on the topic. Nevertheless this was the first time I'd been approached by a government body, and of course I said yes.

The meeting was a three-day offsite chaired (expertly) by the journalist Vivienne Parry. Besides myself, other participants included Margaret Boden, Darwin Caldwell, Kerstin Dautenhahn, Paula Duxbury, Lilian Edwards, Ann Grand, Hazel Grian, Sarah Kember, Stephen Kemp, Paul Newman, Geoff Pegman, Andrew Rose, Tom Rodden, Tom Sorell, Mick Wallis, Shearer West, Blay Whitby, and Alan Winfield, as well as able assistance from Ian Baldwin, Denise Dabbs and Paul O'Dowd. Participants came mostly from robotics, but also from the humanities, law, and social science. We were employed mostly in academia, but also by industry and the research councils (including the then-head of the AHRC, Shearer West).

I was surprised the EPSRC would splash out for so many people for so long in such a nice hotel on this topic, but they made the object of their concern very clear quite early. The EPSRC see robotics as a critical technology for the UK, and does not want to see it face the same fate as other "futurist" technologies have, in terms of public distaste bordering on hysteria that can no longer be addressed by any amount of measured scientific assessment. The EPSRC wants to get robot ethics right from the beginning, to ensure both the safety and the acceptance of robotic technologies.

It became almost immediately apparent that they had succeeded in selecting a very pragmatic and socially-concerned group of experts. The group took a very strong line on what the moral and ethical role of robotics could be, and one that I would not say is the dominant one at typical AISB gatherings.

On the final full day, Internet Law professor Lilian Edwards and I were in a

small break-out meeting together in a session intended to design deliverables as outcomes for the meeting. We decided to make a "real" set of laws for robots. Lillian was keen to have them clearly follow but correct Asimov's laws, while I was keen to include several I'd already developed while writing *A Proposal for the Humanoid Agent-builders League (HAL)* (Bryson, 2000). In the end we settled on five, the first three of which reflect and refract Asimov to the concerns of the group. The group as a whole then refined not only our "laws", but also ordinary language versions of these, and developed a further list of concepts to be communicated to you, our colleagues.

The full version of these documents can now be found by Googling the *EPSRC Principles of Robotics*, and they have become EPSRC policy since April of 2011. The basic principles are reproduced in the box. Below are seven high-level ideas that the group wants to communicate to you, our colleagues. For detailed explanations, please see the website, but I have given the highlights here.

1. *We believe robots have the potential to provide immense positive impact to society. We want to encourage responsible robot research.* We are not a bunch of luddites who "don't get" the real potential of AI. We are concerned professionals who really do want to make AI work and robots real.

2. *Bad practice hurts us all.* We can't ignore the situation if some of our colleagues do things that make all of us look bad.

3. *Addressing obvious public concerns will help us all make progress.*

4. *It is important to demonstrate that we, as roboticists, are committed to the best possible standards of practice.*

5. *To understand the context and consequences of our research we should work with experts from other disciplines including: social sciences, law, philosophy and the arts.* We were all struck by how much we learned from this multi-disciplinary working team.

6. *We should consider the ethics of transparency: are there limits to what should be openly available?* Everyone at the meeting was committed to open-source-software type solutions and approaches, but we came to realise that with robots and AI more generally we do have the obligation to make sure that every "script kiddy" couldn't hack into a system that has information or memory about the private lives of humans.

7. *When we see erroneous accounts in the press, we commit to take the time to contact the reporting journalists.* Most science reporters really don't want to be made to look silly by reporting an "expert" who turns out to be self-promoting or sensationalist. A quiet word or email can often damp hysteria being generated by irresponsible statements.

I would like to thank the EPSRC and also our colleagues who advocated for this meeting. Two of these latter were Alan Winfield and Tom Rodden. Personally I feel extremely proud and happy for my profession and nation that the UK now has an official set of robotics principles that address such important matters. But we are only one country, and there is still much work and advocacy to be done to ensure that intelligent robotics are used appropriately in our society.

Box!

1. *Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.* While acknowledging that even dead fish can be used as weapons by creative individuals, we were concerned to ban the creation and use of autonomous robots as weapons. Although we pragmatically acknowledged this is already happening in the context of the military, we do not want to see these used in other contexts.

2. *Humans, not robots, are responsible agents. Robots should be designed & operated as far as is practicable to comply with existing laws & fundamental rights & freedoms, including privacy.* We were very concerned that any discussion of "robot ethics" could lead individuals, companies or governments to abrogate their own responsibility as the builders, purchasers and deployers of robots. We felt the consequences of this concern vastly outweigh any "advantage" to the pleasure of creating something society deigns sentient and responsible. This was the law we knew would most offend some of our colleagues in AISB — consequently (with David Gunkel) I am running a symposium at AISB 2012 to examine whether this is a reasonable rule. The symposium is called "The Machine Question: AI, Ethics and Moral Responsibility".

3. *Robots are products. They should be designed using processes which assure their safety and security.* This principle again reminds us that the onus is on us, as robot creators, not on the robots themselves, to ensure that robots do no damage.

4. *Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.* This was the most difficult rule to agree on phrasing for. The idea is that everyone who owns a robot should know that it is not "alive" or "suffering", yet the deception of life and emotional engagement is precisely the goal of many therapy or toy robots. We decided that so long as the responsible individual making the purchase of a robot has even indirect (e.g. Internet documentation) access to information about how its "mind" works, that would provide enough of an informed population to keep people from being exploited.

5. *The person with legal responsibility for a robot should be attributed.* It should always be possible to find out who owns a robot, just like it is always possible to find out who owns a car. This again reminds us that whatever a robot does, some human or human institution (e.g. a company) is liable for its actions.

See also (**note to editor**, please combine the web pages & the references into one list for a total of five items):

- The EPSRC Principles of Robotics Website.
- The Machine Question: AI, Ethics & Moral Responsibility.
- Ethics: AI, Robots and Society.
- (Wilks, 2010)

References

Bryson, J. J. (2000). A proposal for the Humanoid Agent-builders League (HAL). In Barnden, J., editor, *AISB'00 Symposium on Artificial Intelligence, Ethics and (Quasi-)Human Rights*, pages 1–6.

Wilks, Y., editor (2010). *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues.* John Benjamins, Amsterdam.