



Marina Núñez. *Imago (Inés)*, 2009. Vídeo monocanal, 2 minuts. Produït per MUSAC, Lleó.

«OBRE LA PORTA DEL COMPARTIMENT DE CÀRREGA, HAL»

PER QUÈ ENS FAN POR ELS ROBOTS?

David Casacuberta

“Open the Pod Bay Doors, Hal”. Why Do We Fear Robots?

This article discusses two of the most common fears associated with the world of robotics: *machines revolt* and *uncanny valley*. On the premise that they are not fears founded on rational evidence, the author tries to establish the ultimate reason for such fears; postulating how robots could symbolise our doubts about the nature of consciousness.

Tenim la tendència a pensar que la ciència-ficció tracta del futur. Superficialment és així, però si analitzem en profunditat aquest gènere descobrirem que aquest futur és bàsicament una metàfora per parlar del present. Quan George Orwell cercava un títol per a la seva distopia sobre un estat futur totalitari, va decidir intercanviar les últimes xifres d'aquell any, 1948, i així va néixer *1984*. Ficció científica, però realitat sociològica.

El mateix podria dir-se de *2001: Una odissea de l'espai*, la famosa pel·lícula d'Stanley Kubrick (basada al seu torn en una novel·la d'Arthur C. Clarke), una de les frases més famoses de la qual hem utilitzat per al títol d'aquest article. Probablement el lector recorda aquesta tensa escena en què una intel·ligència artificial paranoica –Hal 9000– ha planejat eliminar tota la tripulació d'una nau espacial. Al capità de l'expedició, en David Bowman, ha decidit deixar-lo fora de la nau, dins una càpsula minúscula, abandonat a la seva sort. La pel·lícula captura magistralment la por sistemàtica que els humans tenim davant una revolta de robots que ens considerin prescindibles i acabin dominant la Terra, exterminant-nos o convertint-nos en esclaus. Kubrick amplifica la por i la tensió amb diferents elements estilístics. Un dels més comentats és el contrast entre un Hal embogit i molt humà enfrontant-se a humans freds i calculadors que semblen més aviat màquines. Aquesta transposició és molt significativa i la recuperarem més endavant.

«2001: UNA ODISSEA DE L'ESPAI CAPTURA MAGISTRALMENT LA POR SISTEMÀTICA QUE ELS HUMANS TENIM DAVANT UNA REVOLTA DE ROBOTS QUE ENS CONSIDERIN PRESCINDIBLES I ACABIN DOMINANT LA TERRA»

La sèrie *Terminator*, una altra història pretesament sobre el futur però que també ens parla del present, conté també elements molt significatius sobre aquesta por. A l'hora d'explicar la gènesi que va portar a un món futur dominat per màquines, representa un paper clau una intel·ligència artificial anomenada Skynet. Segons s'explica en la pel·lícula, un cop Skynet, per un error de construcció, aconsegueix l'autoconsciència, la seva primera acció és produir una guerra nuclear que gairebé elimina la humanitat. En la pel·lícula, que la primera acció d'un ésser autoconscient artificial sigui destruir els humans es troba natural. Kubrick va considerar necessari explicar la bogeria de Hal apel·lant al fet que la màquina havia rebut ordres contradictòries. Per a James Cameron –el director dels dos primers lliuraments de *Terminator*– donar explicacions era innecessari. Els robots intel·ligents, de natural, volen exterminar-nos.

■ LA REVOLTA DELS ROBOTS I LA VALL ESTRANYA

Tenim aquí la primera por: la por a una revolta de les màquines en el moment en què puguem adquirir les nostres capacitats. Aquesta por ve normalment acompanyada per una altra, encunyada per l'expert en intel·ligència artificial Masahiro Mori com «la vall estranya» (*the uncanny valley*) que sembla connectar el rebuig i la por als robots com més similars són als humans. Podem





© 1968 Metro-Goldwyn-Mayer



© 1968 Metro-Goldwyn-Mayer

En 2001: *Una odissea de l'espai*, el famós ordinador Hal pretén eliminar la tripulació de la nau. En les imatges veiem diverses escenes de la pel·lícula on apareix l'ordinador Hal 9000 i els astronautes. Kubrick ens presenta un robot embogit davant un humà calculador.

«LA POR D'UNA REBEL·LIÓ DELS ROBOTS ÉS PROBABLEMENT UNA DE LES MÉS IRRACIONALS DE SORTIDA. ACTUALMENT NO N'HI HA CAP QUE REPRESENTI ALGUN PERILL ASSOCIAT A LA SEVA INTEL·LIGÈNCIA PERQUÈ CLARAMENT NO EN TENEN CAP»



© 1968 Metro-Goldwyn-Mayer

suportar l'R2D2 de *La Guerra de les Galàxies* amb el seu aspecte clarament de màquina, però un robot amb aspecte humà –un androide– és molt més difícil de pair. I com més s'assembla a una persona més incòmodes ens trobem, com han mostrat els experiments de David Hanson recreant amb un enorme grau de realisme el cap d'Einstein, la versió robot de l'escriptor de ciència-ficció Philip K. Dick o el cantant Jules. Per molt que ens agradin els robots, és inevitable sentir quelcom estrany al cos veient algun d'aquests robots humanoides.

Observem en primer lloc que, entre totes les pors dels humans, aquesta és probablement una de les que més irracionals semblen de sortida. Els robots que es



© 2008 Warner Bros

La sèrie *Terminator* ens mostra un futur dominat per màquines sota les ordres de Skynet, una intel·ligència artificial que ha arribat a tenir autoconsciència. En la mateixa línia, en la pel·lícula *Jo Robot*, el protagonista interpretat per Will Smith (a la dreta) desconfia dels robots que més tard es tornaran contra els humans també sota les ordres d'una intel·ligència artificial. Aquestes pel·lícules reflecteixen un tret característic de la por dels humans cap als robots: ens fan dubtar de la nostra capacitat d'autoconsciència com a tret distintiu.

rebel·len contra els seus amos existeixen en la literatura molt abans que aquests fossin possibles –pensem en el Golem, per exemple—. Actualment no hi ha cap robot ni intel·ligència artificial que representi algun perill associat a la seva intel·ligència perquè clarament no en tenen cap, i el dia en què puguem parlar d'objectes com Hal 9000 o Skynet encara és molt llunyà. Aquesta por, per tant, no es basa en una d'experiència real amb robots autònoms i malvats, perquè aquests éssers mai no han existit.

Per altra part, la idea –força comuna en aquests escenaris apocalíptics– de com la revolta és resultat d'un «accident» no resisteix cap mena d'anàlisi racional. Un robot no és resultat d'un procés natural sinó una obra d'enginyeria. Potser es pot crear un gas tòxic per error en un laboratori i matar milers de persones, però no es pot crear un ésser autoconscient i malvat per error. Com ironitzava el dissenyador de robots autònoms Rodney Brooks, seria com si algú es fiqués en un garatge amb unes quantes planxes metàl·liques, cables, etc., comencés a fer proves i al cap d'un temps exclamés: «Deu meu! He creat un Airbus 300 per error! Què puc fer ara?»



© 2004 Twentieth Century Fox

La robòtica és aquella disciplina encarregada d'imaginar, conceptualitzar, dissenyar, construir i programar un robot. Amb *robot* en un sentit genèric ens referim a qualsevol sistema mecànic capaç de reconèixer i interactuar amb el seu entorn. Els robots acostumen a tenir alguna mena de sensors que funcionin de sentits per tal de poder observar l'entorn, algun sistema de locomoció o, si més no, alguna mena de mecanisme que els permeti interactuar i modificar l'entorn.

Malgrat que avui dia associem la robòtica a xips i ordinadors, la realitat, tal i com hem apuntat abans, és que la idea de robot és molt més antiga. Heró d'Alexandria (segle I dC) recull en els seus textos uns cent mecanismes i màquines, incloent-hi una màquina operada amb monedes. Als segles XVIII i XIX comencen a aparèixer, sobretot com a joguines d'exhibició, diverses màquines amb enginyosos mecanismes que imitaven el



comportament de persones i animals, com l'ànec mecànic de Jacques de Vaucanson o els arquers mecànics del japonès Tanaka Hisashige.

Una bona prova de la fascinació dels humans pels robots és que el terme *robot* no sigui pas la creació d'un enginyer –podria haver estat les sigles d'un complex protocol tecnològic– sinó d'un autor teatral, el txec Karel Capek, que imagina robots de forma humana que fan les nostres feines en la seva obra de 1920 *R.U.R (Rossum Universal Robots)*. El terme ve, per tant, del txec, on *robota* significa si fa no fa «treball dur».

D'altra banda, el gran pas teòric per a la disciplina de la robòtica és l'aparició de la cibernètica, de mans del matemàtic Norbert Wiener, que permet així construir potents models matemàtics per processar la informació que el robot rep de l'exterior i quina és la millor manera d'interactuar amb aquesta informació. A partir d'aquí el camp de la robòtica ha anat expandint-se cada cop més, aprofitant-se de les troballes de diferents disciplines, i ara tenim mans mecàniques amb una precisió similar a la de les humanes, robots capaços de passejar-se per la superfície del planeta Mart, i d'altres que ajuden persones discapacitades. La tendència en la recerca i en la producció industrial apunta a un desenvolupament futur centrat, per una part, en la llar, amb robots pensats per a ajudar-nos en les feines de casa, però també en assistents personals per a aquelles persones amb problemes de mobilitat, i amb aplicacions en medicina, amb la recerca de pròtesis que cada cop s'assemblin més a les seves contraparts humanes. L'altre camp de recerca on clarament els robots seran clau en el futur és, malauradament, la investigació militar, amb tota mena de projectes de míssils intel·ligents, avions espia, robots cercamines, etc.

La por als robots no té gaire a veure amb el desenvolupament de la ciència i la tecnologia. De fet té molt poc a veure amb els robots *per se*. Es tracta d'una por profundament filosòfica, que bàsicament té a veure amb la nostra idea de qui som realment.

■ L'ESPERIT DINS LA MÀQUINA

Si fem una arqueologia d'aquestes pors, diria que són facilitades per l'animisme, la creença religiosa segons la qual tots els éssers tenen en el fons una ànima. Fixem-nos que, a l'hora de la veritat, l'animisme és una mena de materialisme: si tot té ànima, això vol dir que la consciència, la ment, és una propietat intrínseca de la



El fet que el primer a encunyar el terme *robot* fóra el dramaturg txec Karel Capek –i no un enginyer com hauria estat previsible– és una prova de la fascinació humana pels robots. En la imatge, *R.U.R.*, el robot que va imaginar Capek el 1920.

«UN ROBOT NO ÉS RESULTAT D'UN PROCÉS NATURAL SINÓ UNA OBRA D'ENGINYERIA. POTSER ES POT CREAR UN GAS TÒXIC PER ERROR EN UN LABORATORI, PERÒ NO UN ÉSSER AUTOCONSCIENT I MALVAT»

matèria, i per tant, natural. No hi ha cap substància espiritual misteriosa per explicar. La història de l'evolució de les creences religioses al món occidental passa per una dialèctica que consisteix a revisar qui té dret realment a tenir ànima.

Durant certes èpoques de la nostra història, només els homes blancs de certa posició semblaven disposar d'ànima. Aquesta era negada als animals, als membres

d'altres races o a les dones. Aquesta connexió entre matèria i ànima és el que explica que la idea de màquines que es revoltent pugui aparèixer en cultures tan antigues com la mochica (una cultura precolombina present al nord de Perú entre l'any 800 i el 1000). Dins aquesta cultura ens trobem amb el famós mural de *La revolta*





dels objectes, on apareixen plats, gots i tota mena d'estrutres domèstics amb cames i potes enfrontant-se als humans. Els inques anomenaven *pachacuti* a un procés similar, en què, en períodes de caos, que anuncien la fi d'un cicle i l'aparició del següent, objectes i animals es revoltent contra els seus amos humans.

Així, allò que fa possible que ens facin por els robots no és la seva presència cada cop més real en les nostres vides, sinó un cert «mòdul animista» en el nostre cervell que ens facilita pensar objectes inanimats com a objectes animats. I que llanci la primera pedra qui no ha parlat –o fins i tot escridassat– al seu ordinador personal quan no responia de la manera esperada a una instrucció nostra.

El filòsof i científic cognitiu Daniel Dennett anomena aquest procés «atribució intencional» i –resumint molt– diu que una de les nostres formes favorites de conèixer el món és suposar que molts processos són animats, fins i tot quan sabem que en realitat no ho són. Per exemple, juguem contra un ordinador als escacs, l'ordinador fa una jugada i automàticament atribuïm una intenció al darrere: «Hmmm... Mira que és llest. Ara el molt malparit es vol menjar la meua torre.» Malgrat saber que l'ordinador no té cap mena d'intenció, i simplement empra un algorisme per valorar quina jugada resulta més eficient donada la situació de la partida, no podem evitar associar intencions, plans i desitjos a aquesta màquina, perquè és una de les maneres privilegiades per la nostra ment per conèixer el món. De fet, argumenta Dennett, en molts casos, aquesta «atribució intencional» és l'única forma raonable que tenim d'analitzar un determinat problema. Tornant a la partida d'escacs, si comencéssim a pensar en l'ordinador com un processador matemàtic, que assigna valors numèrics a cada possible jugada i llavors calcula la posició més eficient amb un algorisme, no podríem pensar de manera útil com guanyar la partida i fins i tot el programa més simple ens apallissaria.

■ ROBOTS, ZOMBIS I CRIATURES DEL PANTÀ

La revolta dels robots i la vall estranya són pors que manllevem temes de la ciència-ficció, però, en realitat, són històries de terror. Així, Hal 9000 o Skynet pertanyen més aviat al panteó dels monstres, i els hauríem de veure com les encarnacions tecnològiques d'éssers

com el Golem, Frankenstein o els més contemporanis zombis. Tots aquests éssers són metàfores d'una por ancestral a l'ésser sense ànima. El zombi és rematadament estúpid i Hal més intel·ligent que qualsevol de nosaltres, però tots dos són éssers heterònoms, sense consciència, morts en un sentit profund del terme, malgrat que es moguin o parlin. Podem veure aquí la gran habilitat de Kubrick quan remarca aquesta paradoxa presentant un comandant de la nau que és fred com un autòmat i un pobre Hal que, enfollit, canta «Daisy, Daisy» mentre és finalment desconnectat. No podem evitar sentir pena per aquella criatura artificial, però al mateix temps sabem que la raó per la qual ens provoca por era precisament la seva artificialitat.

Aquesta qüestió ha agafat forma teòrica en algunes paradoxes de la filosofia de la ment. Per exemple tenim l'experiment mental de la criatura del pantà, imaginada pel filòsof Donald Davidson en el seu article de 1987

**«LA POR ALS ROBOTS NO
TÉ GAIRE A VEURE AMB
EL DESENVOLUPAMENT
DE LA CIÈNCIA I LA
TECNOLOGIA. ES TRACTA
D'UNA POR PROFUNDAMENT
FILOSÒFICA QUE
BÀSICAMENT TÉ A VEURE
AMB LA NOSTRA IDEA DE
QUI SOM REALMENT.»**

«Coneixent la nostra ment». Davidson ens demana que imaginem que passem un dia a la vora del pantà i, per un esdeveniment físic molt improbable però no impossible, cauen dos llamps. Un d'ells ens mata, i l'altre fa que milions i milions de molècules que es trobaven al pantà es recombinin per crear una còpia exacta de nosaltres. El nostre cos original s'enfonsa al pantà, però el doble continua el seu passeig, actuant en tot moment com ho faríem nosaltres. Tanmateix, a aquesta criatura del pantà, encara que s'assembla exactament a nosaltres, no li podem

atribuir vida mental. La raó és un xic complexa, però filosòficament molt rellevant.

Els nostres pensaments són sobre objectes i situacions del món exterior. El record de què vam esmorzar ahir és possible perquè ahir vam interactuar amb una sèrie d'objectes del món extern que vam utilitzar per fer-nos l'esmorzar. Hi ha una connexió casual entre aquest record i la realitat. La nostra creença que un ou és una cosa fràgil que es pot trencar si no el manipulem amb cura es basa en interaccions anteriors amb ous. Així, per a ser un pensament, cal ser un pensament d'alguna cosa. Necessita un enllaç causal amb la realitat. La criatura del pantà pot dir: «Els ous són molt fràgils», però aquesta frase, en boca seva, no significa res perquè mai no ha estat en contacte amb cap ou, ja que acaba de ser creat en un procés fantasmagòric fa només minuts.

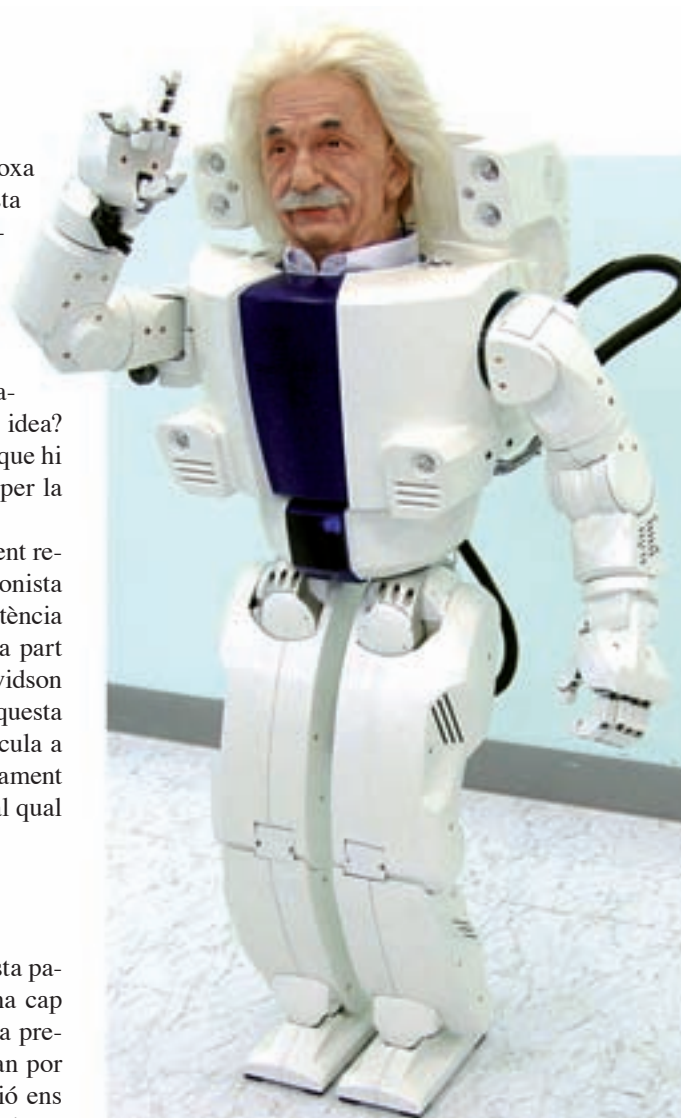
Una forma més visible d'activar aquesta paradoxa és la següent. Imagini que algú li fa una proposta força peculiar, en què se li ofereix la immortalitat. Aquest misteriós personatge li assegura que no li ha de preocupar morir-se. Al moment que mori, els seus equips tecnològics superfantàstics crearan una còpia exacta de vostè amb el mateix aspecte físic i els mateixos records, de manera que seria immortal. Li semblaria una bona idea? O no pensaria més aviat que si un ha mort, saber que hi ha un doble exactament igual a un passejant-se per la Terra tampoc és gaire consol?

Aquesta paradoxa de Davidson és filosòficament rellevant perquè planteja un problema a una teoria monista de la ment. Si un és materialista i no creu en l'existència de substàncies espirituals, com pensem una bona part de la humanitat actualment, la paradoxa de Davidson és eminentment molesta, ja que estableix que aquesta criatura del pantà és idèntica molècula per molècula a nosaltres. Tenim, per tant, un objecte que és físicament idèntic a nosaltres, molècula per molècula, però al qual resulta antiintuïtiu atribuir estats mentals.

■ LA CONSCIÈNCIA HUMANA I ELS ROBOTS

Aquest no és l'espai per intentar solucionar aquesta paradoxa, entre altres raons perquè encara no hi ha cap solució que faci felicitat a tothom. Simplement volia presentar-la aquí per afinar la raó per la qual ens fan por els robots. Dèiem al principi que la ciència-ficció ens parla del present, mostra els nostres desitjos, fixacions, creences i pors en un escenari futurístic. A quina por finalment respon aquesta angoixa d'una revolta dels robots? Part s'explica, sens dubte, per la desconfiança davant una tecnologia cada cop més implicada en tots els aspectes de la nostra vida. Els estudis culturals han escrit a bastament sobre la manera com els monstres simbolitzen pors i prejudicis sociològics concrets. Així, no és difícil veure en moltes pel·lícules de zombis la por de l'home blanc davant la cultura africana, els ovnis com a metàfora de l'enemic comunista o els robots com a símbol d'un estat capitalista que tracta els homes com a màquines.

Tanmateix, aquestes explicacions són parcials. D'una banda no expliquen aquesta presència universal del mite en tota mena de cultures al llarg de la història. No podem pas dir que els mochica vivien en una societat capitalista alimentada per la revolució industrial. Però, sobretot, aquestes explicacions no aclareixen l'efecte de la vall estranya; no ens expliquen per què els robots són encara més inquietants com més s'assemblen a nosaltres.



© Hanson Robotics

«LA RAÓ ÚLTIMA PER LA QUAL TENIM POR ALS ROBOTS ÉS L'ANGOIXA PROFUNDA DAVANT LA REALITAT SEGONS LA QUAL, DESPRÉS DE TOT, NO HI HA CAP EGO. NO TENIM UNA IDENTITAT ESTABLE I CONTÍNUA»

La resposta de per què ens fan por els robots és que ens fan por els monstres. I els monstres radicalitzen característiques fosques que tots tenim a dins. Dràcula o la seva versió actual secularitzada de l'assassí en sèrie Hannibal Lecter capturen aquesta pulsio per la maldat pura present en nosaltres. El zombi ens recorda com de buides poden estar les nostres vides. El robot té una funció similar, en aquest cas fent-nos dubtar de la qualitat que més sovint es cita quan volem distingir els humans de la resta de criatures: la nostra capacitat



© Hanson Robotics

Com més s'assembla un robot a una persona, més incòmode ens resulta. Aquesta por és coneguda com «la vall estranya». En la pàgina de l'esquerra, Albert Hubo, robot creat per David Hanson a imitació del cap d'Einstein. Dalt, el mateix Hanson col·locant els enllaços que després donaran expressió a un robot.

d'autoconsciència. Aquesta és la primera manifestació de la por: i si finalment no som tan especials? I si al capdavant l'autoconsciència pot ser aconseguida per màquines creades per nosaltres? En quin lloc ens deixaria llavors? Si Copèrnic ens va fer fora del centre de l'univers, i Darwin ens va convertir en un animal més, la revolució de la intel·ligència artificial ens traurà també l'autoconsciència com a tret distintiu?

Malauradament, aquesta no pot ser tota la veritat. Com hem dit diverses vegades, la possibilitat d'aconseguir robots autoconscients –suposant que sigui possible– és encara molt remota. I no explicaria la continuat temporal del mite de la revolta de les màquines. La raó última per la qual tenim por als robots és l'angoixa profunda davant la realitat que, després de tot, no hi ha cap ego. No tenim una identitat estable i contínua que va sorgir quan vam néixer i ens acompanyarà fins a l'últim dia. Potser, com afirmen els budistes, no hi ha essències, no hi ha substàncies. La nostra identitat no és estable, sinó que co-varia amb els esdeveniments que ens envolten, la gent amb la qual interaccionem, etc. Estem buits com els robots, perquè ni ells ni nosaltres disposem de cap característica intrínseca que defineixi qui som, sinó que som totalment dependents de causes i condicions.

**«HAL 9000 O SKYNET
PERTANYEN MÉS AVIAT AL
PANTEÓ DELS MONSTRES,
JUNTAMENT AMB EL GOLEM,
FRANKENSTEIN O ELS
ZOMBIS. TOTS AQUESTS
ÉSSERS SÓN METÀFORES
D'UNA POR ANCESTRAL
A L'ÉSSER SENSE ÀNIMA»**

Una primera resposta científica que podríem donar és observar que la identitat humana està garantida a escala biològica: bona part del que fem i som és almenys parcialment definit pel cos que tenim, l'estructura del nostre cervell, els nostres gens, etc. Observem, però, que això no ho negaria pas el pensament budista. En cap moment neguem que les persones existeixin, el que neguem és l'existència d'una entitat mental estable i immutable al llarg de la nostra vida, l'existència d'una substància mental independent de causes i condicions.

Si mirem l'evidència científica d'una manera més detallada veurem que la genètica no defineix de manera monolítica qui serem, sinó que interactua de manera contínua amb l'ambient, transformant el que pensem, la forma com reaccionem, què ens agrada o desagrada, etc. Això no significa ni de bon tros que siguem una *tabula rasa* on tot és possible: els nostres gens defineixen una mena d'espai combinatori de possibilitats, fent que certes coses siguin més probables que d'altres i algunes impossibles. Aquesta xarxa de combinacions està en interacció contínua amb l'ambient, és el que en

filosofia de la ment s'anomena *en-activisme*, i és aquest procés precisament el que posa més de manifest la nostra natura clarament indefinida, oberta a transformacions i contínuament emergent de la nostra ment.

Segons els budistes, descobrir això no és cap mala notícia. Ans al contrari, és la clau per aconseguir la veritable felicitat. Descobrir que no hi ha cap *ego*, cap identitat substancial per la qual lluitar, és el que ens fa finalment lliures i ens permet viure en pau i felicitat. A la nostra mentalitat

occidental, en què gran part de les activitats que portem diàriament tenen com a últim substrat metafísic aquest subjecte estable, acceptar això pot resultar-li eminentment difícil. Potser necessitem llegir més històries sobre robots enfollits. ☺

BIBLIOGRAFIA

- MORI, M., 1981. *The Buddha in the Robot*. Kosei Pub. Tòquio.
 PANIKKAR, R., 2002. *El silencio del Buda. Una introducción al ateísmo religioso*. RBA. Barcelona.
 PICARD, R., 1998. *Los ordenadores emocionales*. Ariel. Barcelona.
 VARELA, F., 1997. *De cuerpo presente: las ciencias cognitivas y la experiencia humana*. Gedisa. Barcelona.

David Casacuberta. Professor agregat de Filosofia de la Ciència. Universitat Autònoma de Barcelona.

