

Irrational Machines: Robots, humans and the monsters of reason

Michael Szollosy¹

Abstract. Robots and AI exist — both in concrete terms and, most especially, in the human imagination — at the contested intersection of reason and emotion. On the one hand, our intelligent machines are the epitome of reason — they are thinking machines without the messy complication of emotions. They thus exist at the apex of our rationalist aspirations. On the other hand, there is a great deal of fear and even expectancy that we will soon as emotionally-aware robots and AI. In the meantime, we relentlessly anthropomorphise machines, attributing to robots and AI all sorts of motivations, and they bear the burden of so many human anxieties and aspirations about the future: from massive job losses to (trans)human enhancement to human genocide. This paper asks, what are the consequences of this situation of robots and AI at the intersection of reason and emotion, rationality and irrationality? It is vital that we come to some understanding of how robots and AI have come to occupy this position, and not simply to dismiss the public imaginations as errors that need to be corrected: the popular perception of robots tell us a great deal about both our present human selves and our future with robots and AI.

1 INTRODUCTION

When we talk about robots and artificial intelligence (AI), we need to distinguish between robots and AI as they really exist, in the labs (and factories and hospitals) of the twenty-first century, and those that exist in the cultural imagination. These are very different, though intricately related entities. However, it is a tragic mistake — truly, as it is so often firmly rooted in *hubris* — to simply dismiss the public perception of robots and AI as errors that can be, and need to be, corrected. As I regularly remind people: robots were invented in the public imagination long before they were ever built in laboratories, and the legacy of those first robots — the *robota* of Karel Čapek’s play *R.U.R.*, the *Maschinenmensch* of Fritz Lang’s *Metropolis*, Televox, Asimov’s Robbie, Hal 9000, Roy Batty, T-800 Model 101 and the rest — on both public perception and on roboticists’ actual efforts cannot be underestimated. In many cases, the public’s imagination has influenced or even dictated the design and implementation of robots and AI (e.g. [15] [24] [23]).

We need to expect and prepare for the fact that the public’s imagination will play a crucial role in the development and implementation of robots now and in the future. Consider, for example, the notorious case of genetically modified foods or organisms (GMOs), which should by now be an object lesson for all researchers in robotics and AI: GMO technologies were expected by many researchers to bring massive benefits to all ‘stakeholders’, but since the 1980s GMOs have been increasingly met with suspicion and hostility by those

very groups, such as farmers and consumers — i.e. *the public* — that were meant to benefit from their introduction and mass application [26]. Whatever the merits of GMOs, or the motivations of the scientists and corporations that tried to bring this research to markets, it is widely agreed that there were serious problems with the manner in which GMOs were introduced to, or some might say *foisted on*, the public. As we discussed at length at the AISB conference in Sheffield in 2016, the Engineering and Physical Sciences Research Council (EPSRC) convened the retreat that led to their ‘Principles of Robotics’ [2] precisely to prevent new developments in robotics and AI foundering in the same way that GMOs had done [5]. It is encouraging that the words ‘responsible research’ are now so often heard in or at least near engineering and computer science departments, though the depiction of robots and AI in films, video games and, in particular, popular newspaper headlines show that we still have a long way to go, and much to be concerned about.

Robots and AI are situated at a very paradoxical place, at a highly contested and anxiety-inducing intersection between reason and emotion. By looking at how robots and AI exist in the cultural imagination, we can learn a great deal about human beings, the way in which we perceive and even construct ourselves as rational and irrational beings; we can learn which domains we reserve for reason and which for emotions, and how the contested spaces between the two give rise to great anxiety, about who we are and what lays in store in our future. Additionally, the cultural perception of robots and AI has and will continue to shape and govern research into robotics and AI. These perceptions may have negative impact upon research and the future implementation and use of robots in our societies, and only by examining in detail these perceptions can we prevent damaging limitations on beneficial technological development, and create robots and AI that are genuinely welcome and useful in society.

This paper begins with a short historical contextualisation of the invention of robots in the cultural imagination, and explains how this history still has a tremendous impact upon the public’s anxiety that could undermine valuable robotics research. I will then consider more specifically how these ideas about robots and AI have particular impacts on social robots and human-robot interactions, and play an important role in shaping policy, looking at the EPSRC’s 2010 Principles of Robotics [2] and the more recent EU Committee on Legal Affairs report with recommendations to the Commission on Civil Law Rules on Robotics [9], which was accepted by the European Parliament’s Legal Affairs Committee in January 2017. Finally, I will consider ways of approaching our confused emotional relationships with robots, in an effort to avoid the worst pitfalls that these relations threaten to drag us into: one from Sigmund Freud’s thinking on how to confront the paradoxes of emotion and reason and, secondly, from the very (science-)fictions that give voice to our anxieties.

¹ University of Sheffield, UK, email: m.szollosy@sheffield.ac.uk

2 CONTEXTS

If we are to appreciate the importance of understanding what robots and AI mean in the public imagination, we first need to understand how these particular images of robots and AI have come to be lodged there. Unfortunately, this entails nothing short of an overview of two thousand years of Western history [16] and comparing the ideologies of world religions [21]. A grasp of how both capitalist media and political parties manipulate public fear and anxiety would certainly help as well (e.g. [14]). However, a shorter version of this — covering little more than two hundred and fifty years — can suffice for our purposes here.

It is no accident that human beings began to fear rational monsters such as robots and AI at the point that we invented the very idea of 'reason' and decided to make ourselves more rational. As Min-soo Kang [16] explains, the very idea of the automata underwent a radical transformation in the Enlightenment: from being a marvelous *machine capable* of independent action to a *human being incapable* of independent thought and action. The ideal, rational man of the early Enlightenment, which was enamoured with the mechanistic view of the world, became the incomplete being, an emotionless, unimaginative conformist in the later, more idealist Enlightenment [16].² We can thus in the eighteenth century already see the ground being laid for our present tensions between robots and humans, between the rational and the emotional: should humanity aspire to emulate the perfect reason of machines, to be governed by rationality and science? or are the consequences of abandoning our emotion, animalistic selves too great, and instead should we take as our ideal the irrational passions of the individual genius?

These tensions and concerns continued in the nineteenth century with increasing concern about the effects of the Industrial Revolution, manifested most famously in the Romantic movement, which provided the prototype that even today dominates our popular conception of robots with Mary Shelley's 1818 novel, *Frankenstein: A Modern Prometheus* [20]. It is important to remember, however, that while Frankenstein's monster has become the basis of so many modern robotic monsters — from the False Maria of Lang's 1927 film *Metropolis* [17] to Ava of Alex Garland's 2015 film *Ex Machina* [13] — it is actual Dr. Frankenstein, the archetypical 'mad scientist', who is the subject of Shelley's novel. And while the Romantics were enamoured of the 'modern Prometheus', the brave new superhumans that stood above mere mortal man and would drag the human race up from our barbarity, they were also ambivalent about such figures. Shelley's novel expresses fears about the effects of both rationality and the ir/rational pursuits of the mad scientist. Thus in Dr. Frankenstein we can see deep ambivalence and anxiety about what the pursuit of science and reason are doing to human beings.

The fear that human beings are being de-humanised, resembling the soulless automata capable of nothing more than executing a basic programme, is shared by many important thinkers in the late nineteenth century (e.g. Marx [19]), and the concern that human beings are forever condemned to suffer in an eternal struggle between rational demands and irrational desires is the fundamental assumption made by Freud and his myriad of followers [3]. These ideas are carried into the twentieth century, when we see the first robots invented in Karl Čapek's play *R.U.R.* [7], from where we get the word *robot* itself, the Czech word *robota* meaning something akin to a forced labourer or a slave. Though Čapek's robots are not actually mechanical in the way we understand robots today — they are more akin

to clones or genetically constructed humans — we see in his play a continuation of the eighteenth- and nineteenth-century fears and the foundations of our own robot mythologies:

- reason and science are transforming human beings into machines, i.e. something less than human, and missing some essential human quality
- the transformation of work, turning people into nothing more than cogs in an industrial machine, dehumanises humans
- at the same time, paradoxically, it is feared that robots (and later AI) will take human jobs
- eventually, robots (and later AI), having made humans redundant, will destroy human beings and take over the world

The last point may be the dystopian *telos*, though most commentators today would accept the first three points as valid concerns. It is this transformation of human beings into sub-human machines that has given rise to our anxieties about robots and AI, more than anything that real robots and AI themselves have done. This is why the psychopathic, genocidal robots of science-fiction are so often indistinguishable from the humans they seek to destroy (e.g. *Blade Runner*, *Terminator*, *Battlestar Galactica*, *Ex Machina*); these robots are just a reflection of the rational monsters we ourselves are becoming. We fear the blurring of that boundary; we fear not being able to tell what is human and what is machine, where one starts and the other ends.

As versions of us, these machines are not merely reflections of human rationality: they *also* embody the most irrational, animalistic, impulsive aspects of humanity. This makes robots all the more terrifying, because they represent both these anxieties about the consequences of *reason* — the inflexible doctrine of rationality, the precision and productivity of technology, the dispassionate methodology of science — *and* our savage, animal impulses: the irrational violence, the impulsive desires for domination and control.

3 EMOTIONALLY ENGAGING WITH (SOCIAL) ROBOTS

The consequences of this situation of robots and AI at the intersection of reason and emotion, rationality and irrationality are manifold, but perhaps nowhere are these fears about the vicissitudes of emotion and reason more manifest in twenty-first century robotics than in the case of social robots. While industrial robots and AI pose a much more immediate and realistic threat to the structural stability of human societies (e.g. [11]), the heaviest burden of popular fears of robots are carried by the idea of social robots. Social robots such as Paro and Zeno and Nao disarm and reassure members of the public (and in Sheffield we often bring these robots to meet people at public engagement events), but it is ultimately what these robots represent that strikes at the very heart of public fears about robots and AI more generally: that one day robots will come to be undistinguishable from human beings, and that they will come to replace, or even surpass, human beings completely.

The fear of social robots is also caused in part because of the inevitability of humans anthropomorphising (e.g. [10] [31]), which means that we will attribute more (general) intelligence and other human attributes, such as emotional responses and motivations, to robots that even slightly mimic human appearance or behaviours. This is fed in part, too, because the over-inflated belief that near-human robots are just around the corner (fed by popular science-fiction such as *Ex Machina* and *Humans*). But these causes are not independent of the historical basis for fears of robots and AI; science-fiction representations of advanced humanoid robots are merely giv-

² In fact, Kang demonstrates how in some writers, such as Diderot, these two conceptions of man and machine exist simultaneously.

ing voice to fears that already exist, and the tendency to anthropomorphise exacerbates our tendency to see ourselves in (in)animate machines. And as social robots are built specifically to exploit humans' tendency to anthropomorphise, it should come as no surprise that such robots magnify our pre-existing expectations and anxieties.

Social robots have tremendous potential in teaching children (e.g. [27] [6]) and care of the elderly (e.g. [8] [22]), but there is a great deal of controversy surrounding their usage that is symptomatic of both this debate and the anxieties they represent. I would like to present two cases to illustrate this: The 2010 EPSRC Principles of Robotics [2] and the EU Parliaments Legal Affairs Committee draft legislation from January 2017 [9].

In 2010 the EPSRC brought together experts 'from the worlds of technology, industry, the arts, law and social sciences' to devise a set of principles for responsible research into robotics. The shadow looming over this meeting, as already discussed, was that of GMOs [5]; the EPSRC was determined that robots and AI were vital to the future of the British economy, and were working to prevent a repeat of what happened with GMOs. Putting aside any judgements of the Principles on the whole (e.g. Prescott, in press), Principle 4 is intriguing:

Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead, their machine nature should be transparent. [2]

This legal phrasing is translated to a general audience in the following way:

Robots are manufactured artefacts: the illusion of emotions and intent should not be used to exploit vulnerable users. [2]

There is perhaps nothing controversial in the principle that robots should not be designed to deceive (despite the fact that a certain degree of deception has always been at the heart of robotics and AI, cf. [24] [10]), but the limitations on the 'illusion of emotions and intent' requires more consideration. The wording of the Principle for general audiences and the commentary make it clear that what is being talked about is social robots: 'One of the great promises of robotics is that robot toys may give pleasure, comfort and even a form of companionship to people who are not able to care for pets, whether due to rules of their homes, physical capacity, time or money' [2]. The commentary further makes it clear that this Principle is meant to protect 'consumers', to prevent unscrupulous manufacturers from eliciting more money from families who have become (emotionally) attached to such a 'toy' (and the EPSRC Principles seem to allow for social robots to be little more than mere 'toys'). That the category of 'vulnerable users' implicitly includes all consumers suggests that the Principle's authors do not see a terrific future for robots that present the 'illusion of emotions'. And while the Principle allows that 'it is permissible and even sometimes desirable for a robot to sometimes give the impression of real intelligence', there is a concern that the 'machine nature should be transparent', meaning that it should not be possible to confuse a robot with something that is not a robot, i.e. *something real*.

There are some that are very keen to exploit our human, irrational, emotional attachments and to use robots as a way of making connections with consumers for reasons that might be best described as less than entirely honourable; I accepted an invitation to speak last year at a marketing 'huddle' where there was a great deal of interest in using social robots and exploiting the 'science of cute' — i.e. the human capacity to anthropomorphise — as a tool for marketing

products to potential customers. However, there are myriad of potential benefits of using the human tendency to anthropomorphise that are legitimately being explored by researchers, e.g. robots as teaching assistants, in elderly care, as companions. It is difficult not to perceive that the EPSRC Principles are, in general, either denigrating or discouraging these possibilities.

It is helpful, too, to see how the question of social robots is treated in the EU Parliaments Legal Affairs Committee draft legislation on robotics and AI which were passed by an overwhelming majority of MEPs in January of 2017. The first paragraph of the report's Introduction begins:

A. whereas from Mary Shelley's *Frankenstein's Monster* to the classical myth of *Pygmalion*, through the story of Prague's *Golem* to the robot of Karel Čapek, who coined the word, people have fantasised about the possibility of building intelligent machines, more often than not androids with human features;

B. whereas now that humankind stands on the threshold of an era when ever more sophisticated robots, bots, androids and other manifestations of artificial intelligence ("AI") seem poised to unleash a new industrial revolution, which is likely to leave no stratum of society untouched, it is vitally important for the legislature to consider all its implications; [9]

This introduction is important, because it tells us a great deal about the origins of the assumptions underlying the report, what the report's authors see as the issues that need to be addressed, and its overall context. This report is, clearly, firmly rooted in the historical tradition of robots and AI that we can trace back to the Enlightenment: *Frankenstein* and Čapek are named checked, and the threat of a 'new industrial revolution' is the once and future context of robots in Europe.³ The report re-enforces many of the familiar concerns about robots, for example that 'the development of robotics and AI may result in a large part of the work now done by humans being taken over by robots', and some new concerns, that could not have been predicted by Shelley or Čapek, such as questions of data protection, but can be nevertheless be seen as part of the familiar contexts; for example, in the case of data protection, we can observe the usual fear that something uniquely human, i.e. the sense of privacy and a unique sphere of interiority, is being lost in the blurring of boundaries between the human and the technological.

On the question of social robots, however, the Legal Affairs Committee's report has very little to say. There is the acknowledgement that 'the "soft impacts" on human dignity may be difficult to estimate, but will still need to be considered if and when robots replace human care and companionship' [9]. Under the subheading of 'Care Robots', the report merely states: 'Points out that human contact is one of the fundamental aspects of human care; believes that replacing the human factor with robots could dehumanise caring practices'.

What this illustrates is a further, additional, articulation of our long-standing fears of robots. Clearly, here is the fear that robots will replace human labour, not just in the assembly lines but in the caring professions. There is the fear articulated that the increased contact with technology will dehumanise human beings, that we will somehow become less human through our contact with machines. Our dehumanisation reflects this anxiety that we are becoming the

³ The additional mention of the myths of *Pygmalion* and the *Golem* illustrate that this context is not limited to the history of Europe since the Enlightenment, but these are outside the scope of this paper.

machines that we fear. This transformation of the human is particularly poignant in the case cited by the EU report, as the jobs that are imagined to be filled by robots in this case are precisely those jobs that are normally assumed to involve high degrees of emotional ties between humans.

There are, of course, legitimate concerns that can and *should* legitimately be raised regarding the use of social robots, particularly as it relates to care. One may, of course, legitimately question whether the lack of human contact can have adverse impacts on our emotional and/or psychological well-being (e.g. [28] [25]). It is right not to conflate robots and AI that can *imitate* emotional responses with *genuine* emotional responses, especially as such distinctions are important to determining the sorts of rights and responsibilities we wish to bestow upon robots and AI [4]. It is also right to wish to protect consumers, though as long as robots and AI are nothing more than *products*, as they will be for the foreseeable future, the rights of consumers will be guaranteed with existing legislation and protections.

But it is impossible to separate these issues from their historical contexts. By anchoring these discussions on the myths of Frankenstein and Asimov, in speaking about *dehumanising* people and *humanising* robots (e.g. [4]), we are engaging in a long-standing debate, and these assessments of the role of robots and AI in our present and future human societies are dominated by the familiar fears, barely re-articulated, and the fundamental human questions that we have been grappling with since the Enlightenment, namely, what are the consequences of reason? How can we negotiate the vicissitudes of reason and emotion, and what path is in our best interests, as individuals, as a species?

In certain ways, therefore, perhaps too many, our debates and discussions about robots are just a symptom of a two-hundred-and-fifty year debate about what it means to be *human*.⁴ We have created robots in our imagination and cast them as characters in this drama. We use them as containers for our anxieties and aspirations. We doubt the authenticity of our relationships with robots (e.g. [28]) as if human emotions directed towards robots are less real, uniquely inauthentic in a way that our relations with other non-human creature or things are not. This is not to say that we can simply ignore these contexts, or proceed as if, having understood this, we can all simply correct these misunderstandings and move on. There are very real dangers in misrepresenting robots and AI, but likewise there are very real dangers in ignoring the complex human perceptions of robots, and ignoring the ways robots and AI represent our own anxieties about rationality and irrationality.

4 LESSONS FROM FREUD AND THE TRANSFERENCE

How then are we to move forward? We seem to be trapped in a paradox, or perhaps even more simply an inescapable contradiction, where we absolutely must talk about robots and AI in the context of this centuries-old attempt to locate *the human* on this map of reason and emotion, but at the same time we must try to discuss robots and AI in the context of their twenty-first century realities, understanding how they will come to live with us. We need to heed public concerns, but also construct a suitable ethics for the use of robots and AI that are not simply proxy debates about how we define ourselves. This can be done in part with a renewed commitment to education, information and engagement, but the public must be listened to as well as

lectured at.

For a strategy to help us approach this problem, I would turn to a thinker whose ideas were shaped exactly in response to these tensions between the rational and the irrational. Sigmund Freud's psychoanalytic practice had to confront precisely this very problem of how to come to a rational way of understanding irrational processes. Freud's challenge was to understand, explain and move beyond long-held beliefs and emotions, the chaos of social and personal histories that seemed intractable, and to which both patient and therapist were subject. Freud gave the name *transference* to this projection of past fantasies onto new situations, the messy battlefield where the patient's past conflicts re-engage. However, recognising that he himself was not free of the impingements of the past, he identified the therapist's *countertransference*, which Freud originally saw as the patient's influence on the analyst's unconscious but has come to refer more generally to the analyst's own irrational baggage in approaching the problems before him. [29] [30]

There are important lessons for all cultural studies, histories, philosophies and psychologies in Freud's recognition of the importance of the past in approaching present problems, and perhaps even more so in his self-reflexive realisation that no discourse, no explanation, is free of the influence of past struggles. And as the goal of psychoanalysis is to 'make conscious what is unconscious', or 'filling gaps in the memory' [12], our goal must be the same: drawing the hidden histories of imaginary robots and AI into the light, so that we can rationally examine the irrational ways we have built robots and AI in the cultural imagination, understanding the way they exist as symptoms of our anxieties about humans as caught between reason and emotion.

Freud's conception of the transferences underwent three distinct phases [18] [29] [30], all of which are relevant in different contexts to our understanding of robots and AI. Freud first conceived the transferences as errors that needed correcting or eliminating altogether, much as we might regard unfounded fears about robots as inaccuracies to be redressed or dismissed. However, Freud learned that it was much better to listen to the transferences, to understand them so as to exploit that knowledge when engaging with patients. Similarly, we need to take seriously the public's anxieties and aspirations about robots and AI. We should not regard the public's imagination as impediment to clarity and rational discussion; instead, we must seek to understand these transferences of imaginary robots onto those in the real world and learn from them, so as to engage more effectively with public perception and, ultimately, to build robots and AI that will be more readily accepted by all stakeholders (public, consumers, industries). Finally, Freud and post-Freudian psychoanalysts came to appreciate the transferences' vital importance in enabling meaningful and productive communication between patient and analyst. Likewise, we need to come to appreciate public perceptions of robots as not just things that need to be understood and grappled with, but as ideas that can be embraced, and with which we can engage.

5 CONCLUSION

Finally, there may also be lessons from the original master of science-fictional robots, Isaac Asimov, who not only did so much to invent robots in the public imagination but is still an important touchstone in our contemporary discussions. (As already highlighted, both the EPSRC Principles and a recent EU draft legislation begin by citing Asimov's famous 'Three Laws of Robotics'.) In his short story, 'The Evitable Conflict' [1], Asimov portrays a simmering, impending global conflict between Machines and an anti-Machine movement,

⁴ Bryson also acknowledges this fact, citing that at least part of the reason robots are 'overly personified' is because 'we are not certain what it means to be human' [4].

The Society for Humanity. In this world, the world Co-ordinator notices that the artificial intelligences that, for all intents and purposes, run the planet and human affairs seem to be acting irrationally. However, as robopsychologist Susan Calvin explains, it turns out that the Machine is not malfunctioning, but merely *pretending* to be acting irrationally in order to function more effectively, so as to appease humans for whom such a rationally perfect machine inspires not confidence but increased anxiety. In other words, a perfectly rational machine acts irrationally to achieve the rational aim of appeasing the irrationality of the human beings that it serves. More than merely providing humorous irony, perhaps this story can offer some lessons in how we should be looking to design robots and AI so that they can better fit in with our noblest human rational aspirations *and* our basest irrational impulses.

ACKNOWLEDGEMENTS

The author would like to thank the referees for their comments which helped improve this paper. This work was supported by European Union Seventh Framework Programme (FP7-ICT-2013-10) under grant agreement no. 611971.

REFERENCES

- [1] Isaac Asimov, 'The evitable conflict', *Astounding Science Fiction*, **45**(4), 48–68, (1950).
- [2] Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, et al., 'Principles of robotics', *The United Kingdoms Engineering and Physical Sciences Research Council (EPSRC). web publication*, (2011).
- [3] Josef Breuer, Sigmund Freud, and James Strachey, *Studies on hysteria*, Basic Books, 1955.
- [4] Joanna J Bryson, 'Robots should be slaves', *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, 63–74, (2010).
- [5] Joanna J Bryson, 'The Making of the EPSRC Principles of Robotics', **133**(133), 14–15, (2012).
- [6] David Cameron, Samuel Fernando, Abigail Millings, Michael Szollosy, Emily Collins, Roger Moore, Amanda Sharkey, and Tony Prescott, 'Designing robot personalities for human-robot symbiotic interaction in an educational context', in *Conference on Biomimetic and Biohybrid Systems*, pp. 413–417. Springer, (2016).
- [7] Karel Capek, *RUR (Rossum's universal robots)*, Penguin, 2004.
- [8] Paolo Dario, Paul FMJ Verschure, Tony Prescott, Gordon Cheng, Giulio Sandini, Roberto Cingolani, Rüdiger Dillmann, Dario Floreano, Christophe Leroux, Sheila MacNeil, et al., 'Robot companions for citizens', *Procedia Computer Science*, **7**, 47–51, (2011).
- [9] Mady Delvaux, 'Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))', Technical report, European Parliament, Brussels, (2017).
- [10] Brian R Duffy, 'Anthropomorphism and the social robot', *Robotics and autonomous systems*, **42**(3), 177–190, (2003).
- [11] Martin Ford, *The Rise of the Robots: technology and the threat of mass unemployment*, Oneworld Publications, 2015.
- [12] Sigmund Freud, *Introductory lectures on psychoanalysis*, volume 1, WW Norton & Company, 1977.
- [13] A Garland, 'Ex machina (film)', *Writer and director: Alex Garland*, (2015).
- [14] Edward S Herman and Noam Chomsky, *Manufacturing consent: The political economy of the mass media*, Random House, 2010.
- [15] Jana Horáková and Jozef Kelemen, 'Robots between fictions and facts', in *Intl Symposium on Computational Intelligence and Informatics*, pp. 21–39, (2006).
- [16] Minsoo Kang, *Sublime dreams of living machines*, Harvard University Press, 2011.
- [17] Fritz Lang, 'Metropolis [motion picture]', *Germany: Kino International*, (1927).
- [18] Jean Laplanche and Jean-Bertrand Pontalis, *The language of psychoanalysis*, Karnac Books, 1988.
- [19] Karl Marx, *Alienation and social classes*, na, 1978.
- [20] Shelley Mary, *Frankenstein: Or the modern prometheus*, 1985.
- [21] Ted Metzler and Lundy Lewis, 'Ethical views, religious views, and acceptance of robotic applications: A pilot study', *Association for the Advancement of Artificial Intelligence (www.aaai.org)*, **15**, (2008).
- [22] Elaine Mordoch, Angela Osterreicher, Lorna Guse, Kerstin Roger, and Genevieve Thompson, 'Use of social commitment robots in the care of elderly people with dementia: A literature review', *Maturitas*, **74**(1), 14–20, (2013).
- [23] N Sharkey and A Sharkey, 'Electro-mechanical robots before the computer', *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, **223**(1), 235–241, (2009).
- [24] Noel Sharkey and Amanda Sharkey, 'Artificial intelligence and natural magic', *Artificial Intelligence Review*, **25**(1-2), 9–19, (2006).
- [25] Noel Sharkey and Amanda Sharkey, 'The crying shame of robot nannies: an ethical appraisal', *Interaction Studies*, **11**(2), 161–190, (2010).
- [26] Joyce Tait, 'More faust than frankenstein: the european debate about the precautionary principle and risk regulation for genetically modified crops', *Journal of Risk Research*, **4**(2), 175–189, (2001).
- [27] Fumihide Tanaka and Shizuko Matsuzoe, 'Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning', *Journal of Human-Robot Interaction*, **1**(1), (2012).
- [28] Sherry Turkle, Will Taggart, Cory D Kidd, and Olivia Dasté, 'Relational artifacts with children and elders: the complexities of cybercompanion-ship', *Connection Science*, **18**(4), 347–361, (2006).
- [29] Robert M Young, *Mental space*, Process Press London, 1994.
- [30] Robert M Young, 'The vicissitudes of transference and countertransference: the work of harold searles', *Free Associations*, **5**(2), 171–195, (1995).
- [31] Jakub Zlotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck, 'Anthropomorphism: opportunities and challenges in human–robot interaction', *International journal of social robotics*, **7**(3), 347–360, (2015).