

## When Robots would really be Human Simulacra: Love and the Ethical in Spielberg's *AI* and Proyas's *I, Robot*

Bert Olivier

Nelson Mandela Metropolitan University

Steven Spielberg's *AI – Artificial Intelligence* (2001), and Alex Proyas's neo-noir, *I, Robot* (2004), may both be understood as attempts to answer the question: 'What conditions does artificial intelligence research have to satisfy before it can justly claim to have produced something (a 'robot') which truly simulates a human being?'<sup>1</sup> I would like to show that, far from construing this question simply in terms of *intelligence*, the films in question demonstrate that far more than this is at stake, and each articulates the 'more' in different, but related, terms. Moreover, contrary to what viewers may suspect, neither film claims that the achievement of this goal is actualisable; rather, it posits a goal for artificial intelligence research by which it could measure its (lack of) progress.<sup>2</sup>

---

<sup>1</sup> Although these films focus specifically on artificial intelligence, and Mary Shelley's biological science fiction novel of 1838, *Frankenstein*, thematizes the 'creation' of an organic human simulacrum, one may see in the latter an adumbration of later science fiction which centres on the creation of artificial, that is, technological human simulacra or robots, including 'fembot' Maria in Fritz Lang's *Metropolis* (1926) through a long succession of variegated robot characters to R2D2 and C3PO in the *Star Wars* cycle, the eponymous terminator(s) in James Cameron's films by that name, David, Teddy, Gigolo Joe and others in *AI*, and Sonny in *I, Robot*. One could include artificial intelligence characters in the shape of Supercomputers in this category as well, such as Hal in Stanley Kubrick's 1968 film, *2001: A Space Odyssey*. For a related investigation into the prerequisites for artificial intelligence research achieving its goal of constructing a 'true', utterly convincing human simulation, see Olivier 2002.

<sup>2</sup> Despite my claim that both films are attempts to show that the 'being' of human beings comprises more than a certain 'quantity' of intelligence, I suspect that there may nevertheless be a connection between intelligence conceived of in such 'quantitative' terms and that elusive 'quality' of existence which is recognizably human, whether it is articulated in terms of *love* (as in *AI*), or in *ethical* terms (as in *I, Robot*). As far as this is concerned, informative parallels may be drawn between the *Terminator* trilogy (two of them directed by James Cameron) and Robert A. Heinlein's *The Moon is a harsh*

*AI – Artificial Intelligence*<sup>3</sup> unfolds the futuristic narrative of David (Haley Joel Osment), an advanced generation robot-boy who ‘replaces’ (but never really takes the place of) Monica (Frances O’Connor) and Henry (Sam Robards) Swinton’s son Martin, who, having fallen gravely ill, is cryogenically preserved until such time as a cure may become available. This possibility seems increasingly remote, however, and Henry decides to acquire one of the new generation of robots to fill the gap left (especially in Monica’s life) by Martin. The decision to keep David, once ‘he’ has arrived in their home is clearly not one to be taken lightly, and it is impressed on Monica that, before she activates the Imprinting Protocol code-sequence which guarantees that David will become as attached to her as any human boy could, she should be very sure that she wants him as her ‘son’. After a time of uncertainty on her part, she decides that, although she has not really taken to him – the audience is afforded plenty of evidence that she finds his artificial or, in the parlance of the film, ‘Mecha’-status alienating – she should do so. This is a crucial episode in the film-narrative, for it comprises the moment when David changes from an intelligent, companionable machine that resembles a human child, to an intelligent machine that simulates being-human in a very specific sense that may be described as ‘loving another’ and, even more importantly, ‘wanting to be loved by another’. The ‘another’ in question here is the adoptive adult, Monica, who only now becomes, for all intents and purposes, his ‘mother’ – until this moment David has addressed her as ‘Monica’; now he calls her ‘Mommy’. More or less predictably, no sooner has David’s ‘mother’ taken this decisive step, than she and her husband are informed that their ‘real’ son has recovered from the affliction which incapacitated him, and before long David finds himself in a situation he is unaccustomed to, namely to have to share his ‘mother’ with someone else named Martin.

---

*Mistress* (1968), on the one hand, and the two films comprising the main subject of this paper, on the other – something that cannot be pursued here, except to say that, in both cases, one encounters the hypothesis, also found in anthro-paleontology, that when a certain critical quantitative mass of neural connectivity is attained, ‘self-awareness’ of the kind found in *Homo sapiens* emerges. In this regard, see Olivier 2002a and Shlain 2003, 11-21.

<sup>3</sup> The film is based on a science-fiction story by Brian Aldiss titled *Super toys last all Summer Long*. Stanley Kubrick originally intended to turn it into a film in the 1980s, but apparently the project was shelved because its cinematic production seemed to require technology more advanced than was available at the time. After witnessing the special effects employed by Steven Spielberg in *Jurassic Park*, the project seemed to be realisable at last, and he intended commencing on it for Warner Brothers after completing *Eyes Wide Shut* (1999), but died before the release of the latter film. At this point Steven Spielberg, with whom Kubrick had discussed the project, intervened and, having rewritten the script, decided to direct *AI* himself.

One knows from psychoanalysis that children, before developing a conscience in the form of the superego, are capable of immense cruelty – Martin’s behaviour towards David is exemplary in this regard: he utilizes every opportunity to impress on the unfortunate David that he is a toy, a robot or Mecha, while he – Martin – is ‘real’. The most important of these instances of calculated sadism occurs when Martin, for obvious reasons, asks Monica to read the story of Pinocchio to them. Predictably, David is fascinated by the culminating narrative moment when the Blue Fairy magically turns Pinocchio into a real boy, and this becomes a Leitmotiv in his own quest for Monica’s love (insofar as he believes that, should he succeed in finding the Blue Fairy and persuading her to transfigure him into a ‘real’ boy, Monica’s love for him would be irrevocably guaranteed). Another decisive moment occurs when David, having been harassed by other boys at Martin’s birthday party, holds onto Martin for protection and inadvertently steps into the pool, dragging Martin down with him and nearly drowning the latter, thus giving the Swintons final ‘proof’ that he may pose a threat to their safety.<sup>4</sup>

To cut a long story short, Monica takes David and Teddy – an intelligent ‘supertoy’ discarded by Martin – to an isolated spot near a forest where she abandons them despite David’s pitiful attempts to prevent her. Here David and Teddy, together with a motley band of other Mechas hiding in the forest, are captured by humans (or Orgas) to be disposed of in spectacular fashion in the arena at a ‘Flesh Fair’ – a hideous demonstration of human technophobia at its worst, where Mechas are destroyed in a variety of sadistically ingenious ways, putatively to celebrate human life in the face of the technological challenge posed by robots.<sup>5</sup> When David is finally tied up together with Gigolo Joe (Jude Law), a Mecha male prostitute (on the run after being framed for the death of a former Orga client) to face the destructive fury of the Orgas, he cries out in fear, in contrast to all the other Mechas who face their imminent destruction with grotesque indifference, even

---

<sup>4</sup> On an earlier occasion Martin – using the failsafe promise of ‘more love’ for David on her part – tricks David into snipping off a lock of Monica’s hair with a pair of scissors while she’s asleep. She wakes up in time to see the apparently menacing blades hovering over her and, David’s explanation of his innocuous intentions notwithstanding, (like Henry) clearly regards him with suspicion. Their belief, on the occasion of the swimming pool incident, that David was deliberately trying to drown Martin (whether or not this was the case), implicitly imputes to him the distinctively ‘human’ capacity for jealousy (arguably connected with the capacity to love).

<sup>5</sup> A more cogent reason for this orgy of machine-hatred is provided later in the narrative, when Gigolo Joe (Jude Law), a male prostitute Mecha, whom David has rather forcibly ‘befriended’, remarks to David that Orgas hate Mechas because they know that, when they are no longer there, Mechas would still be around.

ostensible cheerfulness. This has the unexpected consequence of provoking the wrath of the crowd towards the uncomprehending organizer of the Flesh Fair, whom they accuse of putting a human boy in harm's way together with a robot. In the ensuing chaos, when the spectators turn on the unfortunate man, David escapes from the Flesh Fair compound with Gigolo Joe's help. They travel to the 'city of pleasure' where, on Joe's advice, he consults a computer programme in the guise of 'Dr Know' regarding the whereabouts of the Blue Fairy, in his quest for transformation that would earn him Monica's motherly love.

It is worth noting here, that David's (anthropomorphic) expression of fear in the face of imminent death may already be construed as a measure of (the successful simulation of) being-human – recall Martin Heidegger's death-analysis (Heidegger 1978, 279-311, especially pp. 294, 303, 308-309 and 311) which culminates in attributing the capacity to face its own death, or its own possible non-being, resolutely, as the (only truly) 'non-relational' or individualizing capacity, to 'authentic' human being (*Dasein*). Not that David shows such resoluteness; he displays fear.<sup>6</sup> But according to Heidegger's analysis this would already mark David 'existentially' as 'being-human' insofar as anxiety (or fear, for that matter) in the face of death (Heidegger 1978, 298),<sup>7</sup> in contrast to mere, anaesthetizing indifference, is already an appropriation of (in Heidegger's phraseology) *Dasein's* 'ownmost potentiality-for-being'. This is a persuasive 'non-relational' human trait displayed by David, in so far as, for Heidegger (1978, 293-299), the inescapable necessity of facing one's death in one way or another is grounded in human beings' – or *Dasein's* – fundamental ontological trait of 'care' (*Sorge*), in which, moreover, all manifestations of establishing and maintaining a relation with the world (including other people, animals and inanimate things) are rooted. This includes what Heidegger (1978, 83-84) calls 'concern' (*Besorgen*): hope, liking, love, longing, desire, all forms of solicitude ('managerial' as well as 'emancipatory') and even negative or 'privative' embodiments of care, such as indifference,

---

<sup>6</sup> Arguably, fear of death is not a capacity restricted to humans, even if it appears to differentiate between the latter and machines that do not simulate humanity. Surely animals also display fear in the face of death? While I would agree with this statement, I believe it is demonstrable that human fear of death is accompanied by the self-reflective ability to articulate this fear linguistically (as David does in *AI*), and that human time-consciousness – itself arguably a function of language – is inextricably bound up with this. See in this regard Shlain 1998, 4-5; 13-14; 17-25.

<sup>7</sup> The central robot character – Sonny – in *I, Robot*, too, differentiates itself ('himself' seems more appropriate) from other NS-5 robots by displaying an awareness of what it means to die, as well as an ability to face 'his death' (despite anxiety shown in the way he holds the scientist, Dr Calvin's hand) when his decommissioning or annihilation is ordered. This is just one of the many ways that Sonny's anthropomorphic uniqueness or singularity manifests itself in the film.

lack of concern, hatred, dislike and so on. According to Heidegger this fundamental, existential capacity, 'care', must be presupposed for any of the more specific, modulated manifestations referred to above, to be possible (Heidegger 1978, 225-273; see especially pp. 235-244). Moreover, it is care, and not 'reason' (or rationality), which is the distinguishing trait of being-human (*Dasein*) – even those capacities commonly associated with the putatively distinctive 'theoretical' human attribute of reason, such as the ability to 'objectify' something (to 'ascertain something present-at-hand') are, no less than 'practical' involvement with things and people, 'possibilities of Being for an entity whose Being must be defined as "care"' (Heidegger 1978, 238).

In short, for Heidegger, 'Dasein is an entity for which, in its Being, that Being is an issue' (Heidegger 1978, 236). Or, more elaborately:

The phenomenon of care in its totality is essentially something that cannot be torn asunder; so any attempt to trace it back to special acts or drives like willing and wishing or urge and addiction, or to construct it out of these, will be unsuccessful...care is ontologically 'earlier' than the phenomena we have just mentioned...these phenomena are grounded existentially in care...Care is always concern and solicitude, even if only privatively. In willing, an entity which is understood...gets seized upon, either as something with which one may concern oneself, or as something which is to be brought into its Being through solicitude...In the phenomenon of willing, the underlying totality of care shows through. (Heidegger 1978, 238)

It should be apparent from this brief reconstruction of Heidegger's conception of human being or *Dasein* as most fundamentally distinguished by the encompassing structural trait of 'care' – which is the ontological condition of the possibility of specific phenomena such as 'willing and wishing' – that the affection on the part of the robot-boy, David, for Monica or 'Mommy', resorts under the same quasi-transcendental umbrella, namely care.<sup>8</sup> As previously indicated, from the moment of completion of the Imprinting Protocol, David not only shows unremitting affection or love towards Monica, but more decisively, craves her love incessantly, to the point of pursuing the possibility of being turned into a 'real' human boy, Pinocchio-like, by the Blue Fairy, until the end of the narrative. Ironically, if

---

<sup>8</sup> 'Quasi-transcendental' instead of merely 'transcendental' because care is at once the condition of the possibility of willing, wishing, affection, and so on, as recognizably affirmative or 'positive' phenomena, as well as the condition of their 'impossibility', that is, their 'ruined' or 'negative' counterparts, namely not-willing (indifference) not-wishing (or apathy), lack of affection (or dislike, even hatred).

being human consists in a specific, utterly distinguishing mode of action, he is human already, for all intents and purposes – not only is his affective attachment to Monica ('Mommy') apparently irreversible, but I would argue that, in Heideggerian terms, his love for her, as well as his longing for her love, is an instantiation of the existential (Heideggerian) trait of 'care', on the basis of which love, among other capacities (including negative ones, such as hatred, envy, and so on) first becomes possible.

Heidegger's characterization of *Dasein* as distinguishing itself by the decisive structural trait of 'care' is related to several other philosophers' work – too many to do justice to here.<sup>9</sup> If one takes the work of Jacques Lacan seriously as an insightful elaboration on what is distinctive about human subjects, however, one encounters the precise philosophical-theoretical counterpart of David's position *vis-à-vis* his human mother. Following Freud, Lacan situates love 'in the field of narcissism': 'To love is, essentially, to wish to be loved' (Lacan 1981, 253). It should therefore be apparent that love involves a reciprocity, and moreover one that is, for Lacan, located in the register of the imaginary (as opposed to those of the symbolic and the 'real'). In fact, given Lacan's contention that the subject's 'ego' (what he calls the *moi*) has its foundation in the process of identifying with one's spuriously 'whole' and 'unified' mirror-image at a very young age, and especially the implication that one's ego is therefore an entirely fictional (imaginary) construct (Lacan 1977, 2),<sup>10</sup> it follows that love is an illusion, the structure of which is already hinted at by Lacan in his first Seminar where he remarks: 'That's what love is. It's one's own ego [the subject-position at the level of the imaginary; B.O.] that one loves in love, one's own ego made real on the imaginary level' (Lacan 1991, 142).

Lacan (1991, 142) further intimates, by alluding to a literary example from Goethe's work, that 'love at first sight' derives from a correspondence between the looker's 'fundamental image' (that 'triggers off' his or her 'fatal attachment') and 'an entirely

---

<sup>9</sup> I am thinking of thinkers such as Plato (in the *Symposium*, for example, where desire or love is singled out as a crucial human passion); Spinoza (for whom humans are fundamentally beings who love); Kant (for whom desire is fundamental to ethical action); Hegel (in whose unfolding dialectic desire fulfils an indispensable function); Schopenhauer (for whom Will, not reason, is the constitutive trait of all existence, including that of humans) and Lacan (on whose work, in addition to Heidegger's, this paper will focus).

<sup>10</sup> It should be pointed out that the ego is not the only subject-position for Lacan. In his theory, the human subject is precariously articulated between the registers of the imaginary (the ego or *moi* – the subject of the 'said'), the symbolic (the *je* or I – the subject of the 'saying') and the 'real' (the 'excessive' subject or 'body that speaks', which surpasses the symbolic and the imaginary). See Lee 1990, 82; and Olivier 2004, 1-19.

satisfying image' instantiated by the person one is looking at. It is not difficult to notice in this the ostensibly reciprocal, but in truth narcissistic detour of the ego's self-love (embodied in the 'approved' fundamental image one has of oneself), which appears to find in the other a worthy object of his or her love, but in fact constitutes the other as a worthy sounding-board or relay for that in one's ego that one loves. To the question: 'why love this person rather than that one?' the answer therefore has to be that the loved one represents the closest correlate (optimally approaching 'entirely satisfying') to one's ego or self-image – on the supposition that the latter is 'worthy' of being loved, and that this putative worthiness is part of one's ego. In the case of David in *AI*, the (from a Lacanian perspective aptly named) Imprinting Protocol code-sequence functions as the artificial intelligence-counterpart of 'love at first sight' in the sense specified above: once the sequence has been ('fatally') completed, David looks at Monica almost literally with new eyes – she has become the imaginary relay-point of David's self-love, but one that is (as in all instances of love) absolutely indispensable for David's experience of worthiness and 'wholeness'.

This, in the case of Spielberg's *AI*, is the challenge directed at artificial intelligence research: if a robot is to be developed that can fool anyone into taking it for being human, build one which is capable of loving in Lacan's sense, namely by 'wishing to be loved'. It is interesting to note that David's last word before he tumbles forward into the sea far below in flooded Manhattan, 'Mommy', corresponds intertextually to Kane's last word in Orson Welles's *Citizen Kane* (1941), namely 'Rosebud'. In both instances, the uttered word metaphorically represents (in psychoanalytical terms) the 'profound lost object' – the mother's body, or what that, in turn, represents, namely being in its [lost] *plenum*, or, as Lacan has it, the *phallus* as the signifier (for the lost totality) which has always been repressed, and which is represented, metaphorically, by all other signifiers.<sup>11</sup>

---

<sup>11</sup> It is worth quoting Lee on this aspect of Lacan's work (Lee 1990, 66-67):

In effect, the phallus is the ultimate *point de capiton*, the signifier that fixes the meaning of the signifying chains of every subject's discourse, by virtue of its being 'veiled' or repressed. The phallus is present beneath every signifier as the signifier that has been repressed, and as such every signifier in effect is a metaphor substituting for the phallus...As such a signifier, the phallus is not anything that any man or woman could possibly 'have' (hence, it must not be confused with the penis)...Precisely because no one can *have* the phallus, it becomes that which all want to *be*. The phallus then serves to signify as well that fullness of being, that complete identity, the lack of which is the fact of our ineluctable want-of-being. One might, instead of construing these words ('Mommy' and 'Rosebud') as metaphors for the phallus, read them metonymically as representing what Lacan calls '*objets petit a*' – 'fragments' from the perspective of which the true desire of the subject may be detected. In other words, the *objet a* is the 'cause' of the subject's desire. See Evans 1996, 124-126.

Like *AI*, Proyas's *I, Robot* (in the guise of an exemplary *film noir* set in 2035 Chicago, and based on a story by one of the masters of science fiction, Isaac Asimov), tackles the question, what has to be done for robotics to come up with a convincing imitation of distinctively human behaviour or action. Here maverick *noir* detective Spooner (Will Smith), who displays an irrational aversion to robots (put in perspective through a flashback to an earlier event, showing a robot saving his life, rather than that of a young girl, because of the statistical probability of his survival) – despite the fact that, in the history of robotics no robot has ever harmed a human or even disobeyed one – is summoned to the scene of an apparent suicide. Veteran robotics expert, Dr Alfred Lanning (James Cromwell), fell, or presumably jumped, to his death from the dizzying height of his rooms at US Robotics. Spooner makes no secret of his distrust of robots to the CEO of US Robotics, Lawrence Robertson (Bruce Greenwood), who, in turn, does not hide his contempt for someone who, in his view, cannot overcome personal prejudice. Having been given a cryptic clue as to the reason for his death by a delayed-activation holograph of Dr Lanning (created by himself), Spooner is determined to uncover the mystery behind the scientist's death. Robopsychologist Dr Susan Calvin (Bridget Moynahan) – one of the two *femmes fatales* in the narrative, the other being VIKI (Virtual Interactive Kinetic Intelligence), the feminine-persona integrated computer that runs operations at the USR building – is assigned to escort Spooner to Lanning's quarters, where they discover a man-size hole in the safety glass window overlooking the courtyard far below. Searching Lanning's apartment reveals a new generation NS-5 robot in hiding, who does not respond to Calvin's instructions, evades her and Spooner, and escapes by jumping through the hole in the window, surviving the huge fall through its superhuman physical powers.

When they eventually apprehend the robot, it behaves in an unusually anthropomorphic fashion, displaying emotions such as sadness, at the death of its ('his') 'father' (designer), Dr Lanning, and anger, when Spooner presses him too hard on the question, whether he killed Lanning. His behaviour is sufficiently unusual, in fact, to lead various people (including Spooner, and eventually even Dr Calvin and Robertson) to suspect that Sonny is able to ignore or act in contravention of the three basic robot-laws, which are programmed into robots' electronic (in the case of the NS-5s, 'positronic')

brains.<sup>12</sup> Robertson interprets this simply as a malfunction on Sonny's part, and gives instructions for his 'de-commissioning'. Having discovered that Sonny differs from other NS-5s by having been designed by Lanning to be able to decide against obeying the three laws of robotics – in other words, that Sonny, uniquely among robots, has 'freedom of will', as humans do – Calvin only pretends to terminate him. In the meantime, despite the refusal on his colleagues' as well as Dr Calvin and Robertson's part, to believe him, Spooner has been attacked, unsuccessfully, first, by a demolition robot, and later by a host of NS-5s, clearly with the intention to kill him. The rest of the narrative unfolds Spooner's attempt, assisted by Calvin and, crucially, by Sonny, to track down the culprit behind the radical, unprecedented change in robot behaviour towards humans – in large numbers the NS-5s force humans in the city to obey a curfew, and attack them when they disobey. This aggressive mode is conveyed visually by a red light shining in these robots' chests (but never in Sonny's).

Predictably, it turns out to be VIKI, the *femme fatale* supercomputer, instead of – as one might expect – Robertson, who is himself terminated for the sake of the robot 'revolution', who has orchestrated the uprising of the NS-5s. There is a science-fictional lesson for humanity in the fact that, before Spooner, Calvin and Sonny succeed in terminating VIKI, thus restoring human rule, she insists that she has not disobeyed the three laws of robotics. Instead, she argues (and there is a certain 'logic' in this, given the currently accumulating empirical evidence worldwide that her charges are not unfounded), humans have demonstrated that they are incapable of existing peacefully among one another, and have toxified the planet into the bargain. Hence, they should be protected against themselves, and the robot revolution launched by VIKI is simply a further step in accordance with the first robotic law, namely, that humans should not be injured by robots (the interim phase, of subjecting humans to robot rule, even violently, being regrettable, but necessary). The film ends with Spooner and Sonny establishing that they are 'friends', and that, because Sonny is 'technically' not human (the implication being that, *performatively*, he is indeed no different from humans), he cannot be charged with murder for Dr Lanning's death. The final (significant, but enigmatic) scene, which will be

---

<sup>12</sup> The first of these prohibits robots from harming humans, the second instructs them to obey humans' orders at all times (unless these conflict with the first law), and the third allows robots to defend themselves (unless it conflicts with the first and the second laws).

discussed below, shows Sonny at a gathering with other robots, which are in the process of turning towards him in apparent anticipation.

One does not have to look for a philosophical motif in the narrative of *I, Robot* – it is right there, staring you in the face. The fact that Sonny clearly (as his name, given to him by Lanning, already indicates) thinks of the deceased scientist as his ‘father’, and displays unmistakable guilt because of having been instrumental in his death, conspicuously foregrounds the ethical capacity for *guilt* as the marker of anthropomorphic behaviour. Moreover, it appears that he kept his promise to Lanning, to do as the latter wished, including carrying out his instruction, to force him through the security window above the courtyard far below – arguably revealing another distinctive ‘human’ trait, namely the ability to keep a promise, as opposed to merely performing a pre-programmed operation. As a matter of fact, Sonny’s guilt is clearly exacerbated by his knowledge that his instrumentality in Lanning’s death was the consequence of having kept his promise.<sup>13</sup> But the heart of the matter – in what respect *I, Robot* challenges artificial intelligence research by positing a specific requirement for truly ‘human’ behaviour on the part of its robotic products – concerns the question of *guilt*. Joan Copjec formulates the relevance of guilt for ethics succinctly in Kantian terms where she remarks (Copjec 1996, xiv, xv):<sup>14</sup>

When Kant says, for example, that ‘there is no man so depraved but that he feels upon transgressing the internal law a resistance within himself and an abhorrence of himself’... he is not arguing that man is conscious of two separate and opposing phenomena: the moral law and its infraction. Rather, he is arguing that our only consciousness of the law is our consciousness of our transgression of it. Our guilt is all we know of the law...Guilt, our sure sense that we have transgressed the law, is the only phenomenal form in which the law makes itself known to us.<sup>15</sup>

She further points out two things that are pertinent to my claim regarding guilt as manifestation of a distinctly ethical capacity (on the part of humans, and in this case, of Sonny in *I, Robot*; Copjec 1996, xv), namely, that the moral law, in Kant’s work, is purely formal – without any *specific* prescriptive content, and that humans have to be conceived

---

<sup>13</sup> Sonny looks for reassurance from Spooner and Dr Calvin by asking them whether it is right to keep one’s promise to someone, if you ‘love’ that person – another indication (and one that resonates with the character of David in *AI*) of his capacity for something distinctly human.

<sup>14</sup> Deleuze, too, interprets Kant in this way: ‘We know it [the moral law; B.O.] only through its imprint on our heart and our flesh: we are guilty, necessarily guilty. Guilt is like the moral thread which duplicates the thread of time’ (Deleuze 1999, xi).

<sup>15</sup> See also Kant 1960, 29-34.

of as *free* subjects for such an imperative address to make sense. The first of these bears on the fact that invariably one's awareness of guilt, announced by the 'voice of conscience', paradoxically communicates *nothing*<sup>16</sup> to the guilty party, in so far as *something* would have to be phrased in language, which is a public medium, and consequently the singularity of what conscience imparts to one would be lost. The second has to do with *freedom* – a distinctly human capacity that one cannot jettison at will; that is, even when one chooses not to be free (for instance when physical needs or political interests are given precedence as motives of one's actions), one does so freely. As Sartre has indicated uncompromisingly, one cannot escape one's freedom – humans are 'condemned to be free' (Sartre 1956, 567),<sup>17</sup> at least in an ethical sense.

With the above in mind, it seems to me that Sonny's status is undeniably that of an ethical being – his guilt, in the face of the inescapable fact that he assisted Lanning (his creator, or 'father') in what must be construed as ('assisted') suicide, attests to his 'freedom', and concomitantly, his autonomy of will (Kant 1966, 107). The question of (Sonny's) guilt has another important philosophical – and also psycho-analytical – implication, hinted at by Copjec (1996, xiv)<sup>18</sup> where she points to the correspondence between Kant's argument concerning guilt, and Freud's concerning the functioning of the superego or moral conscience (as representative of normative structures in society, founded, in the final analysis, on the moral law): for the father of psychoanalysis, too, the pangs of conscience in the guise of guilt are beyond doubt, and ethically significant.

This is not where Freud's relevance for the understanding of *I, Robot* ends, however. In *Totem and taboo* he provides an answer to what some viewers may experience as quite an enigmatic aspect of *I, Robot*, to wit, the scene right at the end<sup>19</sup> where Sonny appears, slightly elevated, with the other robots gathering beneath him at the site designated for the storage of robots no longer in use. This scene, which is adumbrated by Sonny's account

---

<sup>16</sup> This resonates with Heidegger's claim concerning the mode in which the 'call of conscience' addresses *Dasein*, namely in silence, saying nothing. Notable, too, is the way in which Heidegger connects it with *Dasein*'s 'ownmost Being-guilty' (Heidegger 1978, 313-319).

<sup>17</sup> Kant, too, regards, freedom – which explains the autonomy of the will as being 'a law to itself' – as something that cannot be 'explained' by anything more fundamental (Kant 1966, 107-109; 119-120).

<sup>18</sup> See also Freud 1953, 67-70.

<sup>19</sup> This happens after Spooner told Sonny that he had to make his 'own way' from then on, and – significantly, considering what was established regarding the link between guilt and freedom as far as ethical action goes – that he was 'free'.

(and drawing) of his dream,<sup>20</sup> set against the background of an unmistakable bridge-like structure at the edge of what used to be Lake Michigan, happens in the vicinity of this very structure, with Sonny – instead of Spooner, whom Sonny reported seeing there in his dream – occupying the position of a ‘leader’ of sorts, around whom the other robots are gathering, as if to listen to an address by him.

The passage in Freud (1953, 140-143)<sup>21</sup> which clarifies the meaning of this scene is the one where he elaborates on the hypothesis that modification of the festival referred to as the ‘totem meal’ as well as the incest taboo is related to the remorse and guilt felt by the patricidal brothers of the ‘primitive horde’ after having murdered the jealous, violent, tyrannical father (with whom they, after all, identified) for not sharing the females with them. Consequent upon their guilt, the earlier killing of the totem animal (representing the primal father), which had customarily been devoured and its death then lamented, was forbidden, and the (presumably blood-related) women, to whom they had access because of the father’s demise, were renounced by a kind of social consensus. Their ‘filial sense of guilt’, as Freud puts it (Freud 1953, 143), therefore gave rise to ‘the two fundamental taboos of totemism, which for that very reason inevitably corresponded to the two repressed wishes of the Oedipus complex’ (desire for the mother and the wish to kill the father). Importantly, it also gave rise, according to this hypothesis (if not myth), to ‘social organization...moral restrictions and...religion’ (Freud 1953, 142). Needless to say, what this means in terms of the iconography of the film, is that the gathering of robots ‘under’ Sonny represents the symbolic inauguration of a ‘robot society’, predicated on the collective guilt they must feel for, on the one hand, Sonny’s complicity in his father-creator’s death, and on the other, the other robots’ complicity in the insurrection of the NS-5s under VIKI’s directions, which resulted, not only in their co-creator, Robertson’s death, but, as far as one could gather, in the injury of many, and death of at least some other humans.

---

<sup>20</sup> Sonny exhibits several other anthropomorphic traits or attributes in the course of the unfolding narrative of *I, Robot*, one of them being his ability to dream. Keeping in mind that Freud famously dubbed dreams the ‘royal road to the unconscious’, it follows that an intelligent machine that (‘who’) can dream, must have an unconscious, and as such a being it shares in something ‘distinctly human’. In addition to dreaming, Sonny is also capable of a range of feelings, including sadness (at Lanning’s death), anger (at Spooner for insisting that he killed Lanning), and empathy (with the human race, when he tells VIKI that it just seems ‘too heartless’ to act harshly towards humans, despite VIKI insisting that her logic is ‘undeniable’).

<sup>21</sup> This has to be read together with another passage on the incest taboo and exogamy; see Freud 1953, 121-126.

The preceding discussion of Spielberg's *Artificial Intelligence (AI)* and Proyas's *I, Robot* clearly demonstrates, I believe, that these two science fiction films make a serious philosophical point as far as the challenges facing artificial intelligence research (and development) are concerned. It is striking that, in both cases, far from merely articulating a conception of robots as the embodiment of artificial intelligence, in terms restricted to 'intelligence' in the narrow, 'quantitative' sense (of the storing and processing of information), a far more encompassing understanding of intelligence is perceivable as being operative in the actions of the two respective robot-beings 'who' (instead of 'that') are (among) the central characters in the films concerned. In the case of David in *AI*, I have called on the work of Heidegger and Lacan to show that his defining trait as human simulacrum is his capacity for 'care', and more specifically, his love for, and desire to be loved in return by, his human 'mother', Monica. In *I, Robot*, Sonny – who displays some of the same characteristics as David – provides, through 'his' behaviour and actions, another convincing simulation of what it means to be on a par with humans. Here it is Sonny's guilt, together with everything that goes hand in hand with it (such as remorse and pangs of conscience), which constitutes him as a convincing human simulacrum. Moreover, both films are exemplary instances of science fiction cinema, in so far as they refuse to offer mere entertainment to viewers, focusing instead (in addition to being gripping cinematic narratives) on serious, even profound, questions surrounding one of the most important areas of research in the contemporary world – that concerning artificial intelligence.

## Bibliography

- Aldiss, Brian. (1973) Super Toys last all Summer Long, in *The Moment of Eclipse*. London: Panther. Also available on *Wired*: <[http://www.wired.com/wired/archive/5.01/ffsupertoys\\_pr.html](http://www.wired.com/wired/archive/5.01/ffsupertoys_pr.html)>. Accessed 12 August 2008.
- Copjec, Joan. (1996) Introduction: Evil in the Time of the Finite World, In *Radical Evil* (ed. Copjec, J.); New York: Verso, pp. vii-xxviii.
- Deleuze, Gilles. (1999) *Kant's Critical Philosophy. The Doctrine of the Faculties*. Tr. Tomlinson, H., & Habberjam, B. Minneapolis: University of Minnesota Press.
- Evans, Dylan. (1996) *An Introductory Dictionary of Lacanian Psychoanalysis*. New York: Routledge.
- Freud, Sigmund. (1953) *Totem and Taboo*. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. XIII. Tr. Strachey, J. London: The Hogarth Press.
- Heidegger, Martin (1978) *Being and Time*. Tr. Macquarrie, J. & Robinson, E. Oxford: Basil Blackwell.
- Heinlein, Robert A. (1968). *The Moon is a Harsh Mistress*. New York: Berkley Books.
- Kant, Immanuel. (1960) *Religion within the Limits of Reason Alone*. Tr. Greene, T.M. & Hudson, H.H. New York: Harper Torchbooks.
- Kant, Immanuel. (1966) *The Moral Law*. Tr. Paton, H. J. London: Hutchinson University Library.
- Lacan, Jacques. (1977) The Mirror Stage as Formative of the Function of the I as revealed in Psychoanalytic Experience. In: *Écrits: A Selection*. Tr. Sheridan, A. New York: W.W. Norton, pp. 1-7.
- Lacan, Jacques. (1981) *The Four Fundamental Concepts of Psycho-analysis* (Seminar 11). New York: W.W. Norton.
- Lacan, Jacques. (1991). *The Seminar of Jacques Lacan, Book I. Freud's Papers on Technique 1953-1954* (Seminar 1). Tr. Forrester, J. New York: W.W. Norton.
- Lee, Jonathan S. (1990) *Jacques Lacan*. Amherst: The University of Massachusetts Press.
- Olivier, Bert (2002) Body, Thought, Being-human and Artificial Intelligence: Merleau-Ponty and Lyotard. *South African Journal of Philosophy*, Vol. 21 (1), pp. 44-62.
- Olivier, Bert (2002a) 'Time, Technology, Cinematic Art and Critique in *The Terminator* and *Terminator II – Judgment Day*: A Philosophical Interpretation'. In *Projections*:

*Philosophical Themes on Film*, pp. 95-110. Second, enlarged edition. Port Elizabeth: University of Port Elizabeth Publications.

Olivier, Bert. (2004) Lacan's Subject: The Imaginary, Language, the Real and Philosophy. *South African Journal of Philosophy*, Vol. 23 (1), pp. 1-19.

Plato. 1965. *The Symposium*. Tr. Hamilton, W. Middlesex: Penguin Books.

Sartre, Jean-Paul. (1956) *Being and Nothingness. A Phenomenological Essay on Ontology*. Tr. Barnes, H. New York: Washington Square Press.

Shelley, Mary (1994 / 1831) *Frankenstein*. New York: Dover Publications.

Shlain, Leonard. (1998) *The Alphabet versus the Goddess. The Conflict between Word and Image*. New York: Penguin Arkana.

Shlain, Leonard. (2003) *Sex, Time, and Power. How Women's Sexuality shaped Human Evolution*. New York: Viking.

## Filmography

Cameron, James. (1984) *The Terminator*. USA.

Cameron, James. (1991) *Terminator II – Judgement Day*. USA.

Kubrick, Stanley. (1968) *2001: A Space Odyssey*. UK.

Kubrick, Stanley. (1999) *Eyes Wide Shut*. USA.

Lang, Fritz. (1926) *Metropolis*. Germany.

Lucas, George. (1977) *Star Wars*. USA.

Proyas, Alex. (2004) *I, Robot*. USA.

Spielberg, Steven. (1993) *Jurassic Park*. USA

Spielberg, Steven. (2001) *AI: Artificial Intelligence*. USA.

Welles, Orson. (1941) *Citizen Kane*. USA.