When Robots Die:

A conversation between Emmet Cole, journalist at Wired magazine UK,
and Nikolaos Mavridis, Asst. Professor of NYU AD

E: I am a journalist working on a piece for Wired U.K. about human robot interaction.

I was talking with HRI expert Adriana Tapus recently, the lead scientist on the HRIAA project to develop a control architecture for companion robots in a therapeutic setting. She was talking about some of her previous research in the area of human-robot interaction (in the therapeutic context) and mentioned, almost as an aside, a story that I found interesting and quite touching.

In essence, it is that an old lady with dementia had been interacting with a robot for several hours a few times a week over a period of about a year. The outcomes were successful, but when the project came to an end and the bot was taken away, the lady was quite distressed and depressed over the whole thing. She had separation anxiety --and (it seems reasonable to conclude), therefore had a meaningful relationship with the robot. {separation anxiety-sympt/diagn}

That got me thinking: Isn't this the kind of small drama that will take place thousands (if not millions) of times over in the future when robots become more commonplace?

N: It might well be the case, that indeed, this small drama will be more commonplace in the future…But, important differences might exist, too!

E: I imagine a future in which loyal robots, which have served us for years, are treated more like pets (to the extent that pets are already treated as family members, at least) than inanimate objects.

N: The crucial point here, which you indirectly point to, is the question of what kind of mental model will humans utilize when it comes to how they conceptualize companion robots – and the interactional affordances of humans with their electronic friends, as well as the resulting behaviors and rituals, will therefore follow. But how many different kinds of such mental models can exist? The simplest view would only postulate too: "things" that we construe as "inanimate objects" and "things" that we construe as "animate, living beings". Of course, upon closer inspection, this is certainly an oversimplification; there is a wide spectrum of entities which we construe as animate. Of course, there is a very strong tendency towards anthropomorphization of the mental models that we use for many different kinds of animate beings; notice for example how easily we attribute for human-like "beliefs", "emotions" or "intentions" to pets, and even to insects – when scientific evidence clearly shows that what we could call an "intention" for example, is quite different between humans and insects. However, apart from the strong tendency towards anthropomorphization which provides a unifying substrate, we also maintain distinctions between the different kinds of mental models that we use for various kinds of animate entities – which also extend to having different emotional reactions to events involving them (for example, accidentally killing a ladybug vs. your cat), as well as different norms and ethics towards them (spraying insects; slaughtering a sheep that you had in your garden), and so on. And certainly, cultural as well as personal variations do exist, too.
Now, robots, and even more so interactive companion robots with which we maintain repetitive longer-term contact, are quite a new entity that will slowly start entering our lives; and the question thus becomes: where will the mental model that we elicit for robots fit within the above spectrum of possibilities, and what will be the emotional implications and ethical rules governing our relations with them?

This is a very interesting question, and at this stage, lacking adequate empirical evidence for a definite answer, we can only speculate. Usually, there seem to be a number of factors that contribute towards the selection of mental model that we use for different entities. First, our ***mode of interaction*** *with it is of primary importance: the frequency, duration, temporal horizon, means (verbal, tactile etc.), as well as naturalness, and also the perceived autonomy of the entity – do we need to direct every little step it does, or will it take intiative? Second, the* ***appearance*** *as well as the* ***behavioral repertoire*** *of the entity – what does it look like, as compared to our existing repertoire of entities, and how does it seem to move or react? Third, the* ***role*** *that we assume to have within the interaction is important: is this a master-slave role, is it an equal partner role, is it a tutor-learner role? And fourth, quite importantly, the* ***pre-existing expectations*** *that we might have formed about the entity, coming usually through fiction or film, although we might have never seen it or interacted with it before. This is especially*

important when it comes to our mental-models for robots: Having been exposed to quite some Sci-Fi, we already have a strong expectation of how a robot should "look like" and "sound like", how it should "react like" and "think like" and so on, although actual robots might well be very different than fictional ones, and this disconnect might create unpleasant or pleasant surprises.

Thus, the above discussion, clearly shows that there seem to be a rich number of factors that affect the mental model that we construe for the entities that we interact with, and also our emotional reactions and ethical norms towards them – and thus a rich terrain exists for open questions waiting to be scientifically investigated! And, numerous early results have started to shine some light on this rich terrain of open questions. For example, Riek has shown that anthropomorphic form seems to lead to much stronger feelings of empathy towards robots: our reactions when we see somebody hitting an android are very different to somebody hitting an industrial robotic arm. Also, many other interesting effects have been observed; for example, perceived gender can be important, as well as

However, one important addition to the above terrain, which I think is quite crucial in order to understand longer-term human-robot relations, is the concept of "Sharing", and more specifically Building and Maintaining a metaphorical ***"Common Locus"*** *("Koinos Topos", in Ancient Greek), which forms the backbone of a meaningful and sustainable Human-Robot relation. But exactly does this mean? Think about human-human relations. What really unites you with a good friend? Arguably, the creation of a shared "Self" between and beyond "I" and "You" is of primary importance. But again, what does this really mean? Some of the more permanent elements that could construct an "I" are not so difficult to list: my "self" consists of my memories, my friends and social circle, my interests and taste, and more. But then, what is this "Common Locus"? It starts, and builds on, exactly on the sharing of these elements: The metaphorical "Common Locus" between two humans (which are in relation), is made up of all these shared elements: shared memories (what they have lived together, and what they have experienced in common), their shared aquaintances and friends (given that we don't live in isolation; but are deeply embedded within our social network), their shared interests and tastes; including also more fundamental shared elements, such as a shared language of communication.*

Thus, the hypothesis is that in order to have a meaningful and sustainable longer-term relationship between a human and robot, a strong "Common Locus" that grows and transforms over time is of primary importance. A simpler version of this hypothesis, was the basis of our "FaceBots" social robots project, which received grant funding from Microsoft ER: we postulated that having a robot that builds "Shared Memories" and "Shared Friends" with humans would lead to much more meaningful relations. And it is not difficult to understand why this seems like a reasonable hypothesis: just think of a friend, and what you talk about when you meet together – and also the deeply bonding feeling of recalling common memories and acquaintances from the past.

Thus, summing up: What kinds of mental models do we construe for different types of entities that we live together and interact with? Quite a spectrum, across the inanimate and human poles; which includes many different kinds of animate entities: from plants, to insects, to pets and beyond. And what factors affect our choice of mental model, as well as our emotional reactions and ethical norms towards them? An interesting range of factors – a non-exhaustive list of which, organized across four categories, was given above: mode of interaction, appearance and behavioral repertoire, role, pre-existing expectations, and so on. And quite importantly, when it comes to longer-term relationships, the concept of a growing "Common Locus": shared memories, friends, interests…

An interesting observation, though, arises: the possibility for an "asymmetric" common locus also exists. I.e. what might really matter in a human-machine relationship is not necessarily the question of whether shared memories, friends, and language do exist bilaterally; it is sometimes enough that unilateral memories held by the human of moments with the machine exist, together with their "imaginary" completion – i.e. by knowingly or unknowingly fooling ourselves that the machine remembers the moments we had together (shared memories), can understand our words towards it (shared language), and so on. Of course, this is not a phenomenon that arises only for human-machine relations (i.e. our old laptop or car and the times we had with it); we could as well have had created an asymmetrical "Common Locus" by sharing memories and words with the mountain whose peak we could see out of the window of the family house that we grew up in, or with our favorite pen during primary school.

E: For instance, when your loyal companion bot no longer works or needs to be replaced-- will we bury them in the back garden? Will we house the old bot in a glass case in the family home? Or keep a head on the mantelpiece?

N: This is a very valid question, and I would not want to presuppose an affirmative or negative answer at this stage. What kind of rituals of farewell or of preservation might develop remains to be seen; chances are, they might start as being similar to the ones that we have for pets or even humans. However, note that there are also many differences: an important one of which is the fact that the "death" of a robot is neither as terminal or as irreversible as the "death" of a human currently is. A robot's "memories" or "mind" could easily be transferred over to a storage medium (hard disk), and then "resurrected" with a different mechanical body and electronic brain. We could even create "selective amnesia" for our robot, and bring it back to the mental state it had some years ago, by restoring an older backup. Furthermore, notice that robots currently have no "offspring" in the human sense of the word (to provide some form of "continuity"), but again their mental content could easily be copied to other robots, edited or combined; thus, human "death" and robot "death" are of quite a very different nature, not so much in terms of the underlying physiological differences, but mainly because of the immediate possibilities for forms of "continuity" beyond death that exist for robots, which do not exist for humans.

E: No says Tapus. Our attachment to bots will be different. When a coffee machine breaks, she told me, we throw it out, we don't bury it.

N: Hmmmm…. But I still do keep my old ZX Spectrum home computer, even if it is not functional anymore. And I do spend some times with it, rarely, but not never. And, I also enjoy playing with an "emulated" ZX Spectrum running within my laptop: so in a sense, I am still "resurrecting" it and interacting with it, every now and then. And this helps me bring back a wide range of "assymetrically" shared memories with it: I can almost instantly feel as fresh as I did during my school years, just by interacting with my "resurrected" old home computer. Thus, I wouldn't be that absolute in my answer to this question. Quite the contrary!

E: I'm not an expert on HRI --that's why I am writing to you-- but I am not so sure. Our relationship with a coffee macihne is very different to the type of relationship one might have with a walking, talking, humanoid companion.

N: I very much agree, on the basis of my previous arguments and experience, as well as on the basis of the existing relevant academic research. A walking, talking, humanoid companion is interactive; exhibits anthropomorphicity in form and partially in behavior; and might even have shared memories and shared friends with us, as for example the FaceBots robots that we have created in my lab do. Thus, many good reasons for the observed difference in the relationship!

E: Do you think bots will be treated like the family pet or the coffee machine? Or something else entirely?

N: On the basis of the theorization on the different kinds of mental models, emotional impact and ethical norms, that I have given above, I think that we will start by re-using elements of our existing kinds of such mental models (the ones we have for humans, pets, and machines we spend a long time with – such as cars – and which seem to "respond" to us), biased with our Sci-Fi and film-induced expectations of what robots should behave like. Then, slowly, as Robots enter our everyday life more and more, and as they become more and more intelligent, there will arise a new special kind of mental model for them; however, this will still be based on elements from the existing kinds of mental models, mixed, adapted and extended with new elements. Thus, slowly it will become something new; with strong human elements; but neither a copy nor something new "entirely".

E: Just what status will robots have in our lives (especially if we are hanging around with the same, presumably humanoid bot, for years on end)? Will we just dump/recycle their parts?

N: Another great question, towards an answer for which the maximum we could do right now is just partially informed speculative prediction; while waiting for a clearer theoretical picture, as well as most importantly for empirical data. I think that the attachment will not be so much with the physical parts ***(body)*** *of our long-term companion robots (which, again unlike humans, could potentially be*

repairable or replaceable to a very large extent); but, we will be attached to the mental parts ***(mind)*** *of our robot friends, i.e. the specifics of the experiences and the behaviors and the words and common memories and all these things that constitute the projection of the "Self" of our robot friends within our mind, which become immediately accessible by interacting with them. But given the possibility of making the mind of such robots "revive" within a new electronic brain, and even in a different body, we will be able to continue interacting with them – as I still do with my old ZX Spectrum home computer, living in "emulation" within my laptop. Thus, even if recycle/dump parts of their bodies, there will be the possibility to still be with their minds and converse with their voices, if we want to.*

E: As mentioned already this is a speculative "thought piece" rather than a hard news story. I am fascinated to know what form you think human robot relationships will take in this not-so-distant future and especially what you think will happen "When Robots Die."

N: Finally, a small side note, which would probably deserve much more space, or a totally special feature. There is yet one more huge difference between the robots and other artificially intelligent entities of the future, and the ones that we are used to today. Classical robots have an electronic "brain" within their body (a computer board with CPU, memory, and the such), and they are running a periodically updated "mental program" software, while keeping their mental contents (interaction memories and the such) in a storage medium. But, already, networked robots have started to appear; which might have many different computers in various locations acting as their brain. Also, their behavior might not be the result of a single "mental program" software; but of the interaction of many. And most importantly, ***networked robots*** *will slowly utilize not only* ***information from the internet*** *(as our Facebots robots do, by accessing and using information from Facebook, or other robots by using google maps etc.), but also* ***services available on the internet*** *(such as a Face Recognition service, a Google Translation service etc.). Thus, slowly the "mind" of such robots will be the Internet, or at least a dynamic part of the Internet. And finally, these web-accessible services* ***might not only be electronic*** *in nature (software programs running on remote computers) – but* ***might contain human components too****. These human components will not be human cells or brains in a vat connected by wires – but rather humans connected through facebook or through their cellphones or through online games, which perform tasks (recognize objects, provide expert opinions on a chest x-ray, and do there such tasks which are currently difficult or sensitive for machines). Thus, the brain of the robots of the future, will reside within a mixture of the cumulative electronic as well as human brainpower of the earth, interconnected through networks – with different cognitive, sensing, and actuation elements fluidly joining and leaving these spontaneous "interconnected minds". And of course, in this way, the robots of the future, will also have distributed bodies as well as minds, consisting of both human and machine elements. This is part of the* ***"Hybrid Human-Robot Cloud"*** *proposal, an extension of cloud computing that includes not only transparently distributed computation and storage services; but also sensing and motor services with human as well as electronic elements, the manifesto of which we are currently drafting with a number of international academic collaborators, covering a range of fields – from robotics and AI, all the way to psychology and ethics. The future will be much more interconnected!*