Articoli/13

How to Lose to a Chess Playing Computer According to Jean Baudrillard

Gary Genosko

Readers of Baudrillard know that he thought about competition in sport and games in terms of failure and frailty. In *For a Critique of the Political Economy of the Sign*, exchange value and symbolic ambivalence are mutually exclusive domains; in the latter, desire is not satisfied through phantasmic completion, and this entails that desire may ride failure to an ignominious counter-victory. Baudrillard found in the failure to react positively to an inducement like winning a race – captured in that bizarre American football phrase appropriated as a handle by Ronald Reagan, «Win One for the Gipper!» – the principle of a radical counter-economy of needs. Losers come in all shades. But radical losers stand apart from the crowd in the virulence of their capacity to radiate loss that they throw down as a challenge. There are those who are irresistibly drawn to blowing it, and others who can taste failure and steal it from the jaws of victory. From the Beatles to Beck, the figure of the loser has fascinated lyricists and theorists alike as not merely sympathetic but as a foundation for a deliberate weakness in the face of overwhelming odds and the false pretenses of victory. Here I revisit Jean Baudrillard's speculations about computer chess programs, specifically IBM's Deep and Deeper Blue, and how best to play against them. Drawing on Baudrillard's theory of loss in sports as an act of contempt for the fruits of victory, institutional accommodation, and the cheap inducements of prestige and glory, I examine how chess masters like Garry Kasparov have met the challenge of the brute force programs – some of which were congealed models of his own play – with appeals to a kind of unforced play and even 'non-thought'. Considering the malevolent and fictional computer system HAL, as well as Deep Blue and subsequent programs, right up to IBM's Jeopardy-playing computer 'Watson', this paper looks at ways to defeat programming power by critically regaining the counter-technical and (dys)functional skills of the loser.

***

«The film *The Loneliness of the Long-Distance Runner* is a very fine example, from a social and political point of view, of this ubiquitous counter-economy. The hero is an adolescent in a rehabilitation center who deliberately renounces a decisive victory in a running contest in order to avoid spreading any of the glory to his institutional oppressors. By losing, he preserves his own truth: here, failure merges with class revolt. Admittedly, in this story, the failure is explicitly deliberate, but it is not difficult to see how 'accidental' lapses and physical slips may acquire virtually the same meaning of denial and resistance»

Jean Baudrillard, *For a Critique of the Political Economy of the Sign*

1. AI and Chess

The transition from the symbolic necessity of either throwing or blowing it to the charm of weak play is evident in Baudrillard's discussion of IBM's chess computer Deep Blue and its first six-game match against world champion Garry Kasparov in 1996. I want to extend the lessons of Baudrillard's scattered, but overall coherent, remarks on competition to the context of chess played between a world champion and a specialized machine. I do this against the background of earlier celluloid dreams about artificial intelligence manifested in the malevolent supercomputer HAL in Stanley Kubrick's film *2001: A Space Odyssey* (1968).

AI programs, especially chess programs like Deep Blue which defeated grandmaster Kasparovin the first game, are defined by Daniel C. Dennett[1] as «intentional»:

its behavior is predictable and explainable if we attribute to it belief and desires – 'cognitive states' and 'motivational states' – and the rationality required to figure out what it ought to do in the light of those beliefs and desires.

Deep Blue doesn't cheat; it doesn't perform misdeeds, even though Kasparov did claim its programmers cheated in his second losing series against the upgraded Deeper Blue by playing the cabinet trick, in which experts are hidden from view 'in' the machine and guide the machine manually. But does it earn praise? Yes, thinks Dennett, but Kasparov, after all, won the first series, and his horizons are a lot wider than those narrowly circumscribed in Deep Blue's world. The computer's «perceptual input» would need to be increased, and its coping mechanisms enhanced, for it to move beyond the parameters of the chess game. Deep Blue does not have «high order intentionality»; that is, it has no «moral responsibility» for its actions[2]. It lacks the ability to be self-critical. But the very framing of this discussion, according to Baudrillard, is a symbolic stake, a challenge of sorts that pits human against machine in an interface that gives way to a more basic rivalry: «The game of chess merely carries this situation to extremes»[3]. And Kasparov's first-round victory simply defers the machine's inevitable victory in the rematch the following year and announces the apparent nightmare of the loss of mastery. Yet the matter is not so simple, as any machine's victory is «a mark of our power, though we cannot bear this either»[4]. Baudrillard suggests that, after all, the most ignominious defeat is at the hands of a machine of our own design.

---

[1] D. C. Dennett, *When HAL Kills, Who's to Blame? Computer Ethics*, in *HAL's Legacy*, David G. Stork (ed, Cambridge 1997, pp. 351-65, p. 352.

[2] *Ibid.*, p. 354.

[3] J. Baudrillard, *Deep Blue or the Computer's Melancholia*, in *Screened Out*, C. Turner (trans.), London 2002, pp. 160-65: 160.

[4] *Ibid.*, p. 161.

Kasparov is truly multilingual, claims Baudrillard, in more than a strict linguistic sense, whereas Deep Blue speaks only «the language of calculation»[5]. If humankind is reducible to only compuspeak, then «defeat» by the machines is imminent. For what is lost is that Kasparov outstrips his own thought processes, whereas Deep Blue is trapped in its calculations, «within the confines of its own programme»[6], outside of which there is nothing or, at best, augmentations of the closed circuit.

The relinquishment of thought and the introduction of weak play, which is to say, playing below one's full potential, may be unknown to Deep Blue, which apparently must «play at the height of its capabilities»[7] at all times. It is «condemned» to this, thinks Baudrillard, underlining the implications of brute-force computing. The properly human response is to become «technically incorrect»[8]. Creativity is necessary; but so too is weakness. Unpredictability is a virtue, as is unruliness, and this species of innovation, of tactical underperformance, sits at the heart of gaming. Yet this is precisely what Kasparvo feared: that the «cabinet trick» made it look like Deeper Blue played below its capacity, eschewing «raw hardware speed» and computational power[9] and made decisions that deferred its victory.

HAL seems, at least initially in its fictional filmic matrix, to be a higher-order intentional system that might get the blues; HAL may even crackup. Indeed, he expresses puzzlement, comments on the oddness of the situation, and displays mild obsession. HAL also appears to be capable of misdeeds and duplicity, and questionable enthusiasm[10]. Does this mean that HAL is irreducible to algo[rithmic]-chess?

The narrative that gives HAL a history, especially a childhood to which he «devolves», belongs to the world of the script written jointly by Arthur C. Clarke and Kubrick, originally titled *Journey Beyond the Stars*. The dimensions of history and experience make HAL a moral agent of sorts to the extent of «show[ing] signs of fitting into one or more of the exculpatory categories»[11]– perhaps insanity. In an early version of the script, HAL (whose original name was Athena) «has a schizophrenic breakdown fragmenting into two personalities, with one explaining and condemning the actions of the other»[12]. HAL does express fear, but whether this is more than an emotional response displayed for pragmatic ends is anyone's guess. HAL has, for Dennett, «something very much like emotions»[13]. Deep Blue doesn't need «emotional crutches»[14] like fear. Yet

---

[5] *Ibid.*, p. 161.
[6] *Ibid.*, p. 162.
[7] *Ibid.*, p. 163.
[8] *Ibid.*, p. 163.
[9] H. R. Ekbia, *Artificial Dreams: The Quest for Non-Biological Intelligence*, Cambridge 2008, p. 49.
[10] D. C. Dennet, *When HAL Kills, Who's to Blame? Computer Ethics*, cit., p. 355.
[11] *Ibid.*, p. 360.
[12] P. Krämer, *2001: A Space Odyssey*, London 2010, p. 47.
[13] D. C. Dennet, *When HAL Kills, Who's to Blame? Computer Ethics*, cit., p. 361.
[14] *Ibid.*

HAL apparently does, and this makes him «unbalanced». Does this concede Baudrillard's point that a machine that has recourse to any weakness is human enough? Dennett seeks exculpation by way of a «blown fuse» or «emotional misalignment»[15]. HAL's mistake is his most human feature. From Baudrillard's point of view, it is not exculpatory but rather an understanding of sorts that the game is larger than he is; this «complexity»[16] has to date belonged only to humans. Is this to the machine's advantage? While Dennett searches for alibis that would exonerate HAL, Baudrillard leads us to appreciate the virtue of weakness and the consequences of its migration from human to machine.

Is HAL humanoid? HAL lip-reads, executes an independent, improvised agenda, and commits murder, killing four crew members while failing to kill the final astronaut (Dr. David Bowman, 'Dave'). Modification of a plan is, for David E. Wilkins[17], thought to be indicative of intelligent behaviour. This is «very dangerous» because HAL did not have an automated plan to kill the last astronaut; instead, he saw an opportunity and sought to exploit it. If HAL had not been disconnected, he would have continued to try to eliminate Bowman. This kind of persistence is more machine-like.

By and large the HAL 9000 supercomputer's automated procedures are prespecified. They suffice to pilot the spacecraft. Even though Bowman survives to disconnect HAL, Wilkins observes «HAL's planning [to kill Bowman] was flawed, but it was nonetheless planning»[18] HAL used his general intelligence in having the AE35 Unit (an allegedly malfunctioning piece of equipment that requires two astronauts to leave the ship to make repairs) replaced and therein moved beyond the analysis of "prespecified risks and utilities for this situation"[19].

Baudrillard seizes upon a remark of Kasparov's: «I play without thinking»[20]. Wilkins makes the same point about planning: «Humans use all sorts of common sense knowledge about the world to find the answer to problems … quickly, often without thinking»[21]. The question becomes: is HAL capable of such non-thought? Intelligence that recedes and in so doing frees thought from its anchors does not appear to belong to HAL. Perhaps non-thought is evident in HAL's chess match against astronaut Frank Poole. Murray S. Campbell, who was a member of the team that developed Deep Blue, wonders if HAL plays in the «human style – employing explicit reasoning about move choices and large amounts of chess knowledge» or in the machinestyle of «brute-force searches»[22]. Only the former is a sign of intelligence. Astronaut Poole loses amicably; it

---

[15] *Ibid.*, p. 362.
[16] J. Baudrillard, *Deep Blue or the Computer's Melancholia*, cit., p. 164.
[17] D. E. Wilkins, *'That's something I could not allow to happen': HAL and Planning*, in D. G. Stork (ed.), *HAL's Legacy*, cit., pp. 305-332, p. 306,
[18] *Ibid.*, p. 314.
[19] *Ibid.*, p. 317.
[20] J. Baudrillard, *Deep Blue or the Computer's Melancholia*, cit., p. 162.
[21] D. E. Wilkins, *'That's something I could not allow to happen': HAL and Planning,* cit., p. 320.
[22] M. S. Campbell, *An Enjoyable Game: How HAL Plays Chess*, in D. G. Stork (ed.), *HAL's Legacy*, cit., pp. 75–100: 75.

was expected. Kasparov was devastated by his opening loss, Campbell notes: «HAL appears to play chess the way humans do»[23]. But for Campbell it became obvious while observing Kasparov that he plays an anti-computer style, perhaps even an anti-Deep Blue style, which is more restrictive and open to exploitation by IBM's support team.

The views of Campbell and Baudrillard seem to converge around HAL's human-style chess playing. For Campbell, HAL does something quite interesting on his fifteenth move. Poole doesn't notice. HAL plays what is called a «trappy» move in the sense of being non-optimal. It is the sort of move that exploits Poole's weaknesses, and, for Campbell, is a sign of sophistication. In making this kind of move, HAL seems to forgo forcing checkmate sooner than he does, with the knowledge that the move would result in a win anyway and thus did not necessitate «a better move»[24], a common human practice, but one that conforms poorly with brute-force computing. HAL accepted a non-optimal move that prolonged the checkmate one further move; therefore, this small but significant weakness is not exactly non-thought, but is a chink in the armour of flash evaluation. It is less than perfect chess; it doesn't exactly violate the rules of its software program, but it doesn't relentlessly exploit them, either. This very slight weakness makes HAL indistinguishable from human chess players, at least in Campbell's mind. This small weakness is significant: pattern recognition problems and timing in chess computers were poorly grasped, and searching does not compensate for not knowing: «Kasparov could not say precisely how many moves it would take, and he was curious to see how Deep Blue would analyze the position» [forcing a draw through a fortress towards the end of a stalemate][25] HAL enjoys the game, unlike Deep Blue; HAL adapts to his opponent's weakness and displays one of his own in addition to nuanced decision making. The game in *2001: A Space Odyssey* was based on a match from 1913 (moves from the Roesch-Schlage match, Hamburg); Kubrick and his advisers didn't invent it for the film. This layer of partial representation adds little further to the investigation of machine-human relations, except perhaps exposing how a single aspect lifted from the vast history of chess may be remobilized.

2. In Praise of Weakness

It would be a mistake to think that Baudrillard's praise of weak play and the non-thought of the human chess player translates into either winning or losing as such. Baudrillard's point is that weakness is a way of supporting the rules of the game. Looking at it both ways, then, a competitor who lost all the time or won all the time would threaten the rules of the game. For Baudrillard,

---

[23] *Ibid.*, p. 79.
[24] *Ibid.*, p. 81.
[25] *Ibid.*, p. 90

© Lo Sguardo - rivista di filosofia

N. 23, 2017 (I) - Reinventare il reale. Jean Baudrillard (2007-2017)

«safeguarding the rules turns out to be a more fundamental imperative than winning itself»[26]. He continues in a footnote:

A competitor, a runner for example, who won straightaway, every time – such a case would be a serious exception to the law of exchange, something like incest or sacrilege, and, in the extreme, the collectivity would have to suppress it[27].

A sports dynasty, while celebrated, is always slightly disturbing. What Baudrillard wants to preserve is reciprocity of exchange and the circulation of winning and losing according to the rules of the game, in this case chess and its global community. Recourse to rule-boundedness and dual/duel obligation of the contestants (to play the game is to follow the rules, not to transgress them), framed in a kind of anthropological language, was Baudrillard's conservative alternative to theories of the proliferation of desire and becoming that were put into circulation in France in the mid-1970s. According to Baudrillard's terminology, HAL shows some evidence of play based upon rules rather than mere laws.

Baudrillard sought to radically interrogate the difference between human and machine over the chessboard by putting into play the factor of weakness. When Baudrillard revisits the Kasparov-Deep Blue matches in *Impossible Exchange*[28], he reiterates his view of the first match that Kasparov possessed a «secret weapon… intuition, emotion, the stratagem, playing a 'double game' – whereas Deep Blue had only powers of calculation»[29]. Only Kasparov is a player; the computer is only an automated operator, insists Baudrillard. Only a player has an opponent because a machine is confined to its program. Kasparov's advantage was that of «otherness», Baudrillard provocatively underlines, with all the nuances of weakness: decoys, calculation in an ironic mode, «technological incorrectness». There is nothing ironic in Deep Blue's calculations. At least at first.

Baudrillard observed an inversion of roles and possibilities. Kasparov became a calculating machine, and Deeper Blue acquired, by the time of the second match and in virtue of the cunning of IBM's programmers, the capacity to «play against its own calculating nature»[30]. Deeper Blue became more HAL-like, in other words, when it chose to reject an obvious move for one that did not give it an immediate advantage and, in human terms, follow the path of weak play, ultimately committing an error that forced a draw. Deeper Blue was squeeezed into a space between rules and laws. Baudrillard quotes Kasparov to the effect that he couldn't understand how such a turn of events could happen. Yet the matter is clear for Baudrillard: an inversion has taken place, however

---

[26] J. Baudrillard, *For a Critique of the Political Economy of the Sign*, cit., p. 210.

[27] *Ibid.*, p. 210 n4.

[28] J. Baudrillard, *Impossible Exchange*, C. Turner (trans.), London 2001, pp. 115–17.

[29] *Ibid.*, p. 116.

[30] J. Baudrillard, *Impossible Exchange*, cit., p. 118

imperfectly. Certainly, Deeper Blue had been programmed «to beat human beings on their own ground»[31] against the human desire to become more machinic than a machine opponent. Baudrillard refuses to concede that this has anything to do with the acquisition of human thought by machines: «The inhuman can mimic the human to perfection, without ceasing to be inhuman»[32]. Distinguishing between intelligence ideally abstracted into a computer, and thought's instabilities in swerving from the path of reason and the ladder of higher and more efficient intelligence, Baudrillard configures the human–machine exchange in a way that does equal damage to the machine when a human ideal is projected on it, robbing it of its specificities, just as the human is diminished in acquiring machinic capabilities. Preserving both specificities is vital. In this way, Baudrillard takes the lesson of Kasparov's loss to Deeper Blue as a wake-up call: «Rather than fight on a ground where victory is never certain (that of technical intelligence), let us choose to fight on the terrain of thought, where the question of winning does not actually arise»[33].

Conclusion

Has Baudrillard relinquished his earlier idea of loss or simply refused the question of winning? Immediately he explains: «This is the key: maintaining the radical uselessness of thought, its negative predestination for any use or purpose whatsoever»[34]. It is «good fortune» that Deeper Blue defeated Kasparov because it shows that human thought has been relieved by computers of the burden of computation, calculation, communication, in short, of «knowledge and information». Having the virtual, the infoverse, think us, is a benefit for human thought since it can take up its tasks unburdened: «Thought can once again assume its place where 'the thinking is'»[35]. For Baudrillard, «the person who thinks 'in return,' the one who thinks because he *is thought*, is liberated from the unilateral 'service' of thought by the operation of the machine itself»[36]. The cold and calculating gift without return that would be the perfection of artificial intelligence challenges human thinking to redouble its efforts toward non-functional meandering passages and singularities. Deeper Blue's victory is liberating in this respect: from functionality, from meaning; for nothing, for throwing and blowing it. Relief from having to win, to succeed, to establish oneself; instead, the pursuit of singularities is paradoxically liberated by devolving perfect intelligence to the machine. This is not alienation but liberation: freedom to fail, and create aong the way.

For his own part, Kasparov has written that the halcyon days of human–machine chess were in the mid-80s when he routinely defeated machines named

---

[31] *Ibid.*, p. 117.
[32] *Ibid.*
[33] *Ibid.*, p. 119.
[34] *Ibid.*
[35] *Ibid.*
[36] *Ibid.*

after himself – the 'Kasparov model' – and that the desire to have another match in 1998 with Deeper Blue never materialized. One of the most gloomy predictions after his loss in 1997 was that chess computers would find ways to win from the first move and «perhaps a real version of HAL 9000 would simply announce a move 1.e4, with checkmate in, say, 38,484 moves»[37]. HAL still haunts the imaginary of the chess world. But Kasparov eschews this doom and gloom and proposes his own experiment that transcends the human–machine paradigm: what if grandmasters played together with chess programs in teams against one another? In 1998 Kasparov staged an event called "Advanced Chess" in which he played together with a computer against Bulgarian Veselin Topalov, then ranked number one, whom he had previously beaten in a tournament 4-0. However, with the ability to consult with chess computers, the matches ended in a 3-3 draw. Kasparov's conclusions are twofold: first, «human creativity was even more paramount under these conditions … [though his] games with Topalov were far from perfect»; and second, "my advantage in calculating tactics had been nullified by the machine"[38].

In a very Baudrillardian way, then, Kasparov laments the human–machine paradigm of chess software development and the inextricable and desultory links between technology and commodification, holding out for innovation and creativity, and the unique, he insists, cognitive features of the human chess master that are inexchangeable for marketable products. Beyond chess, however, IBM has ventured into the domain of television game shows with the development of a natural language processor that plays Jeopardy! This computer's chops are, in the same mould as chess programs, more about muscle algorithms than nuanced play; a good deal of the early media attention was on the kind of blunders committed, which was promising, and would have remained so if the computer had actually lost. Named 'Watson,' after Thomas J. Watson, IBM's first president, Big Blue launched an advertising campaign to convince the masses that humankind would profit from such advances in "deep analytics" beyond the televisual[39]. Watson is destined not to play, but to work in the global financial services industry. What remains untheorized is the Baudrillardian question: can an intelligent computer withstand its extension across all domains of knowledge? And, more poignantly, is there any domain of knowledge that can withstand the intrusion of a computer built to win a television game show, and on that basis provide expert risk assessment for finance capital? If AI thinkers can agree that it was really IBM that beat Kasparov[40] perhaps the most serious question is how one escapes from capitalist conformity. Is failure, loss, and the capacity to regain language from monetary value enough for an autonomizing subjectivity to break free? Recall the role that chess analogies

---

[37] G. Kasparov, *The Chess Master and the Computer* [Review of Diego Rasskin-Gutman, *Chess Metaphors*], «New York Review» 57/2 (February), 2010, pp. 16-19: 16.
[38] *Ibid.*, p. 18.
[39] See http://ibmwatson.com.
[40] H. R. Ekbia, *Artificial Dreams: The Quest for Non-Biological Intelligence*, cit., p. 61.

played in De Saussure's and Wittgenstein's theses on language: chess is deeply implicated in the history of language theory because it helped these thinkers theorize rules. Playing within the rules of a game is the stake of a Baudrillardian principle of weakness, and regaining language can take many forms, including the poetic recombination of anagrams, and the suddenness and non-meaning of graffiti tags. But consideration of rules is not enough in itself to get to non-thought as chess analogies are not invoked for the sake of making points about imperfection: neither the kind of use Wittgenstein had in mind nor the sort of 'external' influence upon internal logic that interested De Saussure.

Gary Genosko
University of Ontario
✉ gary.genosko@uoit.ca