

INTEL·LIGÈNCIA A TONES

TON SALES

INTEL·LIGÈNCIA  
A TONES

Ton Sales

ACIA

TON SALES

# INTEL·LIGÈNCIA A TONES

ACIA

Disseny de la coberta: Anabel Pellicer  
Correcció: Alt\_Formació (<http://www.alt-formacio.cat>)

© 2013 Ton Sales

Editat per: Josep Puyol Gruart  
Associació Catalana d'Intelligència Artificial, ACIA  
Institut d'Investigació en Intelligència Artificial  
Campus de la UAB, 08193 Bellaterra, Catalunya

Dipòsit legal: B-23.889-2013

*Primera impressió, octubre de 2013*

El 28 de març de 1994 un «col·lectiu de persones de diverses universitats catalanes», com descrivia el butlletí de l'ACIA número zero, ens vam reunir a Blanes per crear l'Associació Catalana d'Intel·ligència Artificial. Una de les primeres actuacions del consell rector dirigit per l'Enric Plaza va ser crear el butlletí de l'Associació. El número zero va sortir i va ser distribuït aquell mateix estiu. Els números s'han succeït des d'aleshores sota la direcció, primer del Carles Sierra, i després d'en Ramon López de Mántaras, l'Enric Plaza, la Nuria Agell i el Josep Puyol.

Han passat 20 anys i s'han publicat més de 50 números. Durant tots aquests anys el butlletí i l'ACIA han pogut comptar amb uns col·laboradors excepcionals: Enric Plaza, Llorenç Valverde, Miquel Barceló i Ton Sales. Ells ens han acompanyat amb les seccions El Cau del Hàcker, Amorrat al teclat, Intel·ligència Ficció i Intel·ligència a Tones, respectivament. Amb els seus articles han estat l'ànima del butlletí.

L'associació va decidir recollir els articles dels nostres col·laboradors i publicar-los en format de llibre, tant per agrair-los la seva col·laboració com per a donar més difusió als seus escrits. El resultat són quatre llibres amb unes 1000 pàgines. Cada llibre va precedir d'un prefaci i inclou els articles d'un dels autors. Els articles són presentats en ordre cronològic.

La publicació d'aquests llibres ha comptat amb el treball voluntari i desinteressat de molts membres de l'associació que han ajudat a preparar els textos en format L<sup>A</sup>T<sub>E</sub>X a partir dels butlletins escanejats, primer, o butlletins en diversos formats electrònics, després. La coordinació, i una gran feina addicional, ha anat a càrrec del Josep Puyol. Gràcies a tots ells.

Vicenç Torra  
Investigador científic de l'IIIA  
President de l'ACIA

Bellaterra, 4 de setembre de 2013



## *Prefaci*

Quan vam començar a publicar el butlletí de l'ACIA (Associació Catalana d'Intel·ligència Artificial), ara ja fa gairebé vint anys, creïem que tenir columnistes fixos seria un atractiu per als socis. Si aconseguíem que investigadors de prestigi parlessin de manera amena sobre el present i el futur de la intel·ligència artificial els socis esperarien cada número del butlletí com s'espera un croissant assegut a la terrassa d'un cafè de París: amb ànsia. En Ton es va deixar convèncer i la seva columna va esdevenir ràpidament font d'inspiració per a moltes discussions entre nosaltres, a l'hora del cafè, a les sobretauls i a les pauses dels congressos.

En Ton és un vell conegut de la comunitat d'IA catalana. Ha ensenyat lògica a moltes generacions d'informàtics i sempre ha fet les classes denses però entretingudes, en línia amb el seu estil de divulgador. Li encanta explicar històries, farcir d'anècdotes les explicacions de punts importants de la història de la ciència i obrir parèntesis, molts parèntesis, que miraculosament aconsegueix tancar cada vegada. Tancar bé els parèntesis, entre altres virtuts, li ha permès guanyar-se el respecte i l'admiració de la comunitat científica catalana. Ha participat en nombrosos congressos catalans d'IA fent xerrades, participant en taules rodones, i discutint, sempre discutint amb tothom. En Ton s'estima la lògica, la intel·ligència artificial, i tot allò que està a la frontera del coneixement humà. Sempre busca les arrels de cada tema, el seu origen, les possibles influències creuades entre els investigadors.

En moltes de les columnes que es recullen en aquest llibre en Ton manté un to il·lustrat, de professor que vol captivar l'estudiant amb les idees de grans científics. Ens apropa als personatges des d'una perspectiva moderna, actual, fent-nos veure de quina

manera precisa els seus treballs han influït en el desenvolupament de la intel·ligència artificial i com encara influiran en el futur. Un dels personatges singulars que trobareu en aquest llibre, sobretot a les primeres columnes, és en Ramon Llull. En Ton sempre ha treballat per divulgar el paper d'en Ramon Llull com un dels pares precursors de la Intel·ligència artificial. En particular, del raonament simbòlic automatitzat. Això el va portar recentment, una mica instigat per mi, a participar en una sessió monogràfica sobre Llull al congrés IJCAI 2011 a Barcelona. Un gra de sorra en la reivindicació de Llull davant la comunitat internacional. Apart del seu estimat Llull, el lector trobarà freqüents referències als pares de la lògica i la computació, Turing, von Neumann, Gödel, i als pares moderns de la IA com ara Minsky, Simon o McCarthy. No defuig de donar la seva opinió sobre personatges més aviat polèmics com en Kurzweil, i sempre manté un to distès i entretingut que fa que molt sovint el lector vulgui saber-ne més del personatge i de les seves gestes.

Vist en conjunt, aquest llibre és una anàlisi profunda sobre què és la intel·ligència, humana i artificial. Trobem articles sobre neurociències, emocions, eixams, robòtica, i altres temes calents. Però sobre tot hi trobem una reflexió ponderada sobre el futur de la IA i les seves possibilitats per construir un món millor.

Estic segur que qui s'estreni amb aquest llibre com a lector d'en Ton no podrà parar de llegir-lo. Vigileu que no se us cremi l'arrostit!

Carles Sierra  
Professor d'investigació i subdirector del IIIA  
President de l'ACIA del 1998 al 2002

## Notes de l'editor

Editar aquesta sèrie de quatre llibres ha sigut una feina enorme que ha implicat moltes persones. Però l'objectiu s'ho valia. Molta gent de l'Associació deia que s'havia d'agrair i reconèixer la aportació continuada i tan valuosa dels col·laboradors habituals a la revista NODES, i aquest és el resultat. Jo crec que fa goig.

És molt important agrair a una gran quantitat de persones que aquests llibres siguin una realitat: al consell rector de l'ACIA, representat per en Vicenç Torra i la Bea López; als anteriors directors del Butlletí de l'ACIA i de NODES, en Carles Sierra, en Ramon López de Mántaras, l'Enric Plaza i la Núria Agell; i a una llista molt gran de gent –la majoria socis– que ha fet la «traducció» a  $\LaTeX$  de tots els articles dels quatre autors, exactament 145 articles. També volia reconèixer especialment l'ajuda en l'*esprint* final de l'Eva Armengol.

Com que tot això s'ha fet en secret i era una sorpresa, no podíem demanar l'opinió dels autors. Espero que les decisions que hem pres –de bona fe– no s'allunyin gaire de les que haurien pres els autors.

Entre les decisions més importants hi ha les següents: hem eliminat pràcticament tots els enllaços en comprovar que la majoria ja no funcionaven; només hem posat les figures que eren necessàries per a la comprensió del text; no hem unificat totes les referències dels articles; hem intentat donar unitat a la disparitat de formats dels articles, triant l'estil Tufte- $\LaTeX$  (basat en el treball de Edward R. Tufte) i fent modificacions al nostre criteri; i n'hem encarregat la correcció a uns bons professionals.

Hem publicat els articles dels quatre autors continguts als números de l'1 al 50. Els 30 primers números, només els teníem en



paper i, per tant, s'han escanejat prèviament; del 31 al 46 hem fet servir els PDF, i del 47 al 50 ja estaven en format web.

Ha estat un treball col·laboratiu apassionant que ens ha donat a tots l'oportunitat de llegir o rellegir part dels vostres treballs. Ara amb aquest llibre ja els podem llegir tots amb comoditat. Esperem que us agradin!

Josep Puyol Gruart  
 Científic titular de l'IIIA  
 Director de NODES des d'octubre de 2011

### *Col·laboradors directes<sup>1</sup>*

Núria Agell Jané (UPF), Pablo Almajano (IIIA), Cecilio Angulo Bahon (UPC), Josep Lluís Arcos Rosell (IIIA), Eva Armengol Voltas (IIIA), Ramón Béjar Torres (UdL), Félix Bou Moliner (UB), Dídac Busquets Font (Imperial College London), Jesús Cerquides Bueno (IIIA), Francesc Esteva (IIIA), Àngela Fàbregues Vinent (ArteRed), Zoe Falomir Llansola (U. Bremen), Pere Garcia Calves (IIIA), Àngel García Cerdanya (UPF), Karina Gibert Oliveras (UPC), Lluís Godo i Lacasa (IIIA), Jordi Levy (IIIA), Beatriz López Ibáñez (UdG), Felip Manyà (IIIA), Joaquim Melendez Frigola (UdG), Carles Noguera i Clofent (UTIA Praga), Pablo Noriega Blanco-Vigil (IIIA), Jordi Planes Cid (UdL), Josep Puyol Gruart (IIIA), Arnau Ramisa Ayats (IRI), David Riaño Ramos (URV), Juan Antonio Rodríguez Aguilar (IIIA), Raquel Ros Espinoza (Imperial College London), Jordi Sabater Mir (IIIA), Miquel Sànchez i Marrè (UPC), Marco Schorlemmer (IIIA), Beatriz Sevilla Villanueva (UPC), Carles Sierra García (IIIA), Aïda Valls Mateu (URV), Amanda Vidal Wandelmer (IIIA), Ramon López de Mántaras (IIIA), Vicenç Torra (IIIA).

<sup>1</sup> Aquesta llista és la de tots els qui directament han fet feina per els llibres, després d'una tria entre –preferentment– els socis de l'ACIA. Hi ha altra gent que volia col·laborar i no ho ha fet perquè no ha pogut o no ha calgut. Puc assegurar que si fem una llista d'adherits, aquesta seria molt més llarga.

# Índex

Realment, ens ha d'interessar, que hi hagi IA?	13
Quin és el nucli dur de la intel·ligència?	17
La importància de dialogar i de passar-se	21
Una qüestió de sentiment	25
Cal tenir consciència	29
Doncs, què és la intel·ligència?	33
Racionalitat = Raonament + Emoció	39
Lògica, Informàtica i Intel·ligència artificial	43
La intel·ligència, de Llull al llimac	47
De la matemàtica a la IA	55
Llull com a informàtic <i>avant la lettre</i>	81
Lògiques per a formalitzar <i>estats mentals</i>	95
Llimacs, eixams i <i>càlcul col·lectiu</i>	103
La IA: ciència? tècnica? marc teòric?	107
Redacció de fi de curs	115
Sobre el principi (i la fi) de la IA	121

<i>Agents</i>	127
Calcular i somriure o «El que la intel·ligència no és»	139
Ordinadors emotius	145
Xarxes neuronals, cervells i ordinadors	165
La «Intel·ligència Artificial»: el nom no fa la cosa	187
El 2001 no va ser el d'en Kubrick	195
Les «arroves» i la Informàtica	201
Intel·ligència és més que un mot	209
El context de la Intel·ligència Artificial	215
Qüestions irresoltes: Potser que hi tornem?	219
Podrà mai un ordinador emular el cervell humà?	225
A propòsit de conills i meravelles que ningú no s'espera	229
Quin dia tindrem intel·ligència?	235
És perillós profetitzar, sobretot quan es tracta del futur	239
Hi ha una sola intel·ligència?	247
Després de l'hivern ve la primavera	257
Com construir un cervell	265
Encontres a la tercera fase?	271
Robodependència	281
Crònica de robots (continuació)	289
Bons desitjos per al 2007 i altres prediccions diverses	299
Seguint el camí, un any més	305

<b>El que tenim i el que ens espera</b>	<b>311</b>
<b>La dansa dels robotets</b>	<b>319</b>
<b>De sentiments (i de llur absència)</b>	<b>323</b>
<b>Seran proporcionals, els robots?</b>	<b>333</b>
<b>Hauríem tingut crisi, amb robots?</b>	<b>343</b>
<b>Com hem de fer els robots?</b>	<b>353</b>
<b>Calaix de sastre, en tres apartats</b>	<b>365</b>
<b>Persones i màquines, i quin futur ens espera</b>	<b>373</b>
<b>Els robots, els humans i l'Internet</b>	<b>383</b>
<b>És l'estupidesa humana exportable a les màquines?</b>	<b>397</b>
<b>Màquines de Turing i aturats de llarga durada</b>	<b>407</b>



## *Realment, ens ha d'interessar, que hi hagi intel·ligència artificial?*

**N**o cal ser Roger Penrose ni haver escrit articles amb Stephen Hawking per haver vist o intuït que una de les característiques més clares de la intel·ligència humana és la creativitat. Volem fer màquines intel·ligents? Doncs caldrà que siguin creatives. Oi?

Anem a pams. Què és la creativitat? Qualsevol enze ens dirà que aquesta és la qualitat dels genis, o com a mínim dels que passen per originals, d'aquells que han pensat coses que ningú mai abans no havia gosat pensar, de l'ou (o ous) de Colom. Però aquest mateix enze ens dirà que això no es pot prefabricar de cap manera, que la creativitat és indomesticable: no es deixa dominar ni es fa preveure, no s'educa pas ni es planifica. La creativitat ens agafa sempre per sorpresa, apareix quan menys se l'espera, i ens irrita amb obvietats que sempre havíem tingut davant del nas i que mai no havíem sabut veure, o ens indigna amb difícils novetats incomprensibles que, si mai les arribem a entendre, ens mortifiquen per la seva simplicitat i perquè mostren indissimuladament la nostra reconsagrada estupidesa (la síndrome de l'ou de Colom!). Una cosa no es pot negar pas, però: la creativitat és la marca de l'ésser intel·ligent. I com més creativitat, més intel·ligència. Això sembla clar.

Així doncs, si volem fer éssers realment intel·ligents, caldrà que els fem creatius. Ah, però hi ha un problema!: la creativitat és inexplicable –i, per tant, segurament inconstruïble: no en tenim els plànols (de fet, no en tenim ni idea). No és pas, només, que no sapiguem quan ni com ni d'on surt; és que ni tan sols quan

BUTLLETÍ DE  
L'ACIA  
NÚMERO 1  
TARDOR DE 1994

ja ha sortit ens podem explicar –ni amb la nostra millor voluntat– com carat ha passat, ni per què. I no cal tornar a ser cap Penrose per sospitar que potser la creativitat és una d'aquelles coses que Gödel s'inventava per fer-nos patir i que ell batejava religiosament dient-ne «sistemes no recursivament axiomatitzables», és a dir, planerament, coses senzillament «indescrïptibles» o, potser, inaprehensibles... en fi, tot el contrari d'això que tenim en el confortable i ordenat món dels algorismes.

El problema dels creatius és que no són gens educables ni planificables ni pensables ni comprensibles... en un mot: són incontrolables. Això torna a semblar clar. Ara bé, per què volem saber matemàtiques i física, practicar l'enginyeria i construir sistemes si no és perquè ens ajudin a ordenar i controlar una miqueta més el nostre insegur, impredecïble i perillós món que ens envolta? Si hi ha res que exigim de la tecnologia és precisament la fiabilitat i el control i, d'altra banda, no fer-ho així fóra d'allò més absurd: s'imagina algú una màquina que en lloc d'ajudar-nos es dediqués a mortificar-nos de tant en tant o, pitjor, a fer-nos algun mal de manera aleatòria? Les màquines que hem fet sempre han estat controlables, per definició.

Doncs ja ho tenim: mai no podrem fer màquines creatives. Potser sí que les farem intel·ligents. Això sí. I, a més, les hi farem tant com voldrem, però el que mai no farem és passar d'aquell punt just en què comença a haver-hi un bri de creativitat. Per què? Doncs perquè mai ningú no pretindrà fer una cosa necessàriament controlable (una màquina) que sigui gens incontrolable. Això, que per a un filòsof és una *contradictio in terminis*, per a un enginyer és, senzillament, l'acomiadament de la feina. A part que, si bé potser ens plauria la «companyia» d'una màquina creativa (com podríem entendre també que ens plagués el difícil i fàcilment incòmode veïnatge d'un geni aproximadament insuportable), els inconvenients que tindria quan se'ns descontrolés probablement farien que la fésim malbé aviat amb ganes o, pel cap baix, que l'odiéssim amb tota la nostra cordialitat. I si aguantem més o menys pacientment que els humans que hem de suportar al nostre costat es descontrolin tots sols, voleu dir que això ho aguantaríem en màquines que hem construït (i, pitjor, pagat!) nosaltres mateixos?

Com deia una nena de deu anys a Hans Moravec, que li pregun-

tava si li agradaria tenir un robotet amic que li fes companyia, els deures i tota la resta: «no em puc imaginar res de més horrible!» I podia haver afegit que per a això ja tenia amiguets de carn i ossos, i que si aquests es posaven pesats sempre quedava l'ós de pelfa, que s'ho deixa fer tot sense piular. O bé, com deia un dels germans Marx a un pidolaire que li demanava diners per suportar la seva família: «Sí, home! com si jo no en tingués prou de suportar la meua!» Probablement els informàtics hi posarem algun dia seny i ens passarà definitivament aquesta dèria obsessiva, malaltissa i recurrent de pensar que volem fer màquines «realment intel·ligents» (és a dir, creatives) i ens adonarem que tal cosa, a més de lògicament inviable, finitament inexplicable i recursivament impossible és totalment repulsiva. Perquè, per a bojos, ja en tenim prou amb els coneguts (que, si més no, tenen l'avantatge que, essent humans com són, de tant en tant es cansen). I si no ens passa aquesta ceba, potser caldrà que demanem a la gent que ens perdoni perquè, com va dir Jesucrist, no sabem pas què carai fem.

Barcelona, 25 d'octubre del 1994





## Quin és el nucli dur de la intel·ligència?

BUTLLETÍ DE  
L'ACIA  
NÚMERO 2  
HIVERN DE 1994

QUAN a l'any 1960 el matemàtic John McCarthy feia una prospecció dels camps que havia d'estudiar la seva «intel·ligència artificial», veia claríssim que primer es reconeixerien formes (o cares), com fan els nens sense problemes, i que s'acabaria demostrant teoremes, que és la feina, si fa no fa, dels matemàtics. La implicació subjacent era que, ja que tots els matemàtics han començat essent infants (això de vegades és cert, no us cregueu), les coses que fan els matemàtics són una extensió –monòtona– de les coses infantils; ergo el *theorem proving* és una activitat superior i subsumidora, per elevació i excelsitud, de la merament recongnoscitiva (el *pattern recognition*), que és elemental per definició ja que –atenció, que això és circular– fins i tot un nano s'hi veu amb cor. El resultat d'aquesta presumpció ja el sabem: el desmentiment total, històric. Ara que tenim presa la mida a la *teorèmica* (horrible mot que, com fan sovint, van inventar els francesos), el més calent és a l'aigüera pel que fa a allò tan «senzill» de reconèixer cares.

Doncs una cosa semblant ens està passant, crec, pel que fa a les essències últimes de la *intel·ligència* (la intel·ligència lògica, vull dir). Si un lògic volgués explicar com raonem tots plegats, probablement començaria recitant la Lògica «clàssica», continuaria per les clàssiques «lògiques no-clàssiques» i acabaria parlant de *no-monotonia*, més endavant de *probabilitat* i, finalment, de *creences*, com a extensions formals consecutivament finals i supremes, d'una complexitat i dificultat d'implementació forassenyades. Doncs bé, em fa l'efecte que la realitat, com en altres ocasions, és justament la contrària; que per trobar l'ordre adequat de penetració i comprensió cal capgirar la seqüència. O això sembla, pel que darrerament

estan repetint alguns antropòlegs (si més no, els anglesos aproximadament bromistes de *The thinking primate's guide to deception* i els de la colla de Cosmides, de San Diego).

Quan es passa, com en un zoo, dels titís als ximpanzés sorgeix un fet que marca el començament de la *intelligència* com nosaltres l'entendem. Ni més ni menys que la *mentida*, la capacitat d'enganyar. Els ximpanzés tenen ja una autèntica «teoria de la ment» basada en la possibilitat operativa de traslladar-se mentalment a la posició d'un altre individu per simular què es veu o se sent des d'allà; la qual cosa, com és evident, permet, tot tenint en compte el punt de vista d'altri, adequar la pròpia conducta per sortir-se'n amb èxit sense ser interferit. Exemples? Tants com en vulgueu: el ximpanzé que s'amaga darrere d'un pedrot per menjar-se el plàtan tranquil perquè sap que, d'allà estant, el col·lega que el mira no el pot veure; o aquell altre que fa expressament un crit d'esglai perquè els altres fugin corrents de l'hipotètic perill i no vegin la fruita que ell acaba de trobar. Així, la nostra abstracta «teoria de la ment» que bastim sobre les **creences**, les pròpies i les dels altres *agents* que actuen amb nosaltres, que per un lògic d'ofici és la culminació climàctica de la seva excelsa ciència, per un ximpanzé és, simplement, el punt de partida, concret i pràctic. (I, per cert, qui ens havia de dir, després de tants segles i capes de moralitat, transcendència i solemnitat científica, que la *intelligència* es basava en la *mentida* i en l'exercici professional i perfeccionista de l'engany i del no deixar-se engiponar? Després del *dictum* lapidari de Turing segons el qual no hi ha *intelligència* si no hi ha fracàs, només ens faltava aquesta: la sublim *intelligència*, filla bastarda del *fracàs* i la *comèdia*!)

I això no és tot. Si resulta que les creences, aquesta última assignatura de la Lògica, potser n'haurien de ser la *prima ratio*, el primer pas en la seva comprensió, també resulta que la **probabilitat**, aquella bèstia negra estudiada com a penúltima matèria (perquè no n'hi ha de postúltima) de la Lògica, segurament és l'ingredient de la *intelligència* que va en el segon lloc. En efecte, sabeu quins raonaments han demostrat els experiments que ens resulten més eficients i naturals, aquells en què ens equivoquem menys? Doncs, aclaparadorament, tots aquells en què intervenen –mal que sigui llunyanament– bescanvis: *do ut des*, «jo et grato, tu em grates», «jo te'n dono ara, ja m'ho tornaràs després», «ara faig això perquè si

feia una altra cosa probablement no me'n sortiria», etc. I, tot fent pactes imaginaris, anem calculant probabilitats; molt, sovint, i bé. Constantment fem càlculs sobre coses sotmeses a l'atzar: «si faig això, què passarà? doncs ho faig, perquè llavors passarà això altre i allò de més enllà», etcètera. No calculem exactament nombres, però les nostres preferències, un cop traduïdes a xifres, són molt precises i realistes. Vistos els articles de Cosmides i els sempre interessants de Tversky, això sembla comprovat, plausible i fins i tot fàcil d'entendre.

O sigui, que la intel·ligència, com l'avar de Plaute o de Molière, és tot *engany* i *càlcul*. La qual cosa potser no ens hauria d'estranyar pas tant. Al capdavall, des de fa dos-cents mil anys, si més no, som animals socials de mena molt cooperativa, i això marca. Que sapiguem enganyar els altres –com carai quedar-nos tranquils, si no?– i avaluar, en termes de pactes i de guanys, què ens pot portar el contacte amb el proïsme més o menys insuportable no deixa de ser una base bastant entenedora i natural de la intel·ligència. I no sols *natural*, sinó, com hem dit, també *lògica* (en sentit estricte). I encara més, després de les troballes recents dels antropòlegs diversos que hem citat, científicament *versemblant*. Què volem més? Ara, després de posar-nos a explicar la Lògica a l'inrevés, només ens queda engegar i aplicar-ho a les màquines. Primer, màquines d'enganyar (amb un test de Turing propi: com més ens enganyin, més intel·ligents –i més cofois nosaltres pares de la criatura). I, segon, màquines «calculadores» (en el sentit maquiavèl·lic del terme), és a dir, d'avaluar/contrastar les pròpies possibilitats: el Turing en aquest cas fóra l'èxit –mesurat en encerts–, com passa amb els escurabutxaques o els executius). Amb una mica de sort, com amb els humans, podríem arribar a tenir fins i tot alguna màquina solidària.

Doncs bé, a la feina. Perquè si, amb l'experiència que tenim enganyant-nos a nosaltres mateixos i recíprocament, no sabem programar màquines que ho facin, i si, anomenant-nos informàtics, no sabem fer «calculadors» de cap fred, i eficients, ja podem plegar. Ah! els teoremes ja vindran més tard, afegits; no cal pas amoïnar-s'hi.

Barcelona, 26 de febrer de 1995



## La importància de dialogar i de passar-se

**A** nosaltres (els *IA-aïres*, vull dir) sempre ens ha semblat que la intel·ligència era una cosa que, si ens deien d'inocular-la en els nostres ordinadors, l'havíem prèviament de *construir* com els enginyers saben fer, és a dir, com si fos un *producte*. Dissenyant-lo, preveient-ne totes les possibles utilitzacions imaginables, minimitzant-ne el cost i la complexitat i provant-lo per veure si la cosa rutlla o no. A tot això cal afegir-hi encara això altre: el resultat de la cosa no és avaluat segons res que sigui gaire objectiu (a ningú no se li ocorre pas com fer-ho), sinó que el seu *valor* és mesurat per la satisfacció del client i basat en la *sorpresa* que el producte li produeix. Ara bé: quina remota relació té això amb allò que no sabem definir però que sabem reconèixer tan bé i que anomenem *intel·ligència*?

La intel·ligència, més que un producte, sembla més aviat un *procés*; un procés que se sap com comença però –gens– com s'acaba (i si no, com diria Voltaire, és que no és intel·ligència). Si admetem –que és molt admetre, perquè sovint demana un acte de fe– que els homes ho som d'intel·ligents, no estaria pas de més estudiar el nostre cervell per veure què carai hi passa. Sortosament, Gazzaniga i el seu inspirador Sperry (metge, Nobel i traspasat recentment) ja ens ho han fet, i bé –i ens ho han explicat en moltes obres de divulgació. Resumint: el nostre cervell és, com la Trinitat o l'acudit dels quatre evangelistes, dues coses diferents: l'hemisferi dret i l'esquerre. Dit a l'engròs i tal com raja, el primer copsa totalitats (situacions, *atmosferes*, sensacions) però no les sap analitzar, mentre que el segon es fa un tip d'analitzar però no sap què. Això, que ja els passa, aproximadament, a les rates, a nosaltres se'ns sublima en el següent joc (que ens prenem la llicència per simplificar

BUTLLETÍ DE  
L'ACIA  
NÚMERO 3  
PRIMAVERA DE 1995

intolerablement i barroera): l'hemisferi esquerre hi fa una mena de personatge tipus *portera* (per entendre'ns), que parla i parla i ningú no sap què diu, ni a ningú interessa. Es tracta de sotmetre qualsevol cosa a l'escrutini de la paraula, fent-li donar voltes del dret i del revés sense guió ni sentit aparent; dit abusivament: l'esquerre desbarra. En canvi, l'hemisferi dret és molt sensible a l'exterior, i una mica totxo (no entén gens allò que sent tan intensament); representa el sentit de la realitat davant d'un esquerre que tot ho analitza sense adonar-se de si allò que diu té sentit real o es correspon amb l'experiència o és verificable.

Sembla mentida que una cosa tan aparentment impressionant i reconsagrada com això que en diem Intel·ligència (que demana majúscules i temor reverencial) es pugui reduir a un simple joc absurd entre l'Stan Laurel i l'Oliver Hardy corresponents, o entre el Quixot i el seu escuder (que potser són tan universals perquè ens recorden alguna cosa que tenim tots al cap). Però, ben mirat, no resulta pas tan estrany que la intel·ligència neixi d'aquest *joc* de conseqüències naturalment imprevisibles. I no només perquè aquesta teoria ha nascut de nombrosos casos clínics i s'ha pogut verificar clarament en situacions límit (lesions de l'àrea de Broca que converteixen el pacient en un xerraire incontinent sense el més mínim sentit de l'orientació sobre on cau la realitat o, més tràgicament, en infants autistes o en esquizofrènics), sinó perquè aquesta visió constructiva, plàstica, no-determinista i *dialogant* (*dialèctica*?) lliga bastant més que no sembla amb algunes idees que venen del caos (del *caos autoorganitzat*, pròpiament).

Se sap –si més no d'ençà el 1990– que la salut en general, i la intel·ligència en particular, exigeixen treballar en condicions de caos (en el sentit tècnic: caos determinista). Això vol dir que el nostre cervell és un conjunt complex d'unitats que treballen autònomament, el comportament conjunt de les quals neix (*emergeix*) dinàmicament per autoorganització i se situa meta-establement, *críticament*, a la frontera del caos. Com? Doncs com en tots els sistemes tècnicament *complexos*: el sistema està contínuament traspasant els límits del seu funcionament normal i tornant a la conca d'atracció perquè el conjunt de components del sistema s'hi resisteixen; així, si fa no fa. Una mica com funcionen els infants, que irresistiblement proven de fer-ho tot més enllà dels advertiments

aproximadament morals i de les conseqüències previsiblement nefastes; i és així, precisament, com aprenen a situar-se i centrar-se (i, de passada, a descentrar tot altri).

Tornant al xerriaire incontinent de l'esquerra i al sòmines sensible de la dreta, sembla que aquest mecanisme dual, ben conegut i acreditat, pot donar una bona explicació de com nosaltres fem el que fem: doncs, explorant-ho tot, discutint-ho tot, dubtant de tot i acabar fent qualsevol cosa, més o menys limitats per una realitat que no sempre respectem i per unes explicacions que no sempre ens acaben de convèncer (un romanent de les quals és, probablement, la culpabilitat que sovint ens acompanya). Un mecanisme dual com aquest és, a banda de plausible, suficient per crear el caos necessari per al nostre bon funcionament mental; funcionament mental que versemblantment consisteix, doncs, en el joc permanent, superficialment còmic, entre un component del nostre cervell que sent sense saber i un altre que raona sense sentir –desbarra, vaja– incansablement, o entre una part que no para de *passar-se* i una altra que hi posa seny (o, si més no, resistència o enuig o cansament). Segurament no es això el que ens pensàvem que era –o que havia de ser– la intel·ligència. Ens la imaginàvem més digna, més elegantment descriuible, i gens com aquest impresentable i surrealista diàleg entre beneits (i, a sobre, autistes).

Té tot això cap conseqüència per nosaltres els IA-aïres? Em sembla que sí, i bastant òbvia: si hem de muntar sistemes *intelligents*, val més que els anem fent duals ja des d'ara. I aquesta dualitat ha de tenir dos aspectes complementaris i ortogonals: d'una banda hi ha d'haver la parella de l'analític i el sintètic, el desbarrant i el realista, la mare permissiva i el pare frustrant, el policia bo i el dolent, si cal; i de l'altra, la parella del curiós i el conformista, el *passota* i el *punyetero*, l'entremaliat i el gendarme, l'indignador i l'indignat. I si no posem tot això a l'olla, tot serà segurament previsible i segur, però intel·ligent no gaire (i creatiu, gens).

Barcelona, 23 de juny de 1995





## Una qüestió de sentiment

**P**OTSER a algú li semblarà estrany que en una columna com aquesta parlem de sentiments. Sentiments, en màquines? Doncs sí. Després de segles desterrat, el sentiment està tornant a la ciència (i a la tecnologia), ara ja no com a interferència indesitjable o irrellevant, sinó com a objecte d'estudi i, qui sap, de disseny, usat deliberadament i amb contumàcia.

Anem a pams. Què són, exactament, els sentiments? A primer rebot, qualsevol sofert espectador de TV podria definir-los com secrecions procedents de vísceres diverses, de gran utilitat com a matèria primera de serials. Però no: de fet, tot i els epígons lacrimogènics del celebèrrim consultori de la Sra. Francis, els sentiments tenen un pedigrí evolutiu força presentable. Es tracta ni més ni menys que de mecanismes que preparen el cos perquè se situï en una determinada línia d'acció, en casos en què no hi ha una resposta innata que resolgui el problema. Com a mètode de control, resulta prou ràpid i efectiu, i per això l'evolució l'ha afavorit. I com a procés, resulta menys costós (i força més antic) que el complicat mecanisme lògicoinferencial, tan delicat i difícil de mantenir –i tan propens a fer-nos equivocar. A més, els sentiments són comunicables, transparents, normalitzats, inambigus i, per això, universalment compresos (només cal pensar en les cares que fem: tothom les entén sense haver de preguntar).

No se sap si els sentiments són o seran bons per a les nostres màquines (McCarthy diu que no, de cap manera), però si creiem que hem de deixar-les anar soles que s'espavilin pel món com fills emancipats, potser serà bo de plantejar-se quin repertori hi destinem. De fet, sembla constatar-se que, d'evolutivament bàsics, només n'hi ha cinc: (1) *felicitat*, (2) *tristesia*, (3) *ira*, (4) *por* i (5) *repug-*

BUTLLETÍ DE  
L'ACIA  
NÚMERO 5  
TARDOR DE 1996

*nància*, pentàleg que, com les taules que va baixar Charlton Heston del Sinaí, es redueix a dos (que podríem anomenar *pulsions-base*). En efecte, si deixem a part el cinquè, que sembla sobrevingut, els dos primers representen la capacitat de creació (i/o acceptació) (1) i d'anul·lació *rebuig* (2), respectivament, de relacions establertes entre individus, i els altres dos l'atac (3) o la fugida (4) en l'*encontre* topada amb rivals o depredadors. Si encara volem més reduccionisme, tot plegat es redueix al conegut joc del premi i el càstig, o de la pastanaga i el garrot (sona?). Deixant de banda les afiliacions metafísiques o religioses de cadascú (n'hi haurà que diran: «com cal!», «com Déu mana» o «és clar!»), la qüestió de si aquest mecanisme ens interessa o és útil (imprescindible?) per a les intel·ligents i filials joguines que ens entestem a construir queda oberta a partir d'ara, i tothom s'hi podrà esplaiar de gust.

Hi posem o no hi posem sentiment, una cosa ressalta de seguida: els sentiments estan fortament relacionats amb la percepció, i aquesta amb l'activitat cognitiva. Ara bé, aquesta última és –per dir-ne alguna cosa– simbòlica, calculatòria; i els sentiments, siguin el que siguin, no. Són un mecanisme motor, originador, motivador, preparador, el que sigui, però no pas del tipus cognitiu/simbòlic que ens resulta familiar. Els imaginatius ja poden anar pensant com carai es deu programar una cosa així en Lisp.

Si mai ens en sortim, a crear l'esfera emocional i tenir-la ben separada de la cognitiva (és a dir, del *pensament* pròpiament dit), potser algun dia ens podrem arribar a plantejar activitats cognitives complexes i realistes, que no siguin merament la mena de deducció/càlcul/combinatòria aproximadament simplificatòria/simplista amb què programem les màquines avui i que fem passar per *raonament*. Activitats tals com (bufa!) la *imaginació* (activitat cognitiva sense objectiu conegut), el *càlcul* –i la *deducció*– (això no ens ho cal explicar: és potser l'únic que tenim clar) i d'altres activitats majoritàriament gratificants (i no-deterministes) com ara la *inducció* (una mena de deducció pecaminosa caracteritzada per l'augment de la informació semàntica –i també del risc d'equivocar-se) i (agafeu-vos!) la *creació* (o creativitat), que és com la inducció però sense cap objectiu concret identificable.

Com es conjumina tot això? Doncs, en la IA, fer que el mecanisme *sentimentari* (*sentimental*?) engrani bé amb el cognitiu per

potenciar-lo i fer que el conjunt estigui més ben adaptat evolutivament (que se'n surti més bé de tot, vaja) és una de les feines (que no són pas poques) que ens queden. I si algun mascle vocacional troba que això de programar sentiments és poc seriós o fa marieta, doncs que ho deixi, que així tindrem més feina els qui quedem. No sé si la IA que en resultarà serà més divertida que la d'ara, però *sentida* segur.



## *Cal tenir consciència*

**M**OLTS que no estan al dia del que passa en la IA es farien creus de saber que una de les qüestions més rabents aquests mesos és la de decidir si les nostres màquines presumptament intel·ligents han de tenir o no «consciència». Alguns, si ho sabessin, ens preguntarien, francament, si ja ens hem tornat bojós i d'altres, més discrets, ens demanarien si ara la IA és una branca de la Teologia. Ara bé, tot i la frivolitat aparent de la qüestió, la cosa no és per deixar-la córrer sense més ni més. I és que la pregunta no és gens idiota: els hem de posar «consciència» a les màquines que fem?... perquè, per cert, què és la consciència?

Com que no tenim gaire espai, no explicarem què és. (També és veritat que no en tenim ni idea, com tothom). Una cosa sí que podem avançar, però: de la consciència se'n poden dir, si més no, aquestes tres coses:

1. La consciència sembla estar lligada al «sentiment» (sentiment en el sentit de **sentir**, de tenir sensacions) més que no pas a cap percepció concreta ni a cap «pensament» racional. (Així, quan ens anestesien perdem consciència.)
2. La consciència sembla estar lligada al fet de tenir **records**; de fet, sembla ser-ne la precondition: quan passa la cosa que cal recordar, cal estar plenament conscient: si no, no funciona, i el record no es reté (i el record de llarga durada és, no ho oblideu, una condició bàsica per a la supervivència de l'«animal»). Així, la consciència fa viu, «recordable», allò que en un altre moment potser ens caldrà recordar urgentment (per comparar-ho amb la realitat del moment i fer el que calgui).

BUTLLETÍ DE  
L'ACIA  
NÚMERO 6  
HIVERN DE 1995

3. La consciència sembla organitzar totes aquestes sensacions en un «**mapa**» del nostre cos, dels seus límits físics i de la seva situació espacial respecte a tot el que l'envolta. Amb aquest mapa el cervell (o el mecanisme de control de la màquina) resumeix d'una manera molt eficient –«visual», diríem– tota l'aclaparadora informació que li arriba.

I què se'n treu, de tot això? Doncs, segons qui, la conclusió és que aquest mecanisme ha estat afavorit per l'evolució per raons pràctiques evidents i que potser fóra bo que l'incorporéssim a les nostres màquines. I alguns (McCarthy i un nombre creixent d'IA-aïres) creuen que no sols és recomanable sinó *imprescindible*: la intel·ligència *exigeix* la consciència com a precondició; si no hi ha això, no hi haurà allò.

I què n'hem de pensar nosaltres, pobres mortals, que no som en McCarthy ni tan sols déus de l'Olimp? Doncs, de moment, no se sap. Jo suggereixo, però, de llegir i pair mentrestant el que diuen dos neuròlegs sensats: l'italoamericà Gazzaniga i el lusoamericà Damásio. El primer ja fa temps que estudia la lateralitat cerebral i ens ha explicat manta vegada que l'hemisferi esquerre, fabricant incontinent d'hipòtesis, és sistemàticament posat a ratlla pel dret, que n'adapta les paranoies a la realitat possible. La dualitat incessant entre aquests dos mecanismes, l'un generador i l'altre retallador, és el que permet i fomenta la intel·ligència. La *consciència*, en aquest joc, és la resultant del poder fabulador i referencial de la nostra meitat esquerra un cop la dreta l'assumeix globalment, «emocionalment», i la converteix en subjecte (el *jo*).

Damásio, pel seu compte, ens explica com el pensament i el sentiment van sempre neurològicament acoblats, i que el primer és el *regulador* del segon, mentre que el segon és el *motor* indispensable del primer. Per això és tan artificial, si bé a base de repetir-ho se'ns ha fet intuïtiu, considerar, com ha fet tota la Filosofia occidental d'ençà de Plató, que les dues coses són independents i van separades. I si tot això, com sembla, és un error –tot i que sigui històric i universalment assumit– potser caldria anar canviant de perspectiva i ja no pretendre més fer sistemes amb intel·ligència però *sense vivències*, perquè això, que no pot passar amb els humans, tampoc no deu poder passar amb les màquines. I si no pot ser, és que és impossible. Potser fóra millor, doncs, no encaparrar-s'hi més i

passar a construir sistemes «emotivo-vivencials», o alguna cosa que s'hi assembli, en els quals la **intelligència** sigui allò que sempre ha estat i no un «invent» fet per satisfer les nostres fantasies filosòfiques (i que, a sobre, no funciona). I on la famosa **consciència** en sigui el «resultat» natural, no pas un «producte» de disseny.

Així doncs, per trempar els ànims, el que proposo és llegir, pensar i pair. I, en acabat, ja tornarem a plantejar-nos aquest tema de la consciència. Un altre dia. No haurà de ser gaire tard, però, perquè ja fa massa temps que parlem de màquines intel·ligents. I, d'intelligència, francament, no se'ls n'hi veu gaire. Només, com el «valor» als soldats que fan la mili, «se'ls n'hi suposa».





## *Doncs, què és la intel·ligència?*

**D**E la intel·ligència es podria dir, com Churchill deia de la Unió Soviètica (q.e.p.d.), que és «un misteri, embolicat en un enigma». O sigui, que no en sabem res. Algunes bajanades frívoles sí que se'n poden dir, i sense anar més lluny jo n'he dit algunes aquí mateix, amb els següents corollaris precipitats:

1. No podem esperar de màquines dissenyades per nosaltres cap mena de creativitat, i això més aviat per definició, perquè la creativitat implica la possibilitat del descontrol, cosa inadmissible en una màquina.
2. La intel·ligència real sembla implicar el joc entre dos pols: un de generatiu i un altre de retallador.
3. És possible que a una intel·ligència mecànica també li calgui allò que sembla que ens cal a nosaltres: impulsos primaris que actuïn com a motor (en el nostre cas, el cos i els sentiments) i una mena de mapa global que unifiqui tots els processos en una entitat emergent identificable (en el nostre cas, la consciència d'individu).

Ara. Què més es deu poder dir de la intel·ligència en general? Aquí hi ha un problema, i és que nosaltres som els únics que en tenim, i quan apliquem la paraula intel·ligència a algú altre només ho fem per extensió i analògicament, perquè no tenim cap altre model que ens il·lustri sobre la propietat en qüestió. Seguint per aquest camí, algunes coses salten a la vista. Cada cop és més evident que la nostra intel·ligència és una prolongació natural, no traumàtica, de la dels mamífers. Molt fortament adaptada, això sí,

BUTLLETÍ DE  
L'ACIA  
NÚMERO 7  
PRIMAVERA DE 1996

a la vida del paleolític inferior, és a dir, a la caça-recollecció dels nostres avantpassats. Molt fortament, i molt sorprenentment, de tal manera que la nostra intel·ligència sembla feta a mida per a aquesta relativament curta etapa de la humanitat (en què, per cert, també hem estrenat el llenguatge). O això és el que es desprèn d'algunes observacions recents: la intel·ligència humana conté un mecanisme *avaluador* de situacions i de resultats probables molt afinat –òptim per a situacions d'incertesa forta com ara la caça–, i un mecanisme *compensatori*, en termes de l'«avui per tu demà per mi», que és solidari i perfecte per perpetuar el conjunt de l'espècie i equilibrar l'esperança matemàtica al llarg del temps. Això es veu clarament en experiments, perquè és en aquests dos aspectes on fallem menys i funcionem més bé (i hi tendim més sovint i consistentment).

Tot això fa suposar que la intel·ligència humana és una cosa molt concreta, optimitzada i contextualitzada, i alhora fa sospitar que si n'hem de reproduir en una màquina algunes característiques probablement el més sensat és no reproduir-la tota tal com és per a nosaltres. Especialment, perquè la màquina que volem fer no deu haver pas de ser caçadora ni recollectora, que per això avui no ens calen màquines intel·ligents, ni viure en societat.

La intel·ligència humana es basa, s'ha basat sempre, en la generalització de l'associació de fenòmens (el que diem causalitat) i en la magnificació, sovint abusiva, de vivències que es considera que s'han d'evitar o repetir, segons quins siguin els seus efectes percebuts. De fet, la nostra conducta es forma i consolida gràcies a l'*asimetria* entre el passat i el futur. El passat, el tenim a l'abast, per examinar-lo –obsessivament si cal– i fer-lo servir de guia per repetir-lo o evitar-lo. Del futur, en canvi, no en sabem res, i això, que sol angoixar-nos moltíssim, ens terroritza tant que per saber-ne alguna cosa i posar-nos-hi a favor o en contra som capaços de recórrer a qualsevol cosa. El nostre estadi de caçadors-recollectors ens ha llegat l'obsessió, que avui constatem com a component principal de la nostra intel·ligència, consistent a (1) *repetir* –o evitar– allò del qual ja coneixem les conseqüències, (2) *preveure* el futur mitjançant l'avaluació més o menys probabilística de les situacions possibles, i (3) practicar de manera generalitzada la *compensació* com a mecanisme d'assegurança contra imponderables. En tot cas,

la nostra intel·ligència és com és perquè ha tingut èxit i perquè ha permès superar brillantment l'etapa (el paleolític) per a la qual ha estat «dissenyada».

Deixant de banda la qüestió de si la nostra intel·ligència és adequada per a l'època que vivim, tan diferent del paleolític, ara la pregunta és la següent: si la clau de l'èxit d'una criatura intel·ligent és la bona adaptació al seu entorn mitjançant la correcta predicció del futur (per adaptar-s'hi bé), aleshores quina d'aquestes estratègies es pot considerar intel·ligent?:

1. Repetir una conducta que ha tingut èxit en el passat
2. Matisar, i canviar si cal, la conducta anterior si es veu que ja no dóna resultats (els humans això ho fem reelaborant la percepció causal i les avaluacions passades: se'n diu experiència, d'això, i el cas extrem n'és el procés molt depurat que anomenem Ciència).

Que la primera estratègia no és gaire intel·ligent sembla obvi. Però és la que nosaltres inculquem als nostres nens i als nostres ordinadors en forma de regles («si passa això, fes allò»). Pot ser la manera de resoldre problemes en situacions estacionàries (i ho és en societats on no es toleren els canvis). Si les màquines que dissenyem i volem intel·ligents han de treballar en aquestes condicions, no cal patir-hi més: ja les tenim. Aquestes màquines, per cert, responen al conegut model de la màquina de Turing, una idealització convenient però massa paradisiàca per ser de veritat. No sols redueixen –de manera bastant forçada i antinatural– qualsevol activitat intel·ligent a un «càlcul» (en un sentit generalitzat) sinó que entenen aquesta activitat calculatòria d'una manera completament diferent de com ho fa la natura i de com ho fem nosaltres: l'entenen, i s'hi comporten, com si fossin matemàtics (tot ho han de tenir preespecificat en un llenguatge inambigu i finitàriament executable) i, a sobre, «buròcrates» (hi ha un univers perfecte configurat per regles, i els casos ambigus o fronterers són, com a mínim, una nosa). Un model de màquines com aquest, una mena de matemàtics que funcionarien, no ens convé gaire, especialment si la feina demana realment intel·ligència: és poc realista, poc adaptable, intolerant a l'error, gens associatiu, pròpiament

només local i analític, no pot evolucionar ni autoorganitzar-se i demana massa preparació prèvia d'un programador que sigui –ell sí– intel·ligent. En situacions estacionàries, estables o relativament controlades, res a dir. Ara, si hem d'immergir les màquines en situacions canviantes –que són precisament les que demanen més intel·ligència– aleshores fer-nos-hi forts deu ser probablement una de les decisions més poc intel·ligents, valgui la broma, que podem prendre.

I en casos que realment demanin el dinamisme de la intel·ligència quan s'enfronta a un futur pròpiament incert, què fem? Hi posem una màquina que ens imiti, ja que nosaltres som el model de tota intel·ligència realista imaginable? Doncs potser sí, però sense fer-ne un gra massa. Tampoc no ens convenen, per absurdes, màquines que repeteixin el nostre disseny paleolític: actualment cap màquina (ni nosaltres tampoc) no ha d'afrontar pas el futur per la via de la magnificació de vivències, l'anul·lació de l'angoixa i el joc de les compensacions (sovint fallaç, penseu que és la font de moltes de les nostres percepcions enganyoses). Perquè, com deia Valéry, el futur ja no és el que era: abans podia ser magre però avui, pel que fa a la supervivència física, ja no presenta problemes d'intel·ligència sinó d'organització econòmica.

Descartades, doncs, les màquines que volent-nos imitar a nosaltres acabin imitant el paleolític, i també les que imiten allò que diu que va dir Turing com si fos una recepta o les Taules de la Llei (què n'hauria dit ell, que hi era tan sensible, a la perversió de les seves intel·ligents i matisades consideracions?), què queda? Doncs allò que caldria esperar de nosaltres si no fóssim dogmàtics fins a la negació de la naturalitat: les nostres màquines –paleolítiques o turinguianes, humanoides o no– seran intel·ligents en la mesura que –com tot el que ha estat intel·ligent abans que elles– simulin i prevegin el futur i s'hi adaptin amb èxit, incloent-hi les rectificacions que calguin. Segurament ho faran només en aquells aspectes parcials per als quals estaran dissenyades i d'una manera que no s'assemblarà necessàriament gaire o gens a com ho faríem nosaltres.

De fet, i sense fressa, aquestes màquines ja les anem fent. En voleu un exemple? Qualsevol ordinador de reserva de vols (per cert, un dels primers exemples d'aplicació integrada i en temps real)

és, indubtablement intel·ligent (*ningú* que no s'hi hagués acostumat podria negar-ho), actua com nosaltres (calcula, anota, preveu) però sense les nostres peculiaritats (lentitud i errors, inexactituds, poca atenció, etc.) i s'adapta bé al seu entorn sempre que no canviï gaire. Que és una màquina de Turing? Sí, d'acord, però no en té cap obligació, simplement ens ha convingut fer-lo així o potser no en sabíem més. A un sistema com aquest ja ens hi hem acostumat, ara ja no el trobem particularment «intel·ligent», fa la seva feina amb èxit, impressiona –francament, sí– pel que fa, no presenta gaire «creativitat» (tampoc no se n'hi demana) i, encara més, permet fer allò que per la seva complexitat fóra impossible o impensable sense ell. Volem res més?

Previsiblement, els sistemes intel·ligents que anirem fent seran com aquest: no s'assemblaran a nosaltres, seran indubtablement intel·ligents –tal com ho jutjaríem ara (però els seus usuaris ja s'hi hauran acostumat i ja no els ho semblarà)– i faran bé la feina per a la qual hauran estat dissenyats, que és l'adaptar-se a un futur més o menys imprevisible. I, mentrestant, ens anirà passant la dèria aquesta que tenim de saber i prefigurar què i com pot ser una màquina intel·ligent, i també la falòrnia de fer-la ser com diuen els llibres de Teoria de la Calculabilitat (que és, no ho oblidem, una formalització i una hipòtesi, no pas un fet experimental). Ens passarà aquest defici, com ens ha passat l'astorament dels coetanis davant de la impremta (els semblava incompreensible que «s'escrivís» un llibre sense que l'escrivís ningú concret) o davant les màquines mecàniques que multipliquen per mil la nostra força i els nostres gestos (des de les excavadores fins als robots) o davant de les pantalles màgiques dels ordinadors. Ens hi hem acostumat i, que se sàpiga, ningú no s'ha mort del trauma, i la nostra cultura ho ha superat sense gaire trasbals.

Com va dir Clarke, en una cèlebre frase, «qualsevol tecnologia prou avançada és indistingible de la màgia», a la qual cosa es pot afegir «però un cop ens hi hem habituat ens sembla la cosa més natural del món (i, per tant, ja no ens sembla que sigui màgia)». És de preveure, semblantment, que un cop haurem vist les noves màquines intel·ligents funcionant ja no ens ho semblaran, d'intel·ligents, deixarem de tenir aquesta por terrible de ser suplantats per màquines i tornarem a descobrir en la nostra intel·ligència

noves qualitats típicament humanes. És el que ha passat sempre. Ah, i també ens preocuparem de si aquestes qualitats humanes són potencialment assumibles algun dia per una (nova) màquina «realment intel·ligent». Si no patíssim, no ens distrauríem. És l'essència de la cultura.

Barcelona, 13 de juny del 1996

## *Racionalitat = Raonament + Emoció*

BUTLLETÍ DE  
L'ACIA  
NÚMERO 8  
ESTIU DE 1996

**E**L **raonament**, lògic o no, és un procés de decisió que precedeix i selecciona l'acció segons la situació viscuda i com a resposta a algun problema que se'ns presenta. Algunes opcions demanen rapidesa i, per tant, un cert automatisme (exemple: com esquivar un objecte que cau), d'altres són més a mig terme (què fer quan es té gana) o d'altres, encara, són francament secundàries i diferibles (què fer sobre votar, perdonar, casar-se, etc.) La decisió va precedida per una anàlisi tipus cost/benefici (que popularment qualificaríem de «càlcul» o de raonament «lògic», amb èmfasi en «lògic»). Però la **racionalitat** no és sols aquest acte de presa de decisions que anomenem raonament sinó el conjunt de processos i arquitectures que l'envolten i el mouen. És la intersecció de raonament i emoció. Físicament, és una funció localitzada en un punt del còrtex prefrontal, darrere dels ulls, que és on es regula el nostre comportament decisonal i «social», on se sospesen i prenen les decisions.

El joc complex que precedeix i agombola el raonament es basa en els anomenats avisos del cos (els *somatic markers* de Damásio). Se'ns presenten com «imatges» que equivalen a alarmes i que tenen com a efecte pràctic eliminar certes possibilitats d'acció i així augmentar la precisió en la decisió. Es presenten com a això, «avisos» (tipus «perill!» o «ves-hi!»), i inclouen classificacions (exemple: «bo», «dolent») de contingut inicialment instintiu però influïdes per la cultura. Si d'això en diem «emocions» i resumim tot l'anterior dient que les emocions són el motor de la racionalitat, aleshores no fem altra cosa que repetir Damásio (vegeu el seu *Descartes' error*) i d'altres autors que hi convergeixen (com ara De Sousa, Sutherland, Flanagan, Gazzaniga, Johnson o Searle). De moment, això



vol dir una cosa: l'home no té un cervell desconnectable (no el podem ficar en una banyera, vaja). Però no és sols això. Ja fa un temps que sabem, i cada cop ho sabem més, que una cosa típica nostra és que tenim un cervell capaç de curtcircuitar la seva pròpia alimentació sensorial. Els records i les «imatges» desats al còrtex són reintroduïbles al circuit límbic (el que rep les sensacions i crea l'emoció) com si aquella imatge s'hagués vist ara realment. Així podem recordar i, encara millor, simular i preveure accions abans que sigui el cas d'haver d'actuar. «Pensar» és, bàsicament, això.

Que la intel·ligència, almenys la nostra, necessita un cos, és fora de dubte. Això potser ens indica, en primera aproximació, que un robot està més pròxim a la intel·ligència que un ordinador de sobretaula: si més no té un «cos», i això, aquí, és molt important.

No cal dir que el llenguatge és un reforçador potent del sistema d'avisos corporals. A més, aquests s'organitzen en zones confluents del còrtex i hi donen lloc a la categorització de l'experiència passada, a la classificació de contingències de la vida real i a les representacions disposicionals que determinaran la futura estratègia personal de decisions. Aquí pot passar que diferents avisos, normalment adaptatius, s'interfereixin els uns amb els altres (per exemple, una acció vista com a bona pot ser anul·lada per una altra visió –un altre avís somàtic– com ara la necessitat de preservar l'autoestima, o l'obediència, la seguretat, el conformisme social, etc.), la qual cosa explica força bé algunes de les típiques «irracionalitats» que cometem i que tan bé han estudiat els clàssics Kahneman i Tversky. També pot passar, i passa, que

- alguns d'aquests avisos somàtics es tornen automàtics: això evidentment n'augmenta el rendiment (però pot passar que la cosa acabi resultant contraproductiu);
- algunes de les «imatges» creades en el procés s'internalitzen sense que en tinguem cap rastre de consciència: tenim aleshores allò que en diem intuïció, que és tan eficient com (i més ràpid que) altres mecanismes d'avís que, aquests sí, participen explícitament en el joc que anomenem *pensament*;
- alguns altres avisos somàtics sí que són percebuts conscientment, però no com a «sentiments» (no entren en el circuit límbic) sinó que esdevenen allò que en diem pensament abstracte.

En resum, el pensament –o la racionalitat, o la **intelligència**– és un joc complex d'elements relativament simples que necessiten un annex corporal com a font constant de sensacions, estímuls i motivació (constant, sí: se sap que n'hi ha prou que algú passi 48 hores sense sentir-hi perquè la seva activitat cerebral degeneri i es dissolgui totalment). La cosa es podria resumir, si estiguéssim de broma, dient que el cap fa «el que li demana el cos». Això sembla, cada vegada més.

Ara ve la pregunta: si això és la intelligència, té gaire sentit que fem màquines que mai no han tingut ni tindran un cos mitjançant el qual intercomunicar-se i interactuar amb la realitat i posar-ne experiència, estímuls i coneixement? Fins a quin punt una màquina desconnectada del seu entorn pot simular que és intel·ligent si l'única font de la seva activitat és allò que nosaltres, com a medi interposat, li hem dit que *és* –o que passa per– la *realitat* (una realitat, per cert, a la qual li impedim accedir)? Fins a quin punt allò que aquesta màquina «pensarà» no és més que un subconjunt esquifit i desnaturalitzat –«desarrelat», pròpiament– del que nosaltres podríem haver pensat o hem pensat ja? Perquè si és aquest el cas, allò que estem fent amb les nostres màquines intel·ligents és merament deixar que actuïn en condicions molt precàries sobre parcel·les molt delimitades. Si així i tot podem fer que el que «pensin» sigui intel·ligent, benvingudes aquestes màquines (de fet, són les que tenim). Ara, si d'això n'hem de dir immodestament intelligència, per molt que ens afanyem a afegir-hi artificial, potser que pleguem. Potser que reconeguem que no volem, realment, fer màquines intel·ligents. Que, a tot estirar, només volem fer, mandrosos com som, màquines que ens supleixin a l'hora de treballar molt, sense equivocar-se, en parcel·les en què nosaltres no resultem gaire brillants (és això precisament el que hem fet amb totes les altres màquines, des dels motors fins a les calculadores). I que acceptem que d'intel·ligents (de la mena autèntica) ja ho som nosaltres –és un suposar–. I que per a això no ens cal més companyia.

Ara, si del que es tracta és d'anar fent com fins ara la comèdia de fer veure que fem realment màquines intel·ligents sense que ningú de nosaltres sàpiga què carall és aquesta cosa, ni per a què les volem, per a això ja teníem abans les tertúlies dels cafès, que

eren una manera de perdre el temps molt més humanitzada que aquesta. Potser no s'hi guanyaven uns grans sous, això no, però s'hi vivia molt més tranquil, i un s'hi feia una cultureta. I hom acabava sabent parlar -i pensar- de forma, si no intel·ligent, almenys intel·ligible.

Barcelona, 14 de juliol del 1996

# Lògica, Informàtica i Intel·ligència artificial

**L**A Intel·ligència Artificial *clàssica* actual (la «IA») és una filla de la Informàtica del 1960, aquesta és filla de la Lògica del 1930, i aquesta de la Matemàtica del 1900. És natural, doncs, que cada una hagi heretat les preocupacions i manies de les precedents. I com que la Intel·ligència Artificial és la que ve última, sembla inevitable que passi l'escombra de tot, i es carregui de tics. Però ni la IA ve només d'aquests pares ni la seva història comença el 1900. El seu pedigrí inicial inclou també curiosament i directament –i no és pas casual– les preocupacions filosòfiques de l'empirisme anglès i les del racionalisme del segle XVII, totes elles derivades –avui ho sabem– del fort impacte de la Física de Galileu. I, ja que els *IA-aires* no solem tenir una cultura excessiva i tot això no ho acostumem a saber ni sospitar, de tant en tant va bé que ens anem recordant alguns fets bàsics, i que així tot plegat resulti un pèl menys incompreensible.

Perquè, sincerament, amb prou feines s'entendria res si no es recordés el trasbals conceptual que va ocasionar Galileu a primers del segle XVII amb la seva descripció racional de la natura en termes mecànics i formals: a partir d'aquell moment el món, l'home i la intel·ligència van ser mecanismes analitzables i comprensibles –i per tant dominables i reproductibles com ho podia ser una màquina– i tothom es va veure amb cor de pensar que el que havia fet Galileu amb la natura aviat es podria fer amb tot: així Descartes, el 1637, va creure que l'home en el fons era una màquina, que ell i el seu comportament acabarien podent ser fàcilment descrits en termes mecànics, i que eren pensables *màquines* que

BUTLLETÍ DE  
L'ACIA  
NÚMERO 9  
TARDOR DE 1996

fossin com nosaltres. I Hobbes, que creia que la societat i la política acabarien també essent *describibles* analíticament (i per tant al capdavant controlables), va formular per primer cop dinou anys més tard la frase que avui subscriuria qualsevol IA-aire dels del morro fort: «el pensament no és altra cosa que càlcul». I –ja s'intueix– d'aquí a la IA només hi havia un pas: de fet, la IA va començar a gairebé tots els efectes aquell 1656; la resta va ser espera i farciment. Els grans clàssics (els Hobbes, els Hume, els Leibniz) van desbrossar el camí que nosaltres refem, i van analitzar amb força allò que ens sorprèn. Vist així, la IA moderna ve a ser un *replay* de la Filosofia occidental dels últims 350 anys –molt accelerat per l'ús de l'ordinador com a banc de proves de cada hipòtesi (que així queda molt ràpidament validada o refutada)– que ha fet de nosaltres *IA-aïres* d'avui una canalla mal educada, però d'aprenentatge rapidíssim. L'única diferència amb ells, a part l'ús d'ordinadors, és que nosaltres no portem perruca, sempre anem de pressa i tenim una cultura humanística tirant a deficient. Però la idea que la intel·ligència, i en particular la intel·ligència analítica o *lògica*, havia de ser algun dia fatalment mecanitzada, aquesta la compartim amb els nostres clàssics, que no ho van dubtar mai gens. Només esperaven que arribés el dia i la tecnologia adequada.

Però abans de la tecnologia el que va arribar va ser el concepte mateix de *càlcul*, de *calculador* i d'*informàtica*. Com s'explicarà en aquest Butlletí a propòsit de les interrelacions entre Lògica, Informàtica i IA, com a conseqüència d'una sonada baralla entre dos matemàtics cèlebres als anys vint d'aquest segle alguns matemàtics i/o lògics (Hilbert, von Neumann, Herbrand, Gódel, Chulrch, Turing) es van posar febrilment a definir com havia de ser un procés general de validació i de *càlcul*. En va sortir definit el concepte de *càlcul efectiu* i de *màquina universal* (l'ordinador). Tot seguit, el 1946 va arribar amb l'ENIAC la tan esperada tecnologia, però aquest detall ara ja era secundari: abans i tot de l'existència del primer ordinador, la *informàtica* paradoxalment ja havia arribat amb Turing i tenia el llit fet. I mentre els enginyers dissenyaven els primers calculadors, més o menys monstruosos, lògics imaginatius com ara Pitts o von Neumann anaven preparant, amb els seus models aproximadament biològics de *neurona* i d'animal viu (*autòmat*) autoproductor, respectivament, la realització del somni: tenir algun

dia pròxim *màquines intel·ligents*. Per cert, l'expressió *màquines intel·ligents*, que sembla una flagrant contradicció de termes, no és gens ni mica sorprenent si es considera que aquesta idea es covava com s'ha dit des del 1637, preparada per Descartes i els seus epígons... entre els quals cal comptar per cert Leibniz, que el 1666, després de llegir Llull, va proposar un pla d'atac per mecanitzar el raonament lògic, com a farciment de l'humanoide cartesià.

Que la IA té, des del 1956 i per obra de McCarthy, una forta component lògica no hauria pas de sorprendre. Efectivament, és això el que ell, com a alumne del lògic Church, havia volgut que tingués: no endebades va ser McCarthy que va convertir el lambda-càlcul de Church en un llenguatge de programació (el Lisp). Amb això s'oposava a les interpretacions biologitzants de moda a l'època (fruit del respecte i la influència de Wiener i von Neumann) amb una concepció simple i realista: les màquines intel·ligents del futur eren ni més ni menys els *ordinadors*, i perquè aquestes màquines fessin coses intel·ligents calia programar-les. *Ergo* per atènyer algun dia la *intelligència artificial* (una expressió nova i un eslògan seu) cal fer *programes* com més va més complexos fins a generar conductes indistingibles de les humanes. En aquest pla d'atac, que lògicament havia de resseguir les passes de la intelligència humana, el primer punt d'arribada havia de ser el *reconeixement* de formes per màquina, seguit per la comprensió i generació del *llenguatge* i acabant en la part més difícil, el *raonament* abstracte. Però ha resultat que la dificultat real dels problemes ha estat exactament la inversa de l'esperada (primer s'ha resolt la part teòricament més difícil, mentre que del *reconeixement* el més calent encara és a l'aigüera, i del *llenguatge*, si fa no fa). I és que això de programar probablement té els seus límits i ara ens els comencem a trobar. Potser la intelligència, després de tot, s'assembla més, en el seu origen i funcionament, a l'*autoorganització* espontània dels sistemes complexos i a llur capacitat d'*adaptació* que no pas a la preexistència d'un *programa* (mot grec que vol dir, ni més ni menys, «allò prèviament escrit») que la majoria de vegades no és ni preexistent ni predeterminant ni tan sols *escribable*. Si és així, i si, com sospitava von Neumann, la IA del futur ha de recórrer als sistemes complexos, als algorismes evolutius i a coses per l'estil, aquest fet significarà la fi de la col·laboració de la lògica

i la IA en el projecte de fer màquines intel·ligents. I potser això és el millor que pot passar, perquè probablement la intel·ligència, ni que sigui artificial, és massa seriosa per deixar-la en mans dels lògics, o dels informàtics. Els seus models explicatius, sovint acusats d'esquemàtics i ingenus –a part de mortalment abstractes i avorridament formals–, consisteixen essencialment en el supòsit banal i corporativista que tota cosa intel·ligent s'ha d'assemblar per definició a un programador o a un analista d'aplicacions, hipòtesi explanatòria degenerativament simple que lliga poc, o gens, amb aquesta màquina de pensar, lenta i elemental però barata i ubiqua, que és el cervell.

No sabem pas què passarà els pròxims anys amb la IA. És possible que efectivament es vagi desvinculant progressivament de la Lògica i de la seva antecessora, la Filosofia mecanicista del set-cents, trencant així una associació de segles i arrossegant-hi la Informàtica, descendent directa de totes dues. O no. Ara bé, passi el que passi, l'únic segur és això: el que passarà ens sorprendrà. I molt.

## *La intel·ligència, de Llull al llimac passant pels robots sentimentals*

**A**QUESTA columna sempre ha volgut ser (1) *irònica* i (2) *impersonal*. Primer, *irònica*, i no sols pel to general que vol tenir sinó també per l'epígraf que duu («intel·ligència a tones»), brometa que els editors fan a costa del nom de l'autor que, com és evident, quan es tradueix a l'anglès resulta en una absurda proposició comercial («vendes a tones»), molt lluny dels interessos vitals de qui escriu –que en sa vida, ni quan treballava a IBM, ha pensat mai a vendre res a ningú. Al número passat, a més, els editors –o algú de la gresca– van fer aparèixer l'autor d'això com a treballador d'Obres Públiques, cosa que, corro a desmentir, jo no he estat mai ni en els meus millors somnis. La segona cosa que aquesta columna conrea, la *impersonalitat* (és a dir, no fer servir la primera persona del singular) aquesta vegada excepcionalment la deixo de banda, perquè amb les frases anteriors ja he sortit sense voler de l'anonimat i ja no ve d'aquí, i perquè en aquest número us proposo acompanyar-me en un recorregut lògic-sentimental que va de Llull als robots i als llimacs, en un trajecte bastant més directe i natural que no sembla. I és que al número anterior vaig prometre que parlaria de la relació de concatenació històrica entre la **Matemàtica**, la **Lògica**, la **Informàtica** i la **Intel·ligència Artificial** (IA) i ara ho vull complir (amb una mica d'escreix, com veureu). En particular, us hi volia explicar com la **Matemàtica** va engendrar la **Lògica** i de com aquesta va engendrar la **Informàtica** i aquesta la **IA** –en una mena de nou Gènesi laic i abreujat (que per cert, com vaig dir quan ho vaig prometre, em sembla que s'acabarà tard o d'hora en divorci

BUTLLETÍ DE  
L'ACIA  
NÚMERO 10-11  
PRIMAVERA-ESTIU DE 1997



entre la IA i els seus progenitors). En aquest número trobareu, doncs, aquesta història estructurada en còmodes episodis com si es tractés d'una epopeia d'heroiques gestes, que casualment sembla que només ocorren els anys acabats en 6. Però és que, abusant del dret de cuixa que amablement em concedeixen els editors, a partir d'aquest nucli històrico-conceptual he allargat la cosa èpica cap enrere, fins arribar a en **Llull**, que encara que no ho sapigueu és el pare de tots nosaltres. I també ho he allargat cap endavant, per explicar-vos resumidament què s'està fent a hores d'ara per **programar robots** a base de dotar-los, com si fossin nosaltres, d'*estats mentals* –bufa!– com ara creences, desitjos, propòsits –bons, és clar–, intencions –immillorables–, etc. (però, com recomana McCarthy, sense posar-hi cap *sentiment*, no fos cas que en sortíssim tots socarrimats). Ho faig perquè ja ho tenia escrit i també perquè és un exemple més d'allò que deia la vegada passada: sembla que els informàtics no podem deixar de pensar que *ser intel·ligent* no és altra cosa que fer d'analista d'aplicacions. I ja es veu que això no és gaire creïble: dut a l'extrem, voldria dir que tots els intel·ligents d'aquest món, abans d'exercir, haurien de passar per alguna facultat d'informàtica.

Els dos articles que complementen el nucli promès són, el primer, sobre Llull i, el segon (que va darrere dels altres dos) sobre la lògica dels «estats mentals» dels robots. Estan trets, amb permís de l'autor (servidor, que alhora els ha traduït i adaptat), de sengles participacions en sengles congressos d'informàtics programats per aquest maig. El primer treball, el de Llull, és una presentació en un workshop d'«informàtics paral·lels» (ja m'enteneu: programació, control i verificació de programes concurrents), entre els assistents del qual cal comptar els il·lustres Zohar Manna i Amir Pnueli i el semànticodenotacional Manfred Broy. El segon, sobre les hipotèticoputatives «lògiques d'estats mentals»–d'existència problemàtica i potser plausible–, és un tutorial sobre el tema per ser presentat a un heterogeni club de desvagats que inclou des de lògics a psicòlegs passant per metges i filòsofs i que normalment es coneix, sobretot a l'hora de fer llibres i congressos, com a *ciència cognitiva*.

Deixo Llull –que ja al seu temps rebia l'ambigu elogi d'«il·luminat», el pobre!– per a l'article resum que podreu llegir més avall i que amb l'editor hem amanit amb algunes il·lustracions

passadores. L'altre ítem és la promesa cronologia, una explicació succinta del no gaire conegut traspàs per etapes que va de la Matemàtica a la IA, cronologia sumària a la qual he afegit algunes consideracions que ja havia expressat fa uns anys a requesta de la Vanguardia i que, de fet, no són sinó banalitats refregides, res que el lector no sàpiga de sobres. És sobre el tercer article, sobre els «estats mentals, i com programar-los», del que us vull parlar ara. Malgrat el que hi dic –i us juro que m'he pres el tema molt seriosament– no he pogut estar-me de pensar que, com en tantes altres coses, en el fons els informàtics fem trampa. (Això no és el primer cop que ho dic en aquesta columna, però és que hi tinc com una dèria.) La situació és la següent: oi que ens entenem perfectament, quan diem «vull fer això o allò»? Doncs ja està: fem-li-ho dir a un ordinador, i ho farà. Que fàcil, oi? Evidentment, necessitarem una àlgebra que limiti la cosa de manera que no hi hagi irracionalitats ni absurds, com ara persistir a voler fer una cosa demostrablament impossible (i doncs! com si això no fos el que insistentment fem nosaltres, els imperfectíssims humans!). Un cop explicat què (i com) cal *desitjar*, *evitar*, *creure* i *fer* (que és per això que construïm les nostres «lògiques d'estats mentals»), sembla que ja ho tenim tot fet: només cal deixar anar pel món els nostres instruïts robotets, esperant que no s'equivoquin tant com nosaltres, ni tan sovint.

Ja ho he dit manta vegada en aquesta columna, i ho torno a dir. Per fer de Dr. Frankenstein i parir éssers intel·ligents –igual com per ser gàngster, suposo– no s'hi val a ser tendre i raonable. No és tractant el robot com si fos de la família que el farem llest tot esperant que, com els infants, la vida li acabi d'endreçar les coses que nosaltres li hem esguerrat o li hem callat. A base de dir coses que nosaltres entenem, fem que el robot (o un infant) «entengui»? Nosaltres, que programem les màquines, les «fem» per això *intelligents*? o els intel·ligents ho continuem essent només nosaltres, i elles només fan que seguir el nostre disseny (com fa cada dia qualsevol dels milers de perfectes imitadors d'Elvis Presley o dels Beatles, que ningú no confondria ni boig amb la cosa real)? Nosaltres som un munt de neurones sucades en aigua (una sopa tèbia, que deia Turing) i, tanmateix, –per una mena de miracle evoltiu que ha costat segles– intel·ligents. Vol dir això

que podem bufar i posar una ànima en una pila de circuits i dir que l'hem feta intel·ligent? I on posem *l'experiència*, que –amb el seu ingredient fonamental, el *fracàs*– és condició necessària, com deia Turing, per a qualsevol intel·ligència? Aquestes són preguntes que qualsevol IA-aire s'ha de fer cada dia perquè afecta el sentit mateix de la professió. I no són gens filosòfiques ni trivials, de cap manera. Es resumeixen en això: el que fem, és intel·ligència? o simplement *programació avançada*? («avançada» en el sentit que els programes fan coses que semblen intel·ligents o, dit a la manera de Minsky –que és la canònica–, que si fossin fetes per un humà diríem que són intel·ligents). Però per això no calia tanta faramalla. És un criteri antic (el fundacional, de l'any 1956) i modest. No enganya ningú, i si en traiem alguna cosa sorprenent, amb aspecte de realment intel·ligent, doncs millor, això que tenim de guanyat.

Tornant al cas que m'ocupava dels «estats mentals» dels robots: quan parlem de *desig* o de *por* o de *dolor* o de *voler*, és clar, tots ho entenem. Doncs sembla que quan en parlem a un robot bé ho hauria d'entendre ell també, si més no perquè li ho expliquem tan bé com sabem, amb uns enunciats inambigus i formalment incontestables com són els de la Lògica més o menys Matemàtica. Però, ho entén? Els informàtics, acostumats per obligació a veure-ho tot clar, potser no ens n'adonem i no hi vacillem gens ni mica, però els cagadubtes i desconfiats lògics, acostumats com estan a buscar deliberadament, malaltissament, interpretacions noestàndards de les coses, no poden sinó ser escèptics. Imagineu que en comptes d'interpretar tal com sona «vull fer  $x$  perquè acabo de deduir que  $x$  em serà beneficiós i fent-ho no perjudicaré ningú ni trairé cap altre dels meus objectius» ho llegim tot perversament –com és la nostra obligació de lògics– «cerco anul·lar  $x$  perquè acabo de veure que no puc provar que  $x$  serà un escrofulós pragmàtic i perquè anul·lant-lo no deixaré executar cap càlcul ni deixaré continuar cap dels processos que hi hagi engegats». Quina garantia tinc que les coses que jo dic s'entenguin com jo les dic? Aquesta obvietat, que els lògics coneixen de la seva feina diària, ja estat dita moltes vegades des de Lewis Carroll, que feia dir a Humpty Dumpty, davant una perplexa Alícia, que les coses no volen dir res, només el que nosaltres volem que diguin, fins a qualsevol filòsof o humorista amateur que sap el partit que se'n pot treure dels extraterrestres

per fer veure que les coses òbvies, sempre que hi hagi un bon marcià, d'òbvies no en tenen ni el folre.

En termes més profunds o actuals, això que estic plantejant se'n diu, gràcies a la nostra pedanteria innata i als esforços del psicòleg Steven Harnad i d'alguns altres IA-aïres, el *symbol grounding problem*. Resoldre'l és bàsic, si ens volem aclarir en això de connectar els fenòmens de la intel·ligència *simbòlica* (que són, per definició –o convenció– els que exigeixen al nostre cervell més de 100 mil·lisegons de temps de procés) i els altres, els *subsimbòlics* (que són els que en diríem, per entendre'ns –i de manera segurament impròpia–, «connexionistes» o «analògics»). El *grounding*, que és una qüestió molt més seriosa que no sembla, és, bàsicament, trobar el lligam o mecanisme pel qual es connecten els *símbols* (que per naturalesa són discrets i poden significar el que se'ls vulgui fer significar) i les experiències (sensacions? emocions? necessitats fisiològiques? impulsos –instintius o culturals– bàsics?) que els donen contingut i significat. Per cert, la paraula anglesa *grounding* és força expressiva: vol dir 'posar de peus a terra' una cosa que altrament potser es quedaria als núvols o, en termes elèctrics, connectar un circuit «a terra» per evitar enrampades inútils (que potser ens convindria força als IA-aïres, que prenem mal massa sovint).

Com ja he dit altres vegades, el meu escepticisme sobre els enfocaments només lògics és enorme, precisament per això, perquè són lògics i, per tant, deliberadament «flotants», penjats –massa–, i es poden interpretar amb més graus de llibertat dels que ens convenen. I no s'hi val a dir: «és que ja sabem de què parlem», perquè això, a més de ser un ús esbiaixat i fraudulent de la Semàntica, no resistiria pas la prova del marcià, qui, com és obvi, no sabria gens de què parlem quan nosaltres ja sabem de què parlem. La meva curiositat s'inclina per explorar altres possibilitats que potser ens ajudaran a passar el pont aquest del *grounding* i que, ara com ara, em semblen una bona alternativa per desenteranyinar-nos el cervell i fugir d'aquesta autohipnosi de pensar que les coses són clares (per exemple, que sabem què vol dir *creure* o *voler*) només perquè ens ho sembla. Cito algunes alternatives, però no voldria pas que ningú pensés que aquestes sí que són «la solució» del problema; no s'hi val passar d'un extrem a un altre, que per a

messies i solucions màgiques ja tenim altres bàlsams consoladors. Esmento primer la hipòtesi del sovint bescantat Rodney Brooks: la *intelligència és*, bàsicament, el resultat *emergent* d'un seguit de fenòmens elementals de molt baix nivell (que rarament qualificaríem d'intelligents). Una altra hipòtesi, no necessàriament gaire diferent: la de Luc Steels, segons la qual la intelligència és essencialment un fenomen *analògic*, no *categòric* (*categòric* vol dir que és producte d'una categorització prèvia), i apilable via *superposició* fins a arribar a complexitats notables. I una darrera, probablement compatible amb les dues anteriors: la intelligència potser és el resultat emergent d'un seguit de fenòmens més o menys simples però que, per raons d'escala, es combinen macroscòpicament d'acord amb les lleis físicomatemàtiques que a poc a poc anem descobrint de la *complexitat*, de l'*autoorganització*, de l'*adaptació* al medi exterior i d'una difícil i precària *estabilitat* sempre a punt de vorejar –o de caure-hi!– malaltissament la zona crítica i el *caos* autoregulator.

Per il·lustrar-ho, i perquè es vegi com es d'obert el camp que tenim al davant, m'agrada poder-vos presentar el resum d'una tesi feta aquí mateix, a Catalunya, per l'informàtic Jordi Delgado, en la qual es parteix d'algunes d'aquestes bases conceptuals de les quals us parlo, se n'exploren algunes línies, s'hi treballa a fons, i s'hi arriba a algunes conclusions que el doctorand mateix us explicarà. Jo m'he permès d'adjuntar-hi un comentari introductori que pretén continuar i culminar la línia argumental que us proposo en aquest número i que va del venerable i savi **Lluïl**, alhora clàssic i eixelebrat, als nostres **robots** contemporanis, que per raons vàries ens entestem a dotar de voluntat. És en aquesta introducció/prefaci on parlo de **llimacs**, de formigues, d'eixams, de xarxes neuronals fluides i de càlcul col·lectiu. El doctorand no sols en parla (de llimacs no, però) –i bé– sinó que, a més, prova amb la seva millor voluntat de lligar totes aquestes estranyeses amb l'encarcarada informàtica «acadèmica» (un pèl massa circular i autojustificatòria) que li varen –vàrem– ensenyar a la facultat.

En fi, que hi ha camp per córrer, en això de la IA, i qui no es conforma (o no hi troba tema de tesi) és ben bé perquè no vol. Però abans hauríem, penso, de deixar de creure'ns les piadoses mentides que ens anem explicant. Per exemple aquesta (que més bàsica i elemental ja no pot ser, i que lliga –encara que no ho sembli–

amb la qüestió anterior, la dels robots que tenen «estats mentals»): sabem calcular? Vull dir: sabem què vol dir, realment, **calcular**? Alguns potser es pensaran que m'he tornat boig i commiseradament em recomanaran que em llegeixi «la definició» en algun del textos que fa anys que jo mateix explico. Però hi insisteixo: Ho sabem? L'únic que trobo indubtable és que creiem –vaja, n'estem convençuts– que ho sabem perquè som capaços de donar-ne una definició satisfactòriament «matemàtica» (aquesta: Càlcul = procés *prèviament especificat* segons el formalisme canònic de Turing o qualsevol altre de Turing-equivalent). Però ¿és calcular realment això que diem? Searle (i Hillis, com jo mateix) ho dubta. Sembla que ens ha passat desapercebut un detall: per calcular cal tenir –a més del que diu la definició– *voluntat* de calcular i *ignorància* del –o *curiositat* pel– resultat. (I, a més, se suposa que el procés de càlcul és més o menys *conegut* –i ben entès– i *controlable*; si no, quin interès tindria per nosaltres? ) I no s'hi val a dissimular: si no hi ha aquests ingredients suplementaris no hi ha pròpiament *càlcul*; com explica Hillis (el de la *Connection Machine*) qualsevol extraterrestre tindria dificultats per entendre què fem quan «calculem» sense que prèviament li expliquéssim (com?) què volem fer i per què, i la feina seria nostra per trobar-li'n una definició rigorosa que li aclarís mínimament allò tan evident. Si a la recepta estàndard no hi posem (com?) les components «mentals» (tècnicament parlant caldria dir *intencionals*) mancants com ara la «voluntat», l'«interès» pel resultat, la «familiaritat» aproximada amb els mecanisme i la seva «controlabilitat», l'únic que ens resta és un vulgar procediment que és formalment indistingible de molts altres. Dit en termes de lògic: amb la definició matemàticocanònica no n'hi ha prou: sense els ingredients intencionals –que, per cert, són realment formalitzables?– el procés calculatori admet moltes interpretacions legítimes que nosaltres no acceptaríem com a càlcul. Per exemple, no podríem negar que un depredador «calcula» una equació de Volterra-Lodtke quan va darrere la seva presa, que un projectil calcula paràboles, o que la llum calcula la seva pròpia trajectòria. Una aplicació rigorosa de la definició de Turing (però llegiu què deia Turing: més d'un s'endurà una sorpresa) ens diu que tot això són càlculs (tècnicament, són processos Turing-calculables). Però, els acceptaríem nosaltres com a tals?

Nosaltres «sabem» que calculem –com sabem que creiem, que volem o que fem–, però traduir això en semàntica i tot seguit convertir aquesta en formalitat executable, és possible? Quina desil·lusió si, després de perseguir el rigor formal ens fessin caure la cara de vergonya perquè les nostres definicions pretesament caracteritzadores i inambigües fossin insuficients i acabéssim havent-ho d'explicar com qui explica una opinió personal o la pel·lícula de la TV! Que poc matemàtic –i que poc científic– resultaria! Jo no dic pas que algun dia no es pugui formalitzar realment la noció de **càlcul** (si bé ho dubto), però abans caldrà fer-ho, veure-ho i discutir-ho sense benes als ulls. Mentrestant, creure que sí i que ja ho hem fet és probablement somiar truites, però força útil: permet abordar qualsevol cosa «calculable» (en principi, tot l'univers) amb aquestes màquines primitives actuals que anomenem ordinadors i amb/de les quals vivim. Pot ser tot plegat un gremialisme – fructífer, però gremialisme. De fet és, simplement, un **prejudici** metodològic... sobre el qual se sosté, per cert, tota la ciència que ensenyem a les facultats d'informàtica.

I és que tenir prejudicis és tan corrent que els científics, pretesament verges d'aquest pecat, n'anem plens. Amb l'inconvenient que, com que estem convençuts que sempre tenim raó, reconèixer, amb una sempre convenient humilitat, que anem imperdonablement de biaix en biaix resulta difícilíssim, per no dir impossible. Però és que per a un científic o tecnòleg hi ha una cosa pitjor que tenir prejudicis, i és no saber que se'n tenen. Desgraciadament, els IA-aïres, en podríem omplir un llibre. O unes quantes tesis doctorals.

Barcelona, 19 d'abril de 1997

# *De la matemàtica a la Intel·ligència Artificial passant per la lògica i la informàtica. Una història episòdica*

*Episodi 1<sup>r</sup> (1906): Un malson dels matemàtics: la «inefectivitat»*

BUTLLETÍ DE  
L'ACIA  
NÚMERO 10-11  
PRIMAVERA-ESTIU DE 1997

**G**EORG Cantor –que, segons Hilbert, creant la Teoria de Conjunts ens havia construït un paradís– havia demostrat de manera sorprenent però indiscutible (mitjançant un argument de pròpia collita, la *diagonalització* –que avui ha esdevingut el més practicat diàriament pels matemàtics–) que hi ha inenumberablement més nombres transcendents que algebraics (que inclouen els racionals i les arrels com ara  $\sqrt{2}$  o  $5^{1/3}$ ). La gràcia del cas –si se'n pot dir així, perquè consta que a molts no els en va fer gens– és que *tots* (excepte un) els nombres coneguts eren de la segona espècie. L'únic transcendent conegut era *e* (i això només se sabia de feia poc). I ara venia Cantor i deia que aquesta classe desconeguda de la qual podíem esmentar amb prou feines un sol exemplar era infinitament més nombrosa que la dels nombres normals. I tot això, sense donar-ne ni un exemple (*e* i la seva transcendentalitat no hi feien cap paper en la demostració).

Aquest és un dels primers teoremes diguem-ne anòmals de tota una sèrie de resultats més o menys allucinants que la matemàtica



d'últims del segle passat anava aconseguint mitjançant raonaments sofisticats, discutibles però incontestables. En posarem exemple cèlebre, el que dóna resposta a la qüestió: «¿és possible que un irracional elevat a un exponent també irracional doni una potència racional?» L'interessant no és la resposta –afirmativa–, que no deixa de ser raonable, sinó la sorprenent demostració. Va així: suposem que  $a$  i  $b$  són tots dos  $\sqrt{2}$ ; llavors, una de dues, o  $a^b$  és racional, o no ho és. En el primer cas, ja ho tenim: la resposta és sí. Si estem en el segon cas, aleshores prenem aquest nombre  $a^b$ , que acabem de dir que és irracional, i l'elevem a  $\sqrt{2}$ , que també ho és: la potència resulta valdre 2, que evidentment és racional. Resultat: la resposta és sí. La demostració és elemental, fàcil de seguir i tothom hi està d'acord, però ens hem quedat amb un pam de nas: quin dels dos casos és el que dóna la potència racional quan s'eleva a  $\sqrt{2}$ ? el primer, en què la base era  $\sqrt{2}$ ? o el segon, en què era  $\sqrt{2}$  elevada a  $\sqrt{2}$ ? Dit d'una altra manera, sabem (perquè ho hem demostrat) que  $a^b$  pot ser racional amb  $a$  i  $b$  irracionals, però no hem estat capaços de mostrar l' $a$  ni el  $b$  per als quals ho hem «demostrat». Aquest problema il·lustra típicament la situació, odiada pels matemàtics, en què no hi ha manera de decidir «efectivament» quin és el cas que hem resolt. Una cosa així és perfectament comprensible per a un informàtic: en efecte, quan es programa un ordinador no se li pot pas dir simplement: «Hi ha una opció que resol el problema»; la màquina immediatament planteja: «Molt bé: quina, doncs?» Aquest és el problema clàssicament conegut com a *efectivitat*. No n'hi ha prou d'especificar com es faria per resoldre un càlcul. Cal especificar la seva resolució pas a pas de manera que tothom, humà o màquina, hi pugui accedir seguint una sèrie d'instruccions totalment inambigües en un nombre finit de passos, sense haver de trobar-se mai amb decisions no prou especificades ni amb situacions de les quals no hi ha garantia que s'acabin en un moment donat. Fer-ho així és el que els matemàtics anomenen procediment de decisió o procediment efectiu, i el que els informàtics anomenen algorisme (o algoritme). Tots els matemàtics de l'època, molt tocats, van començar a insistir a prescriure l'ús exclusiu de procediments efectius en matemàtiques, mentre que Hilbert no tenia inconvenient a tolerar meravellat aquestes entremaliadures inefectives, pel que tenien de sorprenents i creatives i, sobretot,

potentíssimes. Quan el mateix Hilbert va demostrar l'existència d'un cert nombre per reducció a l'absurd però sense donar-ne característiques ni aproximades, Paul Gordon, el matemàtic alemany que havia proposat el problema «resolt» per Hilbert, va trobar la «demostració» especialment escandalosa, i ho va descriure dient: «Això no és matemàtica, això és teologia»!

### *Episodi 2<sup>n</sup> (1916): Un pèrdua: la intuïció*

Hilbert, davant de la sorprenent potència constatada d'aquests aparents entreteniments lúdicomatemàtics de tombant de segle i constatant que -tot i pseudolúdics o peculiars- eren de fet potentíssims i que depassaven allò raonablement esperable i comprensible, va proposar que a partir d'aquell moment les Matemàtiques ja no es fessin en virtut de la intuïció dels matemàtics –que fins llavors havia estat l'única guia– sinó de la *correctesa* formal –amb un rigor comprovablement absolut– de les demostracions. No hi feia res que la significació d'un «objecte» –del qual s'acabava de demostrar l'«existència»– o d'un teorema fos inimaginable. De fet Hilbert creia que, com ja havia passat històricament, tard o d'hora arribaria algú que hi trobaria un significat –i potser fins i tot una utilitat– i ja ens ho explicaria tot ben explicat. Calia, doncs, convertir la Matemàtica en un procés absolutament rigorós, de fet, en un procés mecànic –i, si es podia, mecanitzable– de càlcul. Davant d'això alguns, com Brouwer (un deixeble de Hilbert) van reaccionar molt malament: ho van considerar com una degradació de la Matemàtica a una activitat no creativa, amb absència total de la intuïció i reduïda a una mena de joc de símbols sense cap significat. Weyl recordava que la Matemàtica és una activitat creativa humana, ajudada i clarificada per la Lògica, però res més; la lògica no hi decidia res. Hilbert contraatacava dient que la Matemàtica d'avui és tan abstracta que l'home s'hi perd si no recorre rigorosament al càlcul lògic formal; i afegia que Brouwer i Weyl feien trampa; si hi havia res que no entenien o sobre el que no tenien intuïcions clares, senzillament se'n desembarassaven.

*Episodi 3<sup>r</sup> (1926): Un dèria: la «construïbilitat»*

Brouwer va ser qui va encapçalar la posició majoritària pro-efectivitat i pro-intuïció dels matemàtics. Tant, que va arribar a enfrontar-se amb Hilbert –fins i tot en el terreny personal per aquesta qüestió. Quan finalment van partir peres, i molt de mala manera per cert, Brouwer (i els anomenats *intuicionistes*) va elevar la seva condemna a graus de fonamentalisme, mentre que Hilbert (i els anomenats *clàssics* o *formalistes*), va elaborar tota una Matemàtica que diríem «sense manies». Ara bé, ho va fer construint-la, com si donés la raó als seus contradictors, mitjançant procediments rigorosament efectius, que ell va rebatejar com a *constructius* i *finitistes*, de manera que ningú no es pogués queixar que no es pogués arribar a qualsevol dels seus resultats mitjançant un nombre finit de passos perfectament especificats i executables. Així va voler demostrar la consistència (= no-contradictorietat) de qualsevol teoria, que per als lògics és un problema vital, per recurs a la construcció d'un model (cosa que, com ell s'havia entestat a fer veure, resolva el problema), que calia que fos «efectiva» per ser indiscutible. Per fer-ho, al 1926 va ampliar el concepte d'inducció de Dedekind i Peano (d'uns 40 anys abans) al d'inducció *transfinita* sobre els ordinals (una generalització dels naturals). (És a dir, Hilbert acabava d'inventar la recursivitat.) A partir d'aquí tota la Matemàtica passava a ser un càlcul gegantí que començava amb l'Aritmètica i acabava, com a mínim, en la teoria de conjunts (que, com tothom sap, és la base de tota la resta).

El seu deixeble Gödel ho explicava així: «La formalització rigorosa de la Matemàtica permet demostrar qualsevol teorema només amb unes quantes regles mecàniques». Fer matemàtiques quedava, doncs, reduït a fer un càlcul mecànic. I viceversa: això obria inopinadament la porta perquè els matemàtics, duts ara per aquesta falòrnia, fossin els primers a definir rigorosament què havia de ser, doncs, un càlcul, i en què consisteix pròpiament això tan intuïtiu i general de calcular.

*Episodi 4<sup>t</sup> (1936): Una capsa de sorpreses: la «calculabilitat»*

Hilbert va plantejar la qüestió, que ell creia que es resoldria de manera afirmativa, de si «donat qualsevol enunciat, és possible «decidir» si pertany (o no) a una teoria donada?» Aquest problema, de particular interès en el cas de l'Aritmètica, és el conegut com a *problema de decisió* (l'*Entscheidungsproblem* dels textos alemanys). Hilbert estava convençut de la seva resposta afirmativa. La seva solució positiva permetria convertir la matemàtica, i per extensió la Ciència en general, en una qüestió mecànica, molt en la línia industrial del segle. La creativitat en la Matemàtica quedava reduïda a trobar bones preguntes: la solució s'obtindria automàticament i la qüestió quedaria resolta. El problema és que de vegades el mecanisme deductiu entra en un cercle viciós (*bucleja*, que diria un informàtic): estem segurs que hem triat i engegat correctament la deducció, que aquesta «va bé» (i, doncs, la pregunta «s'està contestant»), però la cosa no sembla que s'acabi, i l'únic segur és que si s'acabés –o en tinguéssim la garantia– llavors podríem assegurar ja des d'ara que la pregunta té contesta i només caldria esperar que el procés deductiu s'acabés d'una vegada. Però: i si realment haguéssim caigut en un bucle (i la pregunta fos, per tant, incontestable)? Com detectar-ho?

Aquest problema va rebre dues vies de solució. Com sol passar, una (la de Princeton, que descriurem més avall) era «la correcta» –però no el resolía. L'altra –que sí que el va resoldre– era totalment inesperada i venia d'Anglaterra i d'un *outsider*, Alan Turing. La via «correcta» era l'esperable de lògics com Church, Kleene (el seu doctorand) i Gödel, que els visitava a Princeton amb tot els honors. El planteig era el següent: deduïble vol dir 'calculable' d'una certa manera, seguint rigorosament uns procediments prescrits. Per tant, si ens en sortim a definir les funcions calculables de manera que en sapiguem les característiques i les puguem identificar, aleshores qualsevol deducció passa a ser un càlcul (que «s'acaba», o tenim garanties que ho faci). Els de Princeton van tenir la sort de trobar una manera de caracteritzar tot d'una tots els càlculs. Era un subproducte de la casualitat: Church s'havia muntat un sistema axiomàtic per a tota la Lògica que no va funcionar (va resultar que era inconsistent, com va veure Kleene) però que contenia una part

aprofitable: un mecanisme formal, molt abstracte, per descriure qualsevol funció (ara d'això en diem  $\lambda$ -càlcul –o Lisp, si som informàtics–): la gràcia és que, seguint mecànicament un parell de regles molt senzilles (de *reescriptura*) es podia trobar sempre la solució... sempre que la funció fos calculable, és clar. Kleene, preparant el seu doctorat, el 1932 va anar veient no sols que el formalisme de Church era potentíssim sinó que totes les funcions calculables eren reduïbles a unes quantes de molt bàsiques, a partir de les quals totes les altres s'obtenien per operacions simples tals com la *composició* o la *recursió*. Quan Church i Kleene van presentar els seus sorprenents resultats al visitant Gödel, aquest se'n va sorprendre i, més aviat incrèdul, va oferir la seva pròpia manera de construir totes les funcions calculables en un formalisme molt elaborat en què va esmerçar tot el 1934 i que va anomenar *teoria de funcions recursives generals*. Quan ho va haver acabat, Kleene li va demostrar que el seu formalisme era equipotent a (= permetia de calcular les mateixes funcions que) el de Church. Després d'una agradable sorpresa com aquesta, Church va sospitar –i emetre– allò que ara es coneix com a tesi de Church (i que va ser acceptat provisionalment i una mica a contracor per Gödel): «Vist l'estat dels nostres coneixements relatius en aquest moment (1935), sembla que hem trobat el mecanisme més potent per caracteritzar totes les funcions calculables i per tant, mentre no se'n descobreixi un altre de més potent encara, *qualsevol funció calculable ho és mitjançant el  $\lambda$ -càlcul*». Semblava fort i lleugerament incompreensible (i el  $\lambda$ -càlcul un pèl massa abstracte i *impràctic*), però els lògics ho van anar acceptant (bastant a contracor, com Gödel).

El professor de Turing, el lògic Newman, li va explicar el que pretenien a Princeton (caracteritzar totes les funcions calculables), i Turing sense pensar-s'hi va enfocar el problema no en termes abstractes (d'operacions o manipulacions formals sobre objectes matemàtics) sinó mecànics: què cal fer per calcular? S'ho va prendre com un exercici d'abstracció a partir de la vida real, i el resultat va ser la *màquina de Turing*. Primer va voler saber què vol dir calcular, i va veure que amb un mecanisme senzill (una cinta i un capçal) i una llista finita d'instruccions ja feia el fet. Tot seguit va veure que aquest mecanisme era tan potent que permetia (1) executar qualsevol deducció i (2) construir una màquina prou

potent («universal») que pot simular el càlcul que faria qualsevol altra amb unes dades d'entrada donades. A continuació ho va aplicar a l' *Entscheidungsproblem* de Hilbert. Ho va plantejar així: si un càlcul o deducció sembla que bucleja però acaba parant només caldrà esperar i veure'n el resultat; ara bé, si no sabem si realment estem en un bucle no podrem pas decidir res sobre el resultat. Per tant, ens cal algun mecanisme que, analitzada la situació del nostre càlcul, predigui si s'aturarà o no: això resoldrà el nostre problema. Quin pot ser aquest mecanisme analitzador? La resposta –necessàriament relativa– de Turing és: la màquina més potent de què disposem, és a dir, el calculador universal obtingut més amunt. Si ell resol el problema, el *problema de la decisió* de Hilbert queda resolt, i si no, no. Doncs bé, Turing mateix va demostrar que aquesta màquina màximament potent de què disposava no li resolva el problema i, per tant, Hilbert havia somiat truites: el problema de saber si una cosa es dedueix d'un altra és indecidible. Això va representar un cop fort per a unes esperances força raonables, i va marcar un punt i a part.

### *Episodi 5<sup>e</sup> ( 1946 ): Extrapolació i interpolació: de la cèl·lula a l'autòmat*

Passat el desconcert de la frustració, a poc a poc la gent es va adonar que el mecanisme abstracte (abstracte, però mecanisme) descrit per Turing el 1936 era, a més de màximament potent (tant com el de Church), molt senzill i intuïtiu. Així ho va reconèixer no sols un entusiasmat Gödel sinó també McCulloch i Pitts, un parell de científics joves, biòleg l'un i lògic l'altre, que el 1943 van pensar que si la màquina de Turing podia calcular tot allò calculable i el nostre cervell podia fer tot el que nosaltres som capaços de fer, bé devia ser possible imaginar-se el nostre cervell de manera que fos, a la seva manera, una màquina de Turing. I així ho van fer. Se'l van imaginar com un conjunt de «neurones» interconnectades (la primera «xarxa neuronal» de la història) de manera que cada neurona fos, simplement, un element sumador saturat. Amb això n'hi havia prou, semblava, per reproduir tot allò que una màquina de Turing pogués fer (que, per definició i de moment, era tot). Teníem així la versió biològica de Turing i, com un segon ocell

agafat amb el mateix vesc, un model simple però força plausible del nostre cervell.

Von Neumann, que en aquella època assessorava a Los Alamos els físics de la bomba atòmica (proveint-los de tabuladores IBM per a les seves simulacions), alhora parlava amb els periodistes, el 1946 (acabada la guerra i passats, per tant, els rigors de la censura militar), de futures màquines electròniques «intelligents» amb unitats «lògiques» i de «memòria» que «pensarien » com nosaltres. Evidentment pensava en (1) la de Turing, en (2) el «cervell» de McCulloch-Pitts i en (3) l'ENIAC, el primer ordinador electrònic acabat d'inaugurar aquell mateix 1946, la construcció del qual ell havia assessorat i la millora del qual ell mateix estava projectant detalladament (l'anomenat *projecte EDVAC*). Naturalment, les sensates declaracions de von Neumann van ser tretes ràpidament de context pels periodistes, que només hi van veure «cervells electrònics» artificials, de ciència-ficció. Però von Neumann estava realment entusiasmat per la situació (com Wiener, que l'estava extrapolant vers una quasi paradisiàca societat «cibernètica») i de seguida va començar a concebre com havia de ser una màquina viva, és a dir, no sols intel·ligent sinó capaç d'autoreproduir-se (d'aquests autòmats autoreproductors de von Neumann d'últims dels quaranta, avui en tenim dues herències deliberadament i abusivament simplificades: els autòmats finits i els autòmats cel·lulars, tots dos de la dècada següent). L'entusiasme dels Wiener i von Neumann es va transmetre fàcilment a joves com Ashby o un estrany matemàtic afeccionat a la ciència ficció i al disseny de cervells anomenat Minsky. La cosa semblava clara: la intel·ligència era (1) assequible i (2) complexa, en el sentit que, com el nostre cervell, semblava una qüestió de sistemes molt interactius i complexos en perpetu i precari equilibri, potser modulars i, encara que semblés difícil, potser matematitzables. Sobre això últim es van descobrir aviat dues coses interessants. Primera (això ho va descobrir Kleene): el «cervell» de McCulloch-Pitts era massa elemental, ja que el llenguatge més potent que amb una «màquina» com aquesta s'hi podia reconèixer era un que Kleene va anomenar *regular*, molt elemental en comparació amb el que els humans són capaços d'entendre. La segona (i això és de von Neumann): la vida i la intel·ligència han d'incloure l'error i el seu tractament com una

part intrínseca de l'estudi (com, d'altra banda, sempre ha fet la Matemàtica amb els errors de càlcul) i, per tant, cal tocar aquest tema amb eines semblants a les que fa servir la Termodinàmica per tractar el soroll i l'entropia.

Els nous plantejaments havien convertit el problema del càlcul i de la intel·ligència, que Turing havia plantejat d'una manera tan simple, en un problema contra el qual els matemàtics que s'hi enfrontessin haurien d'aplicar la seva màxima artilleria, des dels sistemes complexos d'equacions diferencials a la teoria de sistemes amb realimentacions, interaccions i homeòstasis. No cal dir que no tots els matemàtics es van sentir feliços de saber que la cosa se'ls havia complicat tant. El paradís de les funcions de Church, reposadament executables en simplicitat màxima de Turing, semblava perdut per sempre. O no? Perquè... què era un ordinador IBM 650 (anunciat el 1955) sinó l'última versió del somni de Turing fet realitat? I per què estudiar cervells vius podent fer servir el dels estudiants/universitaris/programadors més brillants (que, a més, s'hi prestaven entusiàsticament, i de franc)? L'any 1955 el matemàtic McCarthy va reunir en un llibre tot un seguit de treballs (incloent el de Kleene i el de von Neumann que hem citat) per portar aigua a aquest nou molí. Això, i el que va venir immediatament després, és tota una nova època: la de la *Intel·ligència Artificial* (la «IA»).

### *Episodi 6<sup>è</sup> (1956): Els joves repatanis i la Intel·ligència Artificial.*

Com tothom sap, la *Intel·ligència Artificial* (IA) va néixer el 1956. Un grup d'universitaris americans joves, majoritàriament matemàtics (i hereus intel·lectuals de Church), compartien una aversió personal, fins i tot física, pels Wiener, Ashby i von Neumann i la seva manera de veure la intel·ligència no com un procés purament matemàtic sinó com una cosa relacionada amb els sistemes vius. L'estiu del 1956 aquesta colla es va reunir a Dartmouth, a New Hampshire, per reenfocar l'estudi, imitació o «reproducció» (si tal cosa es veia factible) de la intel·ligència humana. Hi havia John McCarthy, Marvin Minsky (canviat de bàndol), Herbert Simon (economista, matemàtic i futur Nobel) i altres. El seu plantejament



era clar: com que tots ells tenien ordinadors a les seves universitats (IBM n'acabava de regalar 100) i alumnes entusiastes capaços de programar-los, es van proposar crear equips i elaborar programes que fessin coses normalment considerades «intelligents» tals com jugar a escacs, raonar sobre figures geomètriques o seguir l'esquema de raonament d'algú quan fa una demostració matemàtica. No volien fer màquines ni teories; només programes. (I de «cervells» o sistemes vius, no en volien ni sentir a parlar.) La idea era simple, i el projecte proposat (que McCarthy va batejar, sense donar a entendre res en especial, *intelligència artificial*) incloïa el retrobament periòdic del grup, cada dos o tres anys, per avaluar els programes que cadascú havia construït.

Per ser justos amb tothom, caldria afegir-hi que (1) a l'inesperat èxit del projecte hi va contribuir no poc el fet que un psicòleg militar amic assessorés el grup (pel que fa als tests Stanford-Binet d'intelligència que l'exèrcit americà aplicava des de feia molt de temps) i alhora fes de mitjancer amb els comandaments de l'exèrcit per proveir el grup de subvencions i entusiasmes relativament pròdigs, i que (2) com sol passar sovint, l'èxit espurnejant del nou grup i del nou concepte va anar estrictament en paral·lel amb el declivi inaturable –i, ara es veu, no tan definitiu com semblava– dels grups i conceptes «equivocats» anteriors, de manera que, com explica Arbib de la seva experiència personal, tots els qui treballaven en el camp i ho feien amb la mentalitat original (ara esdevinguda de cop i volta «antiquada», juràsica, i «anti-IA») cap al 1963-64 van haver de plegar o reconvertir-se: cap editor de revistes gosava exposar-se a l'escarni publicant-los res.

*Episodi 7<sup>è</sup> (1966): El somni finalment realitzat: la resolució, els raonadors i els sistemes experts*

En efecte, a la dècada que va seguir la reunió de Dartmouth l'èxit de l'empresa (la «IA») va ser fulgurant, més encara quan es té en compte que només era un programa temptatiu per desencallar una situació, simplificant-la (o, com a alguns els va semblar, banalitzant-la). A les reunions següents, cada equip aportava resultats sorprenents: programes que eren capaços de fer coses «intelligents», tals com, efectivament, les proposades (jugar als

escacs, etc.); però, a més, els programes podien ser discutits i avaluats, se'n podien estudiar les limitacions i –sobretot– es podien posar en comú certs mètodes que havien funcionat bé. Amb el temps, fins i tot s'arribava a situacions tan sorprenents com l'oportunada per Eliza, un programa que merament es proposava simular les reaccions d'un psicoanalista. L'efecte que feia en els qui dialogaven amb el programa era fulminant. Tothom «veia» en el seu interlocutor un ser intel·ligent i fins i tot sensible. La curiosa experiència demostrava la tendència que tenim d'atribuir intel·ligència en els altres (o, simplement, de veure en qualsevol cosa que parli, o que ho sembli, «persones» com nosaltres), i sobre la vacuïtat que poden tenir alguns intents de «crear» intel·ligència a partir de màquines.

Cap a la fi dels seixanta, el cabal de tècniques de què es disposava era impressionant. Els diferents programes, una vegada passat el fort impacte inicial que tots ells provocaven en els observadors, potser no n'eren gaire, d'«intel·ligents», però eren una abundant, i relativament barata, matèria d'experimentació. D'aquesta època són els primers manuals sobre IA que sistematitzen per primera vegada un corpus científic com més va més sòlid. Exemples: *l'heurística*, la *planificació* i la *metodologia dels espais d'estats* i de la subdivisió recursiva de problemes en subproblemes (metodologies que al principi només eren adaptacions dels *arbres de decisió*, de moda entre enginyers i economistes dels anys 50) han acabat essent aportacions metodològiques típiques i sòlides de la IA. I no cal tornar a dir que, paral·lelament a l'èxit dels uns, corria el descrèdit dels altres: en aquesta època esmentar, a propòsit d'intel·ligència, éssers vius, cervells, sistemes complexos o mòduls interactius s'havia tornat una manera segura de merèixer una mirada de compassiu menyspreu.

Una de les tècniques més reeixides va ser, per cert, l'anomenat «mètode de resolució» del lògic anglès Alan Robinson. En realitat, no era res que hagués nascut dins de la IA. Hi era aliè, i venia d'una altra cultura, la pròpiament lògica (la que hem descrit sumàriament abans i, que, engegada amb Llull, va de Hilbert a Turing passant per l'inexcusable von Neumann). Les dèries dels lògics no hi tenien gran cosa a veure, amb la «IA», però aquesta va fagocitar el problema i la seva solució (rebatejada *theorem proving*) i se'n va

aprofitar plenament. De fet, la mecanització del raonament lògic era la prolongació i culminació de la tradició purament lògica de l'escola de Hilbert, entestada com hem dit ja des dels anys vint a mecanitzar el raonament (tot i que per ser justos hem de dir que aquesta cabòria arrenca de Lull i passa per Leibniz, com expliquem en un altre lloc). El mètode de Robinson, del 1964, resolvia de manera satisfactòriament pràctica i efectiva el problema arque-típicament lògic, que és aquest: donat un nombre indeterminat de premisses (per exemple, tot el nostre coneixement que tenim d'«experts» sobre el món), deduir quina conclusió se'n treu, què cal decidir (si es tracta d'accions a prendre) o què podem predir dins una situació concreta... vaja, tot allò que sabem fer quan sabem, pensem i decidim, sobretot en terrenys que dominem. Un exemple extrem d'això és el del matemàtic que dedueix un teorema i el demostra, o l'expert en alguna abstrusa matèria que dedueix vés a saber què després de complicats càlculs (un metge, un geòleg o qualsevol professional amb experiència). (Però no cal anar tan lluny: el cas més normal som nosaltres mateixos quan pensem sobre les conseqüències del que fem o el que diem.) Doncs bé, aquest famós problema a mecanitzar («el somni de Lull», segons Leibniz) es pot considerar definitivament resolt en principi des de l'any 1964.

*Episodi 8<sup>e</sup> (1976): Morir-se d'èxit (però de manera no gaire intel·ligent)*

Després de –i com a conseqüència de– la resolució, un dels grans èxits de la IA dels seixanta va ser la creació de programes «raonadors». La qual cosa va il·lustrar, de passada, com era d'enganyosa la nostra percepció intuïtiva de la intel·ligència humana, segons la qual la capacitat de raonar abstractament és vista com una facultat «superior» (i vet aquí que ara resultava més fàcil de materialitzar que d'altres facultats considerades generalment com a no tan humanes, com ara les de reconèixer cares, o olors). Doncs bé, en IA, per millorar la capacitat deductiva dels sistemes raonadors es va proposar concentrar els coneixements (a partir dels quals s'havien de deduir les conseqüències lògiques) a temes molt concrets, tancats. El resultat és que s'obtenien sistemes que raonaven si fa no fa

com ho faria un expert humà en el tema. El sistema assimilava els coneixements rellevants, expressats en forma de «regles» (del tipus «Si passa això, aleshores fes allò») i en deduïa conclusions, o bé determinava accions a prendre. Els primers *sistemes experts*, com es van dir aquests programes, tractaven de temes restringits com ara el diagnòstic de certes malalties a partir de símptomes (el Mycin) o la predicció de dipòsits de minerals a partir de dades geològiques (el Prospector). Doncs bé, aquest últim va ser catapultat a la fama periodística, i amb ell els sistemes experts en general (i de retruc tota la IA), perquè a primers dels anys setanta va predir –i ho va encertar– un jaciment de molibdè valorat en molts milions de dòlars.

Després d'aquest fet, la IA ja no va ser mai més aquella empresa científica progressiva i modesta que havia estat. Una vegada esdevinguda titular de diari, va sofrir una mutació. Empreses de totes les mides sorgien com bolets per oferir al mercat *sistemes experts* de tota mena. Amb tot plegat, els diaris van començar a anar plens d'*intelligència artificial*, com si aquesta cosa de títol altisonant i fatxenda fos una invenció recent. I per si fos poc, els japonesos ho van acabar de complicar anunciant, a primers dels anys 80, una *cinquena generació* d'ordinadors, que serien essencialment sistemes experts santificats i generalitzats a tots els camps imaginables.

No cal dir que els sistemes experts són el màxim exponent de la intel·ligència vista com a qüestió estrictament lògica, plantejada exactament com ho faria un lògic o un programador de PROLOG. Els components sempre són els mateixos: una descripció de la realitat (el que un lògic en diria una *teoria* i en IA se'n diu *base de coneixements*), un mètode d'inferència o deducció que podem aplicar mecànicament, i unes conclusions, que determinen (a) nous coneixements sobre la realitat, (b) prediccions sobre el seu comportament o (c) accions a fer-hi per satisfer els objectius propòsits inicials. Com es veu, tot molt racional (potser massa i tot). La feina d'un expert queda reduïda a la d'un filòsof (que, com és sabut, no té mai pressa ni cap mena de necessitats o escassetat de recursos). Avalats per la netedat i èxit del plantejament no va faltar qui, com Newell i Simon, afirmessin que aquests sistemes «eren» intel·ligents, tant com nosaltres, i que –recíprocament– nosaltres (incloent-hi la nostra consciència i llibertat aparent) no érem funci-

onalment altra cosa que programes. Aquesta tesi, tècnicament dita *hipòtesi dels símbols físics* o *del nivell cognitiu* (o també *IA forta*) ha estat qüestionada per informàtics mateix com ara Stuart Dreyfus, que diu que el que Newell proposa com a expert és el mateix que nosaltres anomenem buròcrata (en el pitjor sentit de la paraula) o conductor d'autoescola: tots dos actuen aplicant regles de manera rígida, inexperta i àdhuc estúpida, i que n'hi ha prou a considerar què fa realment un autèntic expert (un decisor d'alt nivell o un conductor de fórmula u, respectivament). Només cal veure'ls actuar amb soltesa, experiència i autoritat. I, tanmateix, no solen recordar les regles inicials i ni tan sols foren capaços probablement d'explicitar-les de manera coherent. De fet, la impressió que fan els nostres «experts» artificials és més aviat aquesta: rigidesa general, pobresa intel·lectual i de recursos, falta d'iniciativa i d'autèntica comprensió de la situació o del món, etc.

### *Episodi 9<sup>è</sup> (1986): Aparició de la paret*

Havíem deixat la IA el 1966 en una fase ascendent, i aviat triomfant. Però si els anys 1956-76 van ser de creixença i la dècada 1976-86 l'esplet de l'èxit, cap allà a l'any 86 es començava a respirar una sensació d'atzucac. Què havia passat? Essencialment dues coses:

1. L'èxit dels sistemes experts semblava indicar que la idea de codificar els coneixements a base de regles i de deduir-ne conclusions (raonar) era un idea no sols bona sinó d'aplicació general. Ara bé, inevitablement sorgien alguns dubtes:
  - Tots els camps del coneixement s'hi presten, a ser codificats per regles? Resposta (optimista): sí, si ho fem amb molt de compte. (D'aquest «fer-ho amb compte» se'n diu *Enginyeria del coneixement*).
  - En certs camps del coneixement (per exemple: el comportament lingüístic del parlant d'un idioma) ser «expert» obliga a retenir de l'ordre de vint, potser quaranta, mil regles, la qual cosa és desastrosa pel que fa al rendiment d'un sistema que les hagi de tractar. Resposta (japonesa): farem màquines ultraràpides que processin un nombre elevadíssim de regles

per segon. Resposta complementària: organitzarem les regles de manera que s'hi pugui «navegar» per accedir ràpidament a qualsevol fragment del coneixement. (De les tècniques usades en aquesta organització se'n diu *Representació del coneixement*).

Ara bé: si ja ens vàiem ofegats per una allau de regles, ara es tracta d'organitzar-les i controlar-les (afegint un nou problema al que ja teníem). Com prendre-s'ho? Molts IA-aïres, enfrontats cada dia a la complexitat creixent dels programes, tenien la sensació d'haver arribat al límit de les possibilitats. I d'altra banda, les respostes «optimistes» semblaven una crida a la fe, l'equivalent tècnic del «No t'amoïnis, això no és res!». No és estrany, doncs, que més d'un pensés que havíem topat amb una paret i que el que realment calia era fer un gir copernicà complet.

2. La segona cosa que havia passat ja abans del 1986 és que es palpava una sensació de frustració no ja pel que feia a la construcció o complexitat dels programes sinó pels resultats que exhibien. Si els nostres programes eren intel·ligents, bé ho hauria de notar tothom, això. I, en canvi, la sensació dominant era que els nostres programes no n'eren, d'intel·ligents. La intel·ligència que mostraven era, pròpiament i a tot estirar, només la del programador, i la inherent als mètodes usats en la seva construcció. I, a més, arreu apareixia el molest fenomen de la «fragilitat» (*brittleness*): semblava el programa «feia bondat» intel·ligent, però fàcilment arribava el cas que, perquè inadvertidament fregàvem o depassàvem els límits previstos d'aplicació o pel que sigui, el programa es comportava amb una estupidesa indissimulable o feia alguna bestiesa incomprensible que palesava l'absència d'un mínim de sentit comú. (Solia passar que el sistema es quedava completament «tallat», sense oferir cap mena de sortida, quan passava algun imprevist, encara que fos perfectament raonable i l'alternativa natural a oferir fos senzilla i òbvia.) En resum, era molt fàcil i freqüent que el sistema «ensenyés el llautó». En aquests casos, la sensació de desconcert i decepció de l'observador humà (que imperceptiblement i inevitablement havia assimilat el comportament intel·ligent de la màquina o programa a l'esperable en humans en condicions semblants), solia ser total,

espectacular i difícilment recuperable. La resposta optimista a aquest problema («També els humans ensenyem el llautó quan estem fora del nostre ambient») no era gaire satisfactòria. I encara més: com va dir una vegada Roger Schank (un dels grans pontífexs de la IA) aquells mateixos anys a Barcelona (en una conferència a la Generalitat), resulta penosíssim constatar la impressió d'estúpidesa irrecuperable que fa veure un programa insistir milions de vegades a repetir un mateix resum de premsa (sense «avorrir-s'hi» ni «queixar-se'n»!) o la seva ineptitud a detectar la banalitat d'una conclusió que clarament no porta enlloc, o la incapacitat de «veure» l'interès d'algun fet nou. De fet, és clar, no hi ha res a dir: el programa només fa el que li han dit que fes; però difícilment es pot qualificar d'intelligent un comportament que més aviat recorda el d'un insecte repetint enfollit un a mateixa acció demostradament estèril o poca-solta.

Davant de tot això, i de la desorientació i sensació d'atzucac que va provocar en molts IA-aïres, la pregunta era: com sortir-ne? Hi havia, és clar, una sortida evident: la que consistia a no veure que hi hagués cap crisi i tirar pel dret amb les solucions tècniques proposades: augment –i domini– de la complexitat creixent dels programes, augment de la capacitat i del rendiment de les màquines, i aprofundiment dels nous camps proposats (*Enginyeria el coneixement* i *Representació del coneixement*). Aquesta és la sortida que va adoptar si fa no fa tothom, ja que permetia anar tirant. Però n'hi havia, almenys, dues més, que partien de la hipòtesi de crisi. Van ser, d'una banda, el *Connexionisme* (per dir-ne alguna cosa); d'altra, la d'un grup heterogeni d'investigadors europeus i americans insatisfets amb la situació (i la feina dels quals encara ara no sabem en què acabara, o si portara enlloc).

*A) El Connexionisme (o Xarxes neuronals, o Neuroinformàtica, o IA subsimbòlica)*

Partia d'una idea bàsica: reconèixer una cara li costa al nostre cervell una dècima de segon (= 100 mil·lisegons). Doncs bé: si les nostres màquines tan ultraràpides no són capaces –amb les nostres sofisticades tècniques de programació– de fer una cosa tan senzilla i universal en tan poc temps i, en canvi, el nostre

cervell ho fa sense problemes (i el nostre cervell, cal remarcar-ho, és una màquina molt lenta i que tendeix a cometre errors), això vol dir que hem de canviar de tàctica, almenys quan vulguem fer coses com ara reconèixer cares (o veus, o situacions). Com fer-ho? El cervell ens en suggereix una manera: creem un conjunt d'unitats (diguem-ne «neurones») i connectem-les massivament: tot seguit, fem funcionar el sistema resultant i experimentem amb els resultats. Doncs bé: els resultats de seguida sorprenden, i molt. Hem pogut així construir sistemes («xarxes neuronals», se'n diuen, amb un abús de llenguatge) que sense ser gaire complicats són capaços de recordar moltes diferents imatges (i qui diu «imatges» podria dir «sensacions» o «situacions»), i fins i tot seqüències senceres d'imatges successives. Aquests experiments com a mínim aconseguen impressionar tothom que els veu. Ningú no sap si tot això porta en direcció a la intel·ligència, o si només és una sèrie de fenòmens que presenten una vaga analogia amb els fenòmens cerebrals. Sigui com sigui, ja es va veure de seguida que algunes característiques dels sistemes connexionistes eren noves:

- Memòria «distribuïda»: la memòria aquí ja no és un «receptacle» al qual s'accedeix sinó un conjunt d'«estats» que el sistema «recorda» i que és propi de tot el sistema, globalment, i no localitzable.
- Memòria «associativa»: l'accés a un record concret no es fa per adreçament (=dir on trobar-lo) sinó per contingut (=donar-ne un esquema bàsic, «evocar-lo»).
- Facilitat d'abstracció: el reconeixement actua com un procés d'*autoorganització* de la informació que accentua i aïlla trets bàsics d'una imatge.
- Paral·lelisme implícit en el procés: totes les unitats («neurones») col·laboren en qualsevol tasca.
- «Degradació gràcil»: el rendiment no cau pas en picat pel fet que una part –tot i important– del sistema deixi de funcionar. (A comparar –favorablement– amb la total intolerància a l'error que tenen els sistemes tradicionals i que provoquen la interrupció immediata i el bloqueig total del procés en curs.)



- Absència de «programa»: el sistema no funciona pas perquè se'l programi prèviament (de fet aquí no hi ha «programador») sinó que observa l'input, s'«autoorganitza», i actua en funció del procés viscut (que resta sempre obert a noves experiències).

Aquestes propietats, absents dels sistemes tradicionals de la IA, configuren una metodologia (també una nova IA? o una anti-IA?) que potser és resumible en l'última característica citada. És a dir:

Sistemes tradicionals  $\Leftrightarrow$  **Programació**

[prèvia, per un *programador*]

Sistemes connexionistes  $\Leftrightarrow$  **Autoorganització i Aprenentatge**

[a partir de l' *experiència* adquirida]

El connexionisme no se sap si avança cap a alguna cosa, però almenys representa una alternativa a la IA tradicional. O bé la seva complementació, que podríem descriure succintament així:

IA tradicional (**IA simbòlica**)  $\Leftrightarrow$  Fenòmens cognitius de  $> 100$  ms

—————↑————↓———— **barrera dels 100 ms**

Connexionisme (**IA subsimbòlica**)  $\Leftrightarrow$  Fenòmens cognitius de  $< 100$  ms

Aquí el problema és, ara com al 1986, com comunicar els dos camps a través de la barrera. La IA tradicional treballa amb programes (que tenen accés a una memòria que és la dipositària del coneixement del sistema) i actua sobre objectes que són essencialment símbols, mentre que el connexionisme treballa amb xarxes (el diferents estats de les quals determinen el coneixement del sistema) i actua sobre estats globals d'energia. Com s'ha de fer doncs per anar cap amunt (=convertir estats en objectes simbòlics)? i cap avall (= l'invers)? De moment ningú no ho sap veure. De fet, les habilitats de cada sistema semblen completament diferents i incompatibles.

### B) La resta

No tothom s'ha convertit al connexionisme, ni tampoc es pot dir que la gent de la IA sigui feliç d'haver de triar entre dues alternatives com si hagués de triar entre dues religions. Un element implícit

i positiu del connexionisme és que ha trencat l'estat hipnòtic de pensar que la IA ha de ser vista necessàriament des d'una sola perspectiva (que és, encara, l'original de 1956: IA = fer programes). El connexionisme permet imaginar coses que no són ni tan sols «programables» en el sentit habitual. Només per això ja mereixeria ser considerat com un bon remei contra l'encarcarament. Però és que hi va haver uns quants investigadors no connexionistes que van creure que la «crisi» de la IA era una magnífica ocasió per tirar endavant en la comprensió de la intel·ligència. En citarem només cinc: Steels, Brooks, Hillis, Holland i Toffoli (i, de passada, profitarem per esmentar Gell-Man i Feynman).

Luc **Steels**, europeu de Brusselles (i vell conegut dels ACIA-aires), o Rodney **Brooks**, australià de l'MIT, treballen ja fa temps sobre nous esquemes que permeten actuar sota qualsevol metodologia: tradicional, connexionista o cap de les dues. Es basen en la hipòtesi que el càlcul no és necessàriament simbòlic ni el contrari, i que sovint és l'efecte i resultat d'un procés cooperatiu. Exemple: una fila de formigues «calcula» automàticament la trajectòria mínima entre el menjar i el formiguer, que sol ser –si res no ho impedeix– una recta; ara bé, si colloquem una pedra que interfereixi la trajectòria calculada, les formigues la recalcularan espontàniament de manera que la nova solució eviti l'obstacle i continuï essent mínima. Aquí és ociós parlar de «qui» calcula la trajectòria: la calcula algun «programador»? es calcula «sola»? és realment un «càlcul»? o més aviat es tracta d'«autoorganització»? Millor no capficar-s'hi gaire. Ah! i no es tracta pas d'un afer de formiguers, només: allà on hem dit formigues i pedres podíem haver dit robots i obstacles, o subjectes i situacions.

Daniel **Hillis**, a part ser gendre de Minsky, és el creador de la fa temps famosa Connection Machine, un ordinador paral·lel ultraràpid, d'estructura semblant a un reticle. Hillis va observar que la dinàmica de càlcul de la seva màquina presentava fenòmens curiosos: de vegades, en segons quins càlculs, l'activitat calculatòria es concentrava en una part del reticle, en d'altres es distribuïa a l'atzar o en oscil·lacions periòdiques, o caòtiques. Tot això volia dir implícitament que hi ha molt que no sabem de què és un «càlcul» i de les sorprenents maneres en què un càlcul es pot executar en màquines d'arquitectures no-convencionals.

I encara hi podríem afegir els càlculs genètics, dits així perquè –deixats anar– ells sols s'autoorganitzen o presenten espontàniament fenòmens tals com mutacions, seleccions, variació, creixement, supervivència o extinció que recorden extraordinàriament els comportaments evolutius clàssics de la Biologia. Que aquests processos recordin alguna cosa no treu de cap manera que siguin capaços de calcular, i calcular bé i efectivament. Aquesta idea, que ja porta dècades pensada i elaborada pel veterà John **Holland**, avui està en un procés de creixement imparable. Com ho estan tot un seguit d'iniciatives, treballs i idees que planen sobre Santa Fe, institut creat *ad hoc* justament per reunir-hi gent d'alt nivell i diversa procedència (com ara el celebèrrim Nobel de Física Murray **Gell-Mann** i d'altres de l'alçada corresponent) perquè hi discuteixin i treballin sobre complexitat i autoorganització de processos en general i –no caldria sinó– de càlcul en particular. (Potser serà una sorpresa per algú saber que hi ha un grup de recerca que s'hi dedica a la UPC mateixa, el cap del qual és el físic Ricard Solé, un dels pocs habituals selectes de Santa Fe.)

L'americà Tommaso **Toffoli** i el seu grup de l'MIT treballa de fa temps sobre un model de càlcul bastant diferent de l'habitual (si bé és l'inventat per Von Neumann als anys 40), els «autòmats cel·lulars», que permeten, com la Connection Machine, experimentar comportaments i dinàmiques de càlcul –incloent-hi els «genètics»– tan diferents dels considerats corrents i tan sorprenents (i classificar-los i comparar-los, com ho ha fet Wolfram i tants d'altres). És significatiu que un altre celebèrrim Nobel de Física, el gran Richard **Feynman**, poc abans de morir, quedés encantat amb el que li explicava Toffoli sobre els càlculs tan no-estàndards que anaven apareixent pertot arreu i que hi dedicà les últimes intervencions de la seva vida, en què, tornant a **Turing**, un altre físic fascinat per la Biologia i els fenòmens quàntics, es demanava si el càlcul no és potser essencialment un procés físic (i tal vegada quàntic) estudiable com a tal. Com poc més tard de Turing va fer **von Neumann**, que es demanava si en comptes d'aplicar-nos a definir el que carai deu ser un càlcul (cosa que ell, com Turing, havia, per cert fet més bé que ningú) no devia valdre més la pena estudiar-lo físicament (segons ell, com una part de la Termodinàmica).

En fi, que fóra d'enzes negar que hi ha una insatisfacció difusa,

una intranquil·litat general i que sorgeixen idees en totes direccions. Cap on porta, tot això? Deu anys més tard que hi van aparèixer els primers símptomes d'existència d'alternatives, no ho podem pas dir segur. Només es pot afirmar que la IA encara té molt de camp per córrer, i l'última paraula encara no l'ha dita ningú. En tot cas, faríem bé de preparar-nos per rebre sorpreses. Com diu McCarthy, si la Física no ha resolt certs problemes fins al cap de cent anys d'estudiar-los, la IA –que Déu n'hi do les sorpreses que ens ha donat en 40 anys– per què havia de resoldre els seus en menys?

*Apèndix (1996-97): Què pot voler dir intel·ligència(?) artificial(?) i quin sentit pot tenir per nosaltres (si és que en té cap?)*

Dit tot l'anterior i recapitulant la història, té sentit insistir en la idea històricofundacional que la IA és un assumpte de lògics? Vist com han anat les coses, és evident que sí: si no arriba a ser per ells no tindríem IA, o no la tindríem com la tenim. I té sentit que la IA continuï lligada a la Lògica? La meua opinió particular, com es desprèn del to dels darrers paràgrafs i com ja vaig expressar clarament a l'últim número d'aquest Butlletí, és que no. Però jo ara vull anar més enllà d'aquesta qüestió de filiació històrica, d'amors i desamors, de paternitats i divorci. La pregunta que em faig, i que ja em vaig fer a partir dels primers números a propòsit de les relacions entre intel·ligència i creativitat, és la següent: Té sentit parlar d'«intel·ligència artificial»? És a dir: és pot dir pròpiament intel·ligència allò que fem? i què vol dir artificial en aquest context? No parlo del sentit que pugui tenir una ciència (la IA) que l'estudiï, sinó de l'existència i del concepte mateix del seu hipotètic objecte d'estudi: unes possibles entitats que siguin artificials i –malgrat (?) això mateix– intel·ligents. Des de les definicions humorístiques de la IA («Pròtesi per a l'estupidesa natural») fins a les serioses i meditades («Conjunt de sistemes que permeten realitzar tasques que, si fossin fetes per éssers humans, serien considerades intel·ligents», a la qual cosa es pot –i se sol– afegir «L'estudi dels quals ens permet avançar

en la nostra comprensió del que és la intel·ligència.»), sempre hi ha un element subjacent i implícit que hi apareix i que es pot resumir en aquesta frase: «Intel·ligència és el conjunt d'aptituds intel·lectuals que **no ens sorprenen** quan apareixen en una persona, però que **ens sorprendrien** si apareguessin en màquines». Aquesta definició temptativa, que té poques pretensions i la particularitat de ser inherentment relativa, cospa molt més bé la idea que l'home ha tingut sempre de la intel·ligència i explica que al llarg de la història s'hagin qualificat d'«intel·ligents» –gairebé per definició– coses que en aquell moment sorprenien; en efecte, s'han anat considerant successivament «intel·ligents» –fins que pel mateix fet d'adaptar-s'hi ja no sorprenien i, per tant, deixaven de ser-ne aparells com el rellotge (el primer giny que funcionava «sol», sense aportació humana continuada, fet que tenia bocabadats els filòsofs fins ben entrat el s. XVII, quan encara s'hi referien admirats), la impremta (que es va qualificar d'«escriptura artificial» –en el sentit que les lletres «s'escrivien soles»–) o els autòmats setcentistes de Vaucanson (que, vistos a Suïssa pels romàntics anglesos, els van produir un impacte tan fort que van inspirar en un d'ells, la Mary Shelley, la idea del Dr. Frankenstein), per no parlar de la impressió que les meravelles programables que surten cada dia al mercat fan en la gent d'avui, tot i el costum i les ganes de deixar-nos sorprendre que tots hem adquirit amb el temps i la tecnologia actual. Caldria recordar, potser, que –en expressió d'Arthur Clarke– «Qualsevol tecnologia suficientment desenvolupada és indistingible de la màgia.»... i sembla que, de set de màgia, no en tenim mai prou.

Ara bé, si la definició que hem donat («Intel·ligència = qualitat que ens sorprèn –en una màquina–») és relativament acceptable i prou precisa, té la particularitat de fer que la IA, com a ciència (i tècnica), estigui condemnada a evolucionar sense parar. Perquè, si el seu objecte d'estudi és sempre canviant (ja que, un cop una cosa ja no ens sorprèn –perquè ens hi hem acostumat–, deixa de ser intel·ligent, i ja no és, doncs, objecte d'estudi), això vol dir que la IA sempre estudia coses diferents, i que allò que estudiava ahir avui ja no hi pertany. Doncs bé: és això justament el que ha passat amb la IA d'ençà de la seva fundació l'any 1956: la llista dels seus temes d'estudi és més volàtil que la de cap altra ciència. Durant

els anys 60 i 70 la IA estudiava, per exemple: (A) com jugar als escacs, (B) com resoldre problemes plantejats de manera prou general, (C) com reconèixer formes més o menys definides, o (D) com entendre situacions realment complicades (E) mitjançant una bona descripció del que en sabem i (F) fent un seguiment de l'actuació de cada agent autònom present en la situació. Alguns d'aquests temes encara estan a l'agenda de la IA dels anys 90 (si bé amb noms canviats i objectius més delimitats, o simplement modificats). Dels exemples citats, l'A ha passat a ser un cas particular del B; B s'ha acabat convertint en un tema general, l'«Heurística», que s'ha autonomitzat bastant de la resta i té els seus propis mètodes i objectius; C s'ha independitzat de la IA i avui ja no s'explica necessàriament dins d'aquesta matèria; D ha donat lloc a tècniques d'estructuració i control de programes (informàtics) complexos que avui formen part ja no de la IA sinó de la Informàtica corrent, en la qual la descripció de situacions realment complexes (=E) ha originat el nou camp de la «representació de coneixements» i ha enriquit el de l'estructuració de dades, mentre que els «agents» (=F) s'han convertit en «objectes» construïbles per qualsevol programador. I així tot. Des d'aquesta perspectiva, la IA es pot considerar merament com la frontera d'avanç de la Informàtica en un moment donat.

I pel que fa als problemes **filosòfics** que presenta el concepte «intelligència» quan li canviem el subjecte i l'apliquem a entitats no-humanes, podem fer dues coses: (a) preocupar-nos-en, o (b) desestimar-los. Ara bé, qualsevol que sigui la decisió que prenguem, cal reconèixer que allò que dóna un punt de visceralitat a la discussió és el fet que representa una transgressió de límits: estem acostumats a aplicar «intelligent» a persones, i les persones signifiquen coses que de cap manera signifiquen –o volem que signifiquin– les màquines; quan passem l'adjectiu d'un col·lectiu a l'altre, l'operació, lluny de ser innocent, té ressonàncies i implicacions profundes, perquè «humanitza» o «antropomorfitza» coses que **no** estem gens acostumats –o disposats– a veure en aquesta llum, ni generalment a acceptar-ne les conseqüències. Ens hauríem de plantejar què és el que realment volem fer amb les màquines: fer clons mecànics de nosaltres mateixos? (Té sentit això? i si en té, aleshores quin sentit passem a tenir nosaltres?) Quan fem roses ar-

tificials, amb realisme visual i tàctil i perfum inclosos, segurament volem tenir aspectes de la rosa real sense haver d'estar pendents de les seves fragilitats (reg, pansiment, efimeritat), però ningú, confio, no ho confondrà mai amb la cosa de debò. De fet, és això el que fem amb els nostres ordinadors (com abans havíem fet amb les màquines motrius): els encomanem coses que fins ara només fèiem nosaltres i els hi fem fer amb la plena confiança que, si la solució està clarament aïllada i resolta, la màquina ho farà amb molta més eficiència que nosaltres i sense defallir. I no cal veure-hi més enllà ni sentir-se'n amenaçat: ho hem fet amb els cotxes i amb els PC, que hem integrat de manera natural en les nostres vides: sabem què n'esperem (i el que no) i ara són com de la família –de vegades fins i tot massa. I no val a posar-nos nerviosos i sentir-nos amenaçats pel que ens pugui passar. Per això, potser l'actitud més sana és la del físic-informàtic Dijkstra, habitual de la boutade; quan se li demana «Les màquines, pensen?» contesta, tot arronsant les espatlles: «Els submarins, neden?». I és clar que aquesta última pregunta, que nosaltres probablement mai no considerariem digna d'atenció, és imaginable que pugui ser motiu de preocupació i debat filosòfic profund... per als peixos (si és que consideren envaït el seu terreny). De la mateixa manera que no ens cal tenir cotxes que tinguin cames o avions que siguin com ocells, mentre ens portin allà on volem, ni ens importa que allò que fan se'n digui caminar o volar (o no), per què hauríem de voler màquines intel·ligents que fossin com nosaltres, o preocupar-nos de saber si el que fan és exactament «pensar » o no? Els xinesos, experts en proverbis, en tenen un que fa: «Val més encendre una espelma que maleir la foscor». I és que és força més productiu preparar un bon futur que *demonitzar-lo* i espantar-se'n abans d'hora.

De fet, em sembla que és el que fem la majoria: demanem a les màquines allò per al que les hem fetes, les usem, les dominem, les integrem al nostre paisatge, i fins i tot les estimem o les odiam. Penseu només com patim pels nostres cotxes, com els tractem, com ens hi sentim quan hi som a dintre... ! I qui diu cotxe, diu PC: heu sentit mai la desolació i la desempара que ens envaeix quan constatem que sense ells ja no sabem escriure un text sense un teclat al davant? També la impremta va deixar desolats –i al carrer!– un bon grapat de copistes orgullosos de la seva esplèndida

i meritòria calligrafia, que ningú excepte ells ha trobat a faltar mai gaire, francament; el que per ells va ser una amenaça real –i tant!– per als altres va ser un guany incomparable.

Només alguns es preocupen d'esbrinar si les màquines ens amenacen o se'ns menjaran. Per a alguns d'aquests professionals de la preocupació el tema és tan estirable i elàstic que dona per passar-hi tota una vida. A la resta només sol preocupar-nos de tant en tant, quan toca deprimir-se per alguna cosa i no es troba cap altre tema. De fet, un mercat per a aquesta mena de cabòries ja hi és. Tant, que alguns tenim la sort que fins i tot ens conviden de tant en tant a fer-ne conferències: perquè després diguin que preocupar-se per la IA no té cap interès!





# *Llull com a informàtic avant la lettre*

**R**AMON Llull va ser l'origen –al segle XIII!– de conceptes informàtics bàsics. Les seves idees han arribat fins a nosaltres –i ens n'hem pogut aprofitar– gràcies sobretot a Leibniz, que va admirar i estudiar profundament aquest primer filòsof en llengua catalana, en qui va saber reconèixer la base de la Lògica moderna (que, sense ell, segurament fóra ben diferent).

BUTLLETÍ DE  
L'ACIA  
NÚMERO 10-11  
PRIMAVERA-ESTIU DE 1997

## *Els primers propòsits*

Al pobre Ramon Llull (1232-1316), frare mallorquí de família burgesa barcelonina, n'hi ha passat una com un cove. A les històries de la Filosofia tant el veiem exalçat com ridiculitzat. Tot i haver tingut intuïcions que avui són bàsiques per a tot lògic i tot informàtic, el lloc que ocupa a les històries de la Filosofia, de la Matemàtica o de la Lògica és ínfim. I tot i ser un dels primers filòsofs a basar la Fe (religiosa) en la Lògica, ha acabat sovint considerat una autoritat en Alquímia, en Càbala i en Mística. Se'l pinta com un boig superb i excèntric –n'hi ha prou de veure el que en diu Manin Gardner en un article del 1958–, un garbuix pseudointellectual de pensaments confusos. Però aquests judicis condemnatoris amb prou feines lliguen amb els fets, indiscutibles, com ara que Llull va ser el primer a veure coses que només set-cents anys més tard s'han pogut desenvolupar i entendre. Hom acaba demanant-se, doncs, qui era realment Llull, i què va fer.

Doncs el que Llull es va proposar va ser construir un sistema innovador i complet, i tot seguit anar a convèncer els intel·lectuals de París de l'època del fet que les seves originals, sorprenents i difícils idees bé valien la pena... sense sortir-se'n gaire, val a dir. En la seva primera visita a la Sorbona, el 1289, senzillament no

el van entendre. La seva combinatòria no s'assemblava gens al mètode de raonament que calia aplicar, canònicament, a l'anàlisi lògica (i que era l'«encadenament causal»). Quan hi va tornar el 1309-11 amb un sistema més entenedor el van rebre amb simpatia, però amb poca acceptació real. Alguns hi van veure, com era el cas, la fonamentació de la fe exclusivament en la lògica, i de l'enteniment humà en la raó (una idea molt moderna, que s'oposava a la dels revelacionistes i dels místics i que havia de culminar en els racionalistes i en la Ciència). Però després de la seva mort es va acabar la simpatia, i el lul·lisme va caure víctima de la Inquisició en la pugna que enfrontava franciscans i dominicans al s. XIV. I Llull, que sempre havia primat la lògica per davant de la fe –i ho havia fet proposant idees noves, progressives i rigoroses en la línia del que avui entenem per ciència– va acabar, perversament, convertit en un pensador obscur, reaccionari i confús ridiculitzat pels primers pioners de la ciència moderna (Bacon o Descartes, que en llei li haurien d'haver agraït una bona pila de coses), alhora que, inopinadament, esdevenia l'heroi d'alquimistes, cabalistes i místics en general (pel fet que se li atribuïen obres apòcrifes esotèriques). La crítica condemnatòria que li han fet els «moderns», de Descartes a Gardner, –bàsicament, que el sistema lul·lià és confús– no s'aguanta gaire: de fet, l'embolic entre fe religiosa, motius ètics, apologètica i explicacions naturals era la regla general, més que no pas l'excepció, en la filosofia medieval. I, pel que fa a les seves pròpies confusions personals, és prou conegut que tots els innovadors entenen poc i malament els seus propis descobriments –l'abast dels quals no solen copsar– i que els acostumen a extrapolar sense control. Les innovacions de Llull, que tan estranyes van semblar als seus contemporanis, avui ens resulten ben familiars. Heus-en aquí algunes de les més típiques.

### *Alguns conceptes informàtics*

#### *La idea de càlcul*

Que el raonament lògic és, en certs sentits, càlcul –o, més ben dit, que es pot formalitzar i validar per mitjans controlables– avui és una visió acceptada i clarament explicada pels lògics dels anys 1920

(Hilbert o Herbrand, per dir-ne dos), i efectivament mecanitzada als anys 60. Però aquesta idea ja va ser avançada al segle XVII per Hobbes, que va escriure el 1655 que «El raonament no és altre que càlcul», i per Leibniz, que va pensar el 1658 (i ho va escriure el 1666) que, en el futur, els filòsofs resoldrien llurs disputes com ho fan els comptables, prenent plomes i calculadors (àbacs) tot dient *Calculemus!*. Leibniz deia explícitament que això era ni més ni menys fer realitat el somni de Llull. I ho era. Llull havia avançat aquesta idea el 1274 en observar que, per convertir musulmans (una preocupació obsessiva de l'època), les *disputationes* públiques eren inútils (les dels 1260 havien acabat circularment, sense que ningú convencés ningú). Calia trobar, doncs, un mecanisme que demostrés i generés veritats de manera que, un cop tothom estigués d'acord en els pressupòsits, la mateixa objectivitat del procediment obligaria a tots a acceptar les conclusions. A l'elaboració d'aquest mecanisme va dedicar Llull tota la vida. El seu admirador Leibniz, 400 anys més tard, va explicar la idea bàsica en termes concisos i atractius, d'una manera que el mateix Llull hauria pogut subscriure personalment. A més, el «mecanisme» de Llull no era tan sols un procediment abstracte; tenia un suport veritablement «mecànic», els seus discs rotatoris concèntrics que, avui considerats joguines banals, van ser els primers dispositius mecànics que mai van existir (fig. 1). A partir d'aquest mecanisme elemental, i per simples manipulacions mecàniques, es va crear tota una heurística per la qual es generava combinatòriament una cadena deductiva de veritats, que tot seguit s'explorava i validava.

### *La idea d'un alfabet del pensament*

Quan George Boole, el 1847 i el 1854, va descobrir i formalitzar les «lleis del pensament», el que va fer bàsicament és concebre el «pensament» com un conjunt de manipulacions de conceptes expressats algebraicament. O sigui, com Llull el 1274, tot i que el mallorquí, a diferència de Boole, creia que li calien no sols manipulacions (combinacions) permeses sinó també un conjunt finit de veritats elementals amb què començar. En va dir «dignitats» (per traduir el mot «axiomes» dels grecs) o «principis absoluts», nou en total, més 45 conceptes bàsics addicionals (en grups de nou) que va anomenar

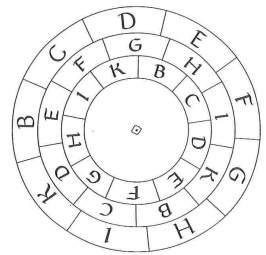


Figura 1: El primer calculador de la història.

«principis relatiu» (que incloïen la consistència i la contradicció), «regles» (que incloïen la quantitat o la modalitat), «subjectes», «virtuts» i «vicis» (fig. 2). Hi va afegir regles manipulatives bàsiques (un precedent del càlcul relacional) i un procediment de validació. Aquest consistia bàsicament a expandir les possibles combinacions i a anar-les seguint fins que, una de dues: o bé dos conceptes es reforçaven mútuament –i així donaven versemblança a la conclusió– o bé apareixia una contradicció –que volia dir que s’havia de negar la hipotètica conclusió. Com Boole, però 550 anys abans, Llull creia fermament que era possible tractar el pensament humà (el raonament lògic) de manera simbòlica i sotmetre’l a procediments normalitzats de control i seguiment objectiu.

Figura 2: The alphabet of the ars brevis.

THE ALPHABET OF THE ARS BREVIS

	<i>Fig. A</i>	<i>Fig. T</i>	<i>Questions and Rules†</i>	<i>Subjects</i>	<i>Virtues</i>	<i>Vices</i>
<b>B</b>	goodness	difference	whether?	God	justice	avarice
<b>C</b>	greatness	concordance	what?	angel	prudence	gluttony
<b>D</b>	eternity*	contrariety	of what?	heaven	fortitude	lust
<b>E</b>	power	beginning	why?	man	temperance	pride
<b>F</b>	wisdom	middle	how much?	imaginative	faith	accidie
<b>G</b>	will	end	of what kind?	sensitive	hope	envy
<b>H</b>	virtue	majority	when?	vegetative	charity	ire
<b>I</b>	truth	equality	where?	elementative	patience	lying
<b>K</b>	glory	minority	how? and with what?	instrumentative	pity	inconstancy

*La idea d’un mètode*

No tots els filòsofs de l’època de Llull se sentien obligats a fixar clarament un mètode –i menys a demanar que fos únic o el més universal possible– i a seguir-lo estrictament. Aquest concepte, plenament modern i que inclou la idea de «procediment efectiu» de Hilbert o la d’algorisme («màquina») de Turing, no s’assemblava gens a res existent el s. XIII, molt abans que Descartes suggerís l’existència i aplicabilitat d’un «mètode» universal (1637). Tot i que de cap manera reconegués el seu deute lul·lià, Descartes –que

coneixia bé Llull– va fer per la filosofia el que el mallorquí havia proposat per a l'anàlisi lògica en general: fixar un conjunt de regles, considerar-les permanents i universals, i seguir-les rigorosament.

### *La idea d'anàlisi lògica*

La idea de Llull era analitzar els conceptes bàsics associant-los uns als altres per veure què passava. Això, per ell, era com penetrar en els secrets del funcionament de Déu i la natura, i així entendre més bé el món (i poder-ne donar una explicació efectiva i objectiva). I si de passada, com a conseqüència del mètode, s'arribava a la Fe (o fins i tot a la Revelació mística), aleshores Llull satisfia el seu disseny últim, el de fonamentar la Fe en la Raó, i de justificar les creences en l'anàlisi lògica. L'originalitat de Llull és que ho aconseguia a la pràctica mitjançant l'execució mecànica de l'expansió iterada d'un conjunt donat de creences inicials (un nucli o «compendi» de veritats) fins que, si arribava el cas, s'hi detectés una contradicció. Postulant l'existència d'un procediment com aquest el que feia Llull era, de fet, anticipar la idea moderna (1955) dels taulers semàntics. (Més avall en tornarem a parlar.)

### *La idea d'heurística i de deducció*

A Llull li interessava trobar noves veritats, i no sols demostrar les conegudes. («Demostrar» volia dir 'ser capaç de convèncer qualsevol d'una veritat donada'.) Aquesta segona part està inclosa en el que normalment anomenem deducció. La primera («trobar»), en canvi, sorprèn una mica. La ciència moderna ha evitat sistemàticament l'anàlisi del procés que seguim per descobrir o inventar. El descobriment s'ha atribuït a la imaginació, al geni o a la casualitat, però ningú no ha pogut explicar, i molt menys controlar, el desenvolupament del procés heurístic, en part perquè només el podem «controlar» a posteriori, un cop ja ha sorgit la cosa: llavors podem verificar si les prediccions que es poden fer a partir d'allò descobert es compleixen o no. Aquesta simple raó era la que va donar Pop per els anys 1950 en la seva polèmica amb Carnap i contra la idea d'aquest d'un «lògica inductiva» que descrivís formalment el fenomen. Encara avui estem orfes d'idees sobre la qüestió fora d'algunes magnífiques intuïcions de Polya

sobre el procés creatiu, i algunes pistes i regles semiempíriques que es basen en l'exploració sistemàtica de totes les possibilitats imaginables. En això consisteix el mètode «morfològic» de Fred Zwicky (consistent a generar associacions de manera combinatòria i exhaustiva) o les seves versions modernes més o menys informàtiques i molt ad hoc que usem en Intel·ligència Artificial i que pretensiosament anomenem «heurística». El que sorprèn més és que un mètode modern com el de Zwicky, amb els seus algorismes i taules d'exploració i aparionament s'assembla moltíssim, fins en aparença externa, a les eines visuals d'en Llull. No cal ni dir que l'Heurística com a ciència, si mai ha existit, avui està en el mateix estat d'indigència en què Llull la va trobar.

### *La idea de sistemes generatius*

De les anticipacions de Llull potser la més sorprenent és la idea de tenir un conjunt finit de regles i un altre, també finit, de veritats –«conceptes bàsics», axiomes o com se'n vulgui dir–, de manera que d'aquests últims es puguin generar un conjunt (normalment infinit) de veritats derivades. Avui descriuríem la cosa més senzillament, i diríem que Llull havia descobert la idea de sistema generatiu. En lingüística d'un dispositiu finitista com aquest se'n diu gramàtica (= conjunt de regles per manipular cadenes d'un alfabet a partir d'uns axiomes inicials) i les cadenes generades són el llenguatge. En informàtica del dispositiu se'n diu màquina i el que s'hi genera és el conjunt de configuracions d'una cinta. Com avui sabem, el mateix mecanisme pot funcionar amb la marxa enrere: la mateixa gramàtica capaç de generar un llenguatge també és capaç d'acceptar-ne o reconèixer-ne les cadenes. O la mateixa màquina que calcula els resultats acceptables també és capaç de reconèixer un càlcul correcte. (Que aquests dos processos duals siguin lleugerament asimètrics en termes de càlcul es, per cert, un corollari del primer teorema d'incompletesa de Godel.) Llull va ser el primer que va veure aquesta dualitat reversible: en les seves paraules, el mateix sistema heurístic que proposava per derivar noves veritats a partir d'un conjunt abreujat (*compendium*) de veritats inicials i que ell anomenava, en el seu català original, «art de trobar veritat» (o *ars inveniendi* quan ho havia de dir en llatí) i

que avui en Lògica anomenem simplement inferència (o *forward chaining*, si volem pedantejar) tenia una qualitat dual que permetia executar-lo cap enrere, de manera que la cosa quedava convertida en un sistema reconeixedor o acceptador que ell anomenava «art de demostrar» (*ars demonstrandi*) i que nosaltres anomenem simplement demostració (o *backward chaining*, o *cerca per objectius* en Intel·ligència Artificial). Així, per a Llull, si mai havíem de demostrar cap enunciat concret no caldria que inventéssim cap nou sistema: el mateix que ens permetia explorar noves veritats seria suficient per certificar la presumpta veritat; la demostració seria ni més ni menys el procediment de certificació mateix.

*La idea de graf*

Llull connectava els seus «conceptes bàsics» amb línies (fig. 3) i prescrivia que, per combinar els conceptes i derivar-ne les conseqüències, calia seguir-les. Això era nou. No pas ara, però; tenim un nom per al dispositiu que Llull va inventar: en diem graf. L'anècdota té dos aspectes interessants: primer, Llull en va donar una varietat dual isomorfa (fig. 4): va reunir tota la informació del graf en forma de taula de doble entrada just el que nosaltres anomenem matriu d'adjacència del graf i, segon, els grafs lul·lians no pretenien representar l'estructura conceptual o taxonòmica d'un objecte o classe (la presentació arborescent dels conceptes ja era corrent a les darreries de l'imperi romà) sinó que Llull les concebia més aviat com una *xarxa semàntica* actual, per ser «seguida», és a dir, executada dinàmicament com si fos un autèntic *programa* descobridor de veritats o l'arbre de decisió (en el sentit informàtic) d'un algorisme o procediment de decisió.

*La idea dels taulers*

Els procediments de «trobar veritat» que Llull va proposar eren dos sobretot. Un anava cap endavant: els conceptes s'hi combinaven (seguint el graf director) i, si es reforçaven mútuament, demostren la conclusió *per analogia*. El dual negatiu consistia en el fet que si mai els conceptes que s'estava manipulant resultaven ser mútuament contradictoris (inconsistentes) la veritat inicialment postulada quedava automàticament refutada, i demostrada la con-

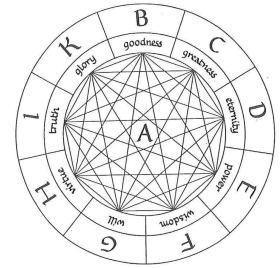


Figura 3: El graf...

BC	CD	DE	EF	FG	GH	HI	IK
BD	CE	DF	EG	FH	GI	HK	
BE	CF	DG	EH	FI	GK		
BF	CG	DH	EI	FK			
BG	CH	DI	EK				
BH	CI	DK					
BI	CK						
BK							

Figura 4: ...i la matriu



trària (fig. 5). Aquesta és la primera vegada que apareix per escrit un a idea semblant a la dels taulers (*tableaux*) semàntics de Beth (fig. 6) del 1955 (o, en ciència, a la idea de refutació de Popper del 1959). (Però aquesta anticipació del s. XIII d'idees actuals no va tenir, a diferència de les altres de Llull, cap influència directa sobre nosaltres.)

Figura 5: Com trobar una contradicció (= «impossible!»)

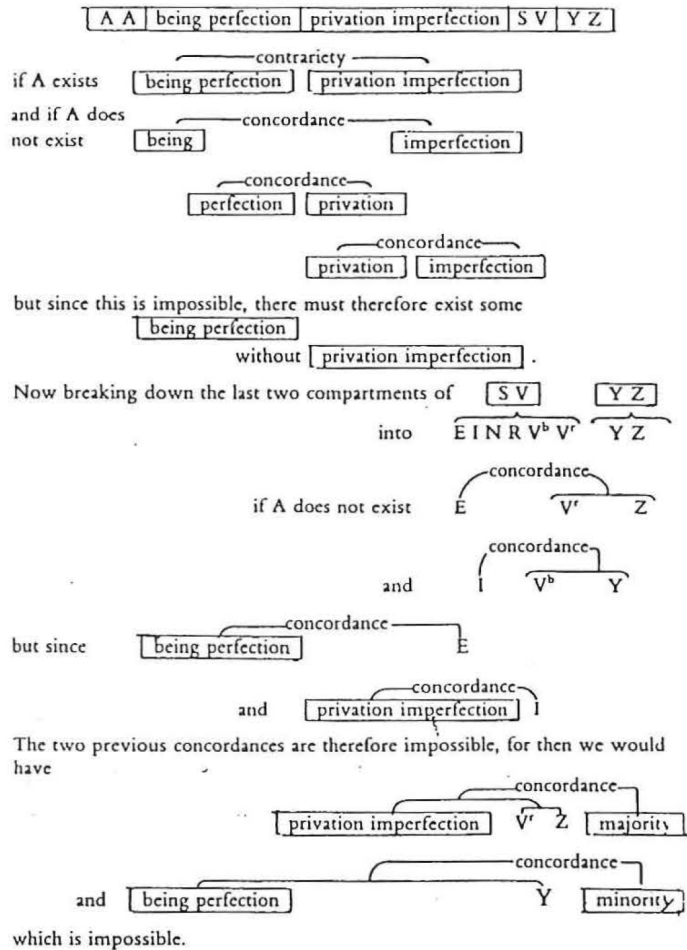


Figura 6: Taulers semàn-  
tics (la versió moderna de  
l'«impossible»)

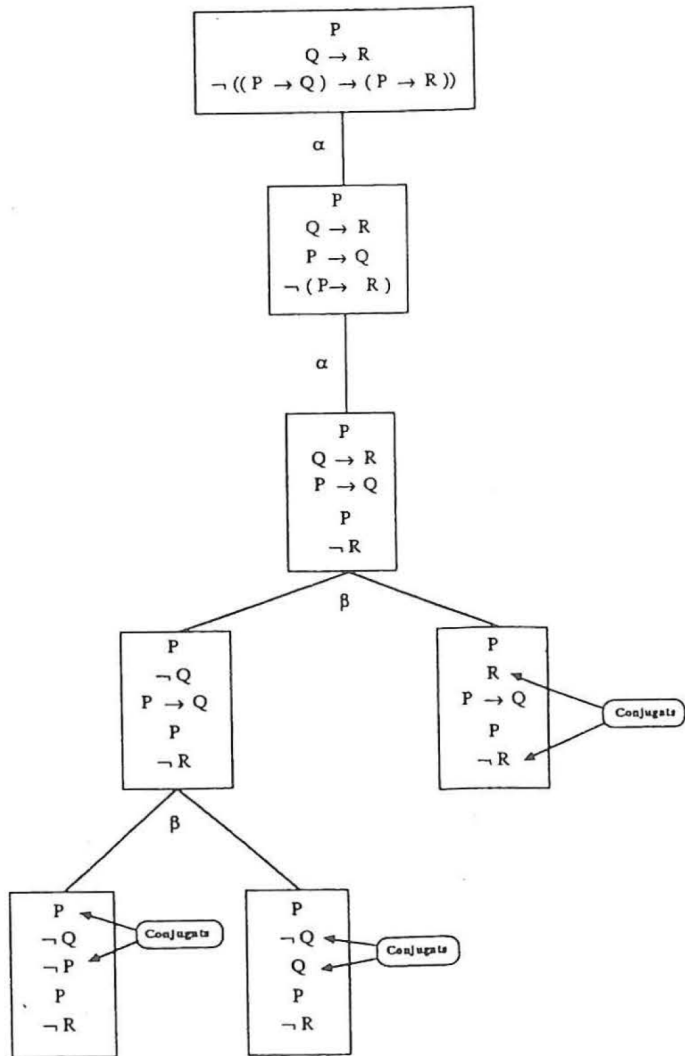
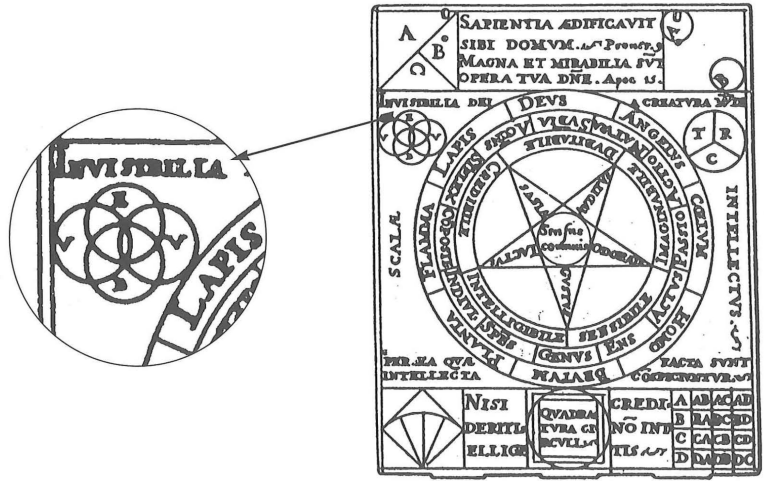


Figura 7: El primer «Venn»



*La idea de xarxa conceptual*

Com ja hem dit, els grafs de Llull no eren ni arbres taxonòmics estàtics ni il·lustracions de com estaven estructurats els conceptes sinó un autèntica xarxa d'enllaços que permetien que l'usuari explorés de manera combinatòria les relacions que hi havia entre els conceptes que es manipulaven. Eren, doncs, una prefiguració dels moderns grafs conceptuals i de les xarxes semàntiques. No eren estàtics sinó que pressuposaven una interpretació dinàmica: per a Llull, saber què volia dir un concepte equivalia a seguir-ne les associacions i explorar-ne les conseqüències. (La visió inherent d'ontologia dinàmica que això donava del món va ser entusiàsticament rebuda per innovadors filosòfics i científics com ara, al s. XV, l'influent Nicolau de Cusa.)

*La idea dels diagrames*

Els universalment cèlebres diagrames de Venn (de fet d'Euler) també tenen el seu origen històric últim en Llull (fig. 7). Va ser ell el primer a tenir la idea de representar gràficament els seus *conceptes* a base de cercles i a enllaçar-los superposant-los i intersecant-los –tot i que el que volia no era ensenyar si hi havia intersecció entre ells sinó mostrar que tenien una afinitat mútua

més o menys forta. (També va enllaçar els termes d'un sillogisme en forma de triangle, en allò que més tard, amb Buridan, es va dir *pons asinorum*.) Els dibuixos circulars de conceptes que feia Llull van esdevenir al s. XVI, a mans de J. L. Vives, una eina didàctica, i van ser millorats al s. XVII per Sturm o Leibniz –que en va fer tota una notació lògica (inèdita)– i finalment, és clar, per Euler (a la dècada del 1760). (Nosaltres ara en diem, impròpiament, diagrames de Venn.)

### *Els orígens*

Per als historiadors és interessant seguir la pista de com i d'on va treure Llull les seves idees peoneres. D'algunes ja en sabem alguna cosa. Per exemple, la de començar amb un conjunt finit de regles i desenvolupar tot un sistema té un antecedent remot en Euclides i en els grecs alexandrins, i un altre de més recent i innovador en el llibre d'«àlgebra» d'Al-Khwarizmi, llibre que Llull cita com a font, i que, traduït al llatí poc abans de l'època de Llull, va provocar un positiu daltabaix amb la seva idea inèdita de manipulacions rígides per regles i d'*algorismes* (concepte i paraula que precisament deriven del matemàtic àrab). D'altra banda, la idea de Llull d'un mètode omnicomprensiu que a partir de regles desenvolupés conceptes probablement era una extensió formal d'un component bàsic, ara totalment oblidat, de l'educació medieval: les complexes i elaborades tècniques per recordar i estructurar els conceptes en la memòria (humana) en una època sense impremta. De fet, el *mètode* lul·lià, un cop desenvolupat en aquest sentit per Petrus Ramus al s. XVI, va ser la inspiració de Bacon i Descartes. Pel que fa als dispositius mecànics (els discos rotatoris (fig. 1)), ara sabem que a la dècada dels anys 1260 a Algèria hi havia uns discos semblants «per respondre preguntes» (de fet, estris d'endevinació), i que Llull ben bé podia haver-ne tingut notícia poc abans de proposar, el 1274, els seus propis discos. Discos que, per cert, han tingut un aprofitament inesperat en criptografia, des del moment en què Alberti els va fer servir per codificar missatges (fig. 8), i encara avui els podem reconèixer en els rotors de la màquina desxifradora alemanya Enigma de la segona guerra mundial (fig. 9), llunyana i inesperada ressonància del mecanisme lul·lià.

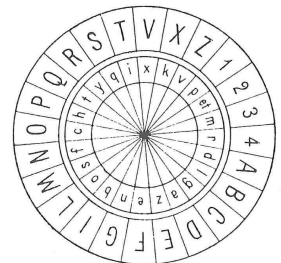
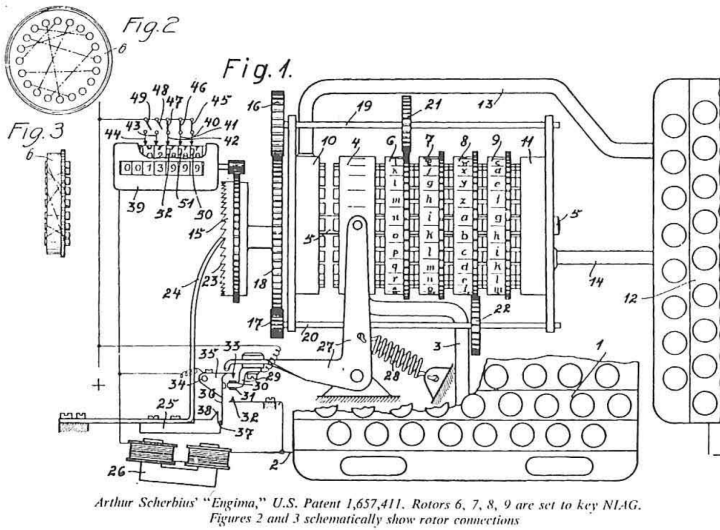


Figura 8: El «disc de xi-frar» d'en Leon Battista Alberti (s. XVI).



Arthur Scherbius' "Enigma," U.S. Patent 1,657,411. Rotors 6, 7, 8, 9 are set to key NIAG. Figures 2 and 3 schematically show rotor connections

Figura 9: L'últim avatar criptogràfic de la roda de Llull.

### Les conseqüències

Llull no és un simple precursor ni, com a anticipador, ha estat oblidat. La seva obra, que va haver de pagar el peatge inesperat de veure's incrementada amb tota mena de llibres apòcrifs que li eren falsament atribuïts, era ben conegut i apreciat per molts pensadors influents del Renaixement i per posteriors. Va exercir una gran influència en gent –que no sempre la va admetre explícitament– com ara Montaigne, Pascal, Descartes o Newton (que tenia Llull a la biblioteca, fet que l'igualava amb el seu arxienemic Leibniz). Giordano Bruno i Leibniz no sols en van rebre la influència sinó que no van dubtar a reconèixer haver-la rebut. Leibniz és el nostre enllaç més directe amb Llull. Tot buscant una notació universal i una manera universal –més o menys inspirada pels mètodes de la matemàtica– d'adquirir i desenvolupar el coneixement (la seva *mathesis universalis*), va llegir i absorbir Llull amb fruïció, el va adaptar críticament i va proposar un sistema objectiu i mecànic per fonamentar-hi sòlidament la lògica i el pensament racional. Hi va fracassar, però, després de deixar un reguitzell de notes inèdites (que inclouen una àlgebra del pensament i un formalisme a base

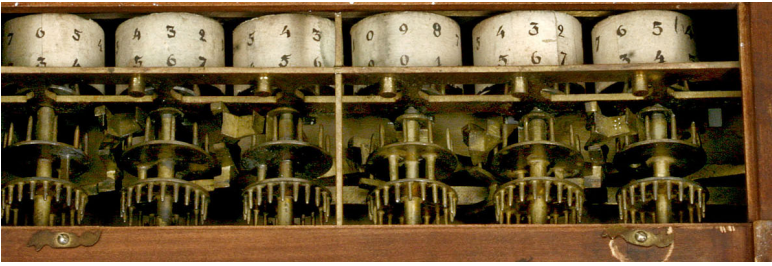


Figura 10: Els discos de Llull ara ja «mecànics» de veritat (aquí, la calculadora de Pascal).

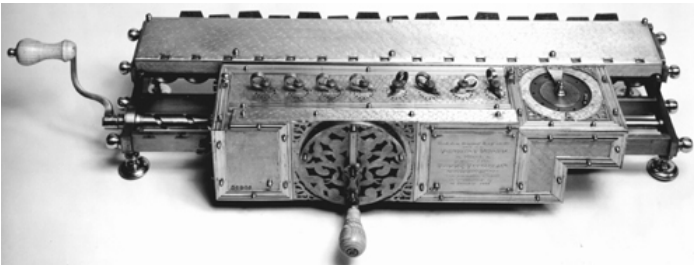


Figura 11: Els discos de Leibniz, encara decimals, podrien haver estat «binaris».

de grafs). Només uns 150 anys més tard el seu frustrat programa va poder ser desencallat per les salvatges intuïcions de Boole. Però altres idees de Leibniz van poder seguir endavant, sobretot el seu intent de descompondre i analitzar conceptes, que va tenir dues derivacions no previstes: (1) l'anàlisi de quantitats menudes (els *infinitèsims*, sobre el desenvolupament i prioritat dels quals la discussió amb Newton va acabar en una acrimònia lamentable) i (2) la construcció efectiva per Leibniz, a la dècada del 1670, d'una màquina de calcular real (la primera multiplicadora pràctica) (figures 10 i 11) que col·lateralment va motivar una inesperable reflexió per part de Leibniz sobre la idoneïtat del sistema binari de numeració per calcular). La reflexiva i elaborada *Dissertatio de arte combinatoria* de Leibniz, publicada el 1666, no sols és una lectura interessant i agradable per a qualsevol lògic o matemàtic d'avui. És la millor crítica i el millor homenatge que Llull ha rebut mai: reconeixent els seus mèrits i adaptant les seves idees a les necessitats modernes de la ciència, Leibniz ho va fer tot i més per incloure Llull dintre la nostra herència científica i, fent-ho, ens va fer un favor a tots.



# *Lògiques per a formalitzar estats mentals. Una brevíssima introducció*

**E**ls esdeveniments físics són objectius, no tenen significat, són controlables per les lleis de la física, i són inconscients, mentre que els mentals són subjectius, tenen significat, són intencionals (és a dir, són «sobre» alguna cosa), i són conscients. La ciència i la tecnologia normals tracten dels esdeveniments físics, mentre que la ciència cognitiva i la Intel·ligència Artificial (IA) tracten la conducta intel·ligent, habitualment lligada als esdeveniments mentals. Així, si volem màquines intel·ligents, les haurem de dotar d'una estructura mental? El fet de pensar els nostres programes (informàtics) en termes d'estats mentals ens fa les coses molt més fàcils, però traduir aquests estats en equivalents màquina –i programar un ordinador a partir d'aquí– no és pas feina planera, tot i que darrerament se n'han fet alguns intents amb una certa ambició. Aquí provarem de descriure'ls resumidament, però abans parlarem del concepte mateix d'**estat mental**.

BUTLLETÍ DE  
L'ACIA  
NÚMERO 10-11  
PRIMAVERA-ESTIU DE 1997

## *LA QÜESTIÓ DELS ESTATS MENTALS*

### *Esdeveniments físics i mentals: intencionalitat*

Els esdeveniments físics són tangibles, objectius i inconscients, no tenen significat, i són controlables per les lleis de la física. (Que no tinguin significat vol dir que no depenen de cap interpretació.) Els mentals, en canvi, són intangibles, subjectius i conscients, tenen significat, i són intencionals. «Intencional» no vol pas dir que hi hagi cap fi o intenció en la cosa sinó aquesta es «refereix a» –que



«va de» o que «és sobre»– algun objecte (potser inexistent) sense el qual la cosa perd tot significat. Així, les creences, desigs, pors o objectius són intencionals. La intencionalitat l'entén tothom força bé i es fa servir amb facilitat, però a l'hora d'explicar-la apareixen dificultats conceptuals, així com també a l'hora de convertir-la en eina de la IA (v. [4,9]).

### *L'anàlisi filosòfica de la intencionalitat*

La filosofia ha abordat el problema de la intencionalitat demanant-se si existeix o no cap correlat físic dels estats intencionals. Dintre de la tradició no dualista afavorida pels informàtics, el punt de vista materialista majoritari es resumeix en la següent frase de Searle: «El cervell és la causa de la ment» i, doncs, els estats mentals (intencionals) són estats cerebrals (físics). Hi ha una polèmica oberta entre els teòrics sobre la naturalesa d'aquesta relació. N'hi ha que pensen que no hi pot haver estats mentals en cap cosa que no sigui pròpiament un cervell, mentre que altres (els *funcionalistes*) opinen que una màquina també podria tenir estats mentals sempre que la construïssim de manera que fos funcionalment equivalent a un cervell. I, sense arribar a això, n'hi ha molts que creuen que el comportament d'una màquina és descriptible per mitjà d'explicacions intencionals, és a dir, a base d'atribuir-li creences, desigs i altres estats mentals propis dels humans. Alguns (sobretot Dennett [5]) ho consideren natural, mentre que d'altres (com ara McCarthy [11]) ho accepten com a metàfora útil, per estalviar-nos precisions innecessàries, o quan no disposem d'una explicació millor. En el supòsit que les màquines realment es comporten intencionalment, de seguida apareixen preguntes: podem expressar adequadament aquesta intencionalitat en termes lògics? podem trobar-li una semàntica adequada per «fonamentar» els estats mentals en alguna realitat física? (això darrer és el *grounding problem*). Tant McCarthy [11] com Dennett [5] hi dediquen reflexions, com fan Searle, Harnad i molts altres.

### *Estats mentals i consciència*

La relació entre els estats mentals i la consciència, evident en els humans, planteja problemes en la màquina: la consciència,

és independent dels estats mentals? n'és condició necessària? o suficient? És accessible a la ciència? La qüestió és prou pertinent, perquè els estats mentals i la consciència –o els seus correlats físics o funcionals– produeixen un efecte real, mesurable, en la conducta final del subjecte, efecte que ens permet predir-la o anticipar-la d'alguna manera.

## INTELLIGÈNCIA ARTIFICIAL I TEORIES D'AGENTS

### *El paradigma dels agents*

Una tendència relativament recent de la IA es basa en el concepte d'agent (vegeu per exemple el text de Russell i Norvig [17] del 1995). Un agent és una entitat lògicament formalitzable capaç de mostrar autonomia, iniciativa, comunicació, cooperació i adaptabilitat al seu entorn, i d'entendre'l. Aquest enfocament sembla més prometedor i flexible que els tradicionals, basats en coneixements i regles (els *knowledge-based systems*), i és el banc de proves escollit en què s'han formalitzat i assajat els diversos conceptes intencionals. (L'article de Wooldridge i Jennings [22] és una panoràmica de com estan les coses en la IA pel que fa als agents.)

### *Com capturar la intencionalitat?*

Els estats intencionals ja havien estat formalitzats en Lògica per gent com Hintikka [8] o Kripke [10] com a *Lògica de la creença i del coneixement*, una varietat de lògica modal que també s'ha fet servir per descriure els processos comunicatius entre agents (v. p. ex. Halpern i Moses [7]). D'altres extensions modals, com ara la lògica temporal i la dinàmica (v.p. ex. Goldblatt [6]), també ens interessen aquí perquè actualment s'estan utilitzant per copsar formalment els aspectes dinàmics dels estats intencionals.

### *Creences, temps, comunicació i intencions*

Combinar lògiques ja conegudes com les que tracten la creença, el temps o la comunicació és relativament fàcil, però els desigs i les «intencions» (volent dir objectius) no són tan senzills de formalitzar. La intenció ha acabat transformada en un operador que

segueix els criteris proposats per Bratman [1], i a partir d'aquí diferents autors han creat una sèrie de formalismes lògics semblants i aproximadament equivalents anomenats *lògiques BDI* (on *BDI* vol dir '*Belief, Desire & Intentions*'), dels quals només esmentem les formulacions de Shoham [18], de Cohen i Levesque [3] i de Rao i Georgeff [13,14,15]. Abans, però, de descriure aquestes lògiques esmentem una adaptació recent del càlcul situacional de McCarthy a l'autoconeixement (d'un robot) feta per McCarthy mateix [12].

### *Lògiques per a estats mentals*

Les quatre teories que triem i que descrivim de manera resumida són totes elles formalismes de representació lògica, la qual cosa no vol pas dir que excloquem altres formulacions lògiques com la de Singh [19] o la de Werner [21] o explicacions radicalment alternatives com ara els sistemes reactius de Brooks (o de Steels) [2,20] o els autòmats situats de Rosenchein i Kaelbling [16]. (Sobre les propostes de Brooks i Steels vegeu l'apèndix que precedeix les referències a la fi d'aquest article.)

Tota la informació important sobre aquests formalismes es pot trobar en dos reports tècnics [4,9] –publicats el 1993 i el 1995, respectivament, per l'AAAI (American Association for Artificial Intelligence) en l'ocasió de sengles congressos monogràfics– així com en treballs anteriors dels seus autors. Aquí només en descriurem els punts principals.

### *La formulació de càlcul situacional de McCarthy*

John McCarthy [12] proposa formalitzar el coneixement que pot tenir un robot d'ell mateix a base d'estendre el seu Càlcul Situacional a operadors proposicionals intencionals com ara *belief*, *know*, *learn* i *forget* així com a operadors d'acció tals com *decide* i *intend*. McCarthy distingeix les proposicions (regides pel metapredicat *holds*) de les accions (amb els metapredicats *occurs* i *result*) i hi afegeix uns axiomes que regulen el comportament del robot-agent (com per exemple *holds (know p, result (learn p, s))*, on *s* és la situació donada).

*Lògiques BDI: la formulació de Shoham*

En un ambiciós article sobre *agent-oriented programming* Yoav Shoham [18] proposa formalitzar conceptes intencionals com les creences, les obligacions, les eleccions (o decisions), compromisos i capacitats mitjançant operadors complementats amb una semàntica normal modal de Kripke, una lògica modal temporal i uns quants axiomes regulatoris i interdefinicions (com quan Shoham defineix decisió com l'obligació de l'agent amb si mateix: « $DEC_a^t\psi = OBL_{a,a}^t\psi$ »).

Alguns exemples (autoexplicatius d'expressions) :

- $holding(robot, cup)^t$  (t és l'instant)
- $B_a^t\phi$  (=l'agent *a* creu la proposició  $\phi$  en *t*).  
Exemple:  
 $ABLE_{robot}open(door)^t = CAN_{robot}open(door)^t$
- $OBL_{a,b}^t\phi \rightarrow B_a^t(\phi \wedge ABLE_a\phi)$  (Principi anomenat de benevolència)
- $OBL_{a,b}^t\phi \leftrightarrow B_a^tOBL_{a,b}^t\phi$  (Principi anomenat d'introspecció)

*Lògiques BDI: la formulació de Cohen i Levesque*

En una formulació paral·lela a la de Shoham, i –com ell– amb la pretensió d'aplicar la teoria clàssica dels actes lingüístics de Searle i altres a les accions intencionals, respectant escrupolosament els criteris de Bratman [1], Philip Cohen i Hector Levesque [3] han elaborat a partir del 1990 una lògica modal amb semàntica estàndard de Kripke, a la qual han afegit temps (lineal), basada en els operadors modals *belief*, *goal* i *intend*, i els temporals *happens* i *done*. Shoham i Cohen-Levesque són molt conscients de les (petites) divergències entre els seus sistemes, com quan observen que la *decision* de Shoham no inclou cap motivació o desig per part de l'agent, mentre que el *goal* de Cohen and Levesque sí.

*Lògiques BDI: la formulació de Rao i Georgeff*

A. Rao i M. Georgeff [13,14,15] construeixen un formalisme que ells anomenen *agent-oriented situated systems* i que és àmpliament

equivalent a les dues lògiques BDI esmentades. Hi caracteritzen els conceptes intencionals *belief*, *desire*, *intention* i *commitment* i hi afegeixen una semàntica de temps bifurcat i l'èmfasi en aspectes pràctics tals com especificació, disseny i verificació d'agents. Alguns exemples d'expressions:

- $\models INTEND\phi \rightarrow \neg GOAL\neg\phi$
- $\models INTEND\phi \rightarrow \neg BEL\neg\phi$
- $\models BEL\phi \rightarrow GOAL\phi$ .

### *Apèndix: sobre un parell de sistemes radicalment alternatius*

Segons Brooks [2], la AI tradicional sempre ha consistit –per voluntat i vocació (cf. la *hipòtesi dels símbols físics* de Newell)– en la **manipulació abstracta de símbols**, on els símbols mai no queden realment ancorats (*grounded*) en la realitat. El control dels sistemes resulta jeràrquic i els sistemes cada vegada més complexos (Brooks diu gràficament que «s'ofeguen en un mar d'incrementalisme») sense que no es pugui identificar cap fragment concret d'«intelligència» en cap mòdul en particular. Brooks proposa una nova IA «situada» on els mòduls o agents elementals, cap d'ells especialment intelligents o complexos, simplement «construeixen» una «intelligència» a base de coexistir i cooperar.

Semblantment, Steels [20] fa temps que insisteix en la idea que en els sistemes intelligents la representació és més «analògica» que simbòlica (discreta) o simplement categòrica (= prèviament categoritzada); Steels creu que la intelligència és més aviat el resultat d'operacions elementals de tipus difusió/reacció (per cert, inventades per Turing als anys 1950).

## *Bibliografia*

- [1] Bratman, M. E.: *Intentions, Plans, and Practical Reason*. Harvard University Press, 1987
- [2] Brooks, R. A. : *Intelligence without representation*, Proc. IJCAI, Sydney (1991)
- [3] Cohen, P. R. & Levesque, H. J.: *Intention is choice with commitment*, *Artificial Intelligence* 42 (1990), 213-261
- [4] Cox, M. & Freed, M., eds.: *Representing Mental States and Mechanisms*. AAAI Technical Report SS-95-05 (1995)
- [5] Dennett, D.: *The Intentional Stance*, MIT Press (1987)
- [6] Goldblatt, R.: *Logics of Time and Computation*. CSLI (1992)
- [7] Halpern, J. Y. & Moses, Y: *A guide to completeness and complexity for modal logics of knowledge and belief*, *Artificial Intelligence* 54 (1992), 319-379
- [8] Hintikka, J. : *Knowledge and Belief*, Cornell University Press (1962)
- [9] Horty, J., & Shoham, Y, eds.: *Reasoning about Mental States: Formal Theories and Applications*. AAAI Technical Report SS-93-05 (1993)
- [10] Kripke, S.: *Semantical analysis of modal logic*, *Zeitschr. f Math. Log. u. Grundl. der Math.* 9 (1963), 67-96
- [11] McCarthy, J.: *Ascribing mental qualities to machines*, in: Ringle, M., ed. *Philosophical Perspectives in Artificial Intelligence*, Harvester Press (1979)
- [12] McCarthy, J.: *Making robots conscious of their mental states*. in [4] (1995), 89-96
- [13] Rao A. S. & Georgeff, M. P.: *Asymmetry thesis and side-effect problems in linear time and branching time intention logics*, Proc. IJCAI, Sydney ( 1991), 498-504
- [14] Rao A. S. & Georgeff, M. P.: *A model-theoretic approach to the verification of situated reasoning systems*, Proc. IJCAI, Chambéry (1993), 318-324

- [15] Rao A. S. & Georgeff, M. P.: Verification of agent-oriented situated systems: a model-theoretic approach, in [9] (1993), 115-124
- [16] Rosenschein, S.], & Kaelbling, L. P. : The synthesis of digital machines with provable epistemic properties, Proc. TARK, Morgan Kaufmann ( 1986)
- [17] Russell, S., & Norvig, P.: Artificial Intelligence: A Modern Approach, Prentice Hall (1995)
- [18] Shoham, Y: Agent-oriented programming, Artificial Intelligence 60 (1993), 51-92
- [19] Singh, M. P.: A logic of situated know-how, Proc. AAI (1991), 343-348
- [20] Steels, L.: Steps toward common sense, Proc. ECAI, Munich (1988)
- [21] Werner, E.: What can agents do together: a semantics of co-operative ability, Proc. ECAI, Stockholm (1990)
- [22] Wooldridge, M, & Jennings, R. N.: Intelligent Agents: Theory and Practice (report submitted Oct. 1994 'to the Knowledge Engineering Review, revised jan. 1995)

# *Llimacs, eixams i càlcul col·lectiu*

## *(Prefaci a un resum)*

**E**N l'article anterior, en què parlava dels estats mentals i les seves lògiques i dels robots amb ment (però «descebrats») i amb sentiments (o no), ja es veia que per a molts IA-aïres programar un robot consisteix a explicar agradósament a l'interessat (el robot) el que ha de fer, el com i el perquè, talment com si l'hi ho expliquéssim a un nostre fill : «fes això, compte amb allò, si vols anar per aquí abans hauràs de preveure que no saps encara si hi ha algú al mig, i per tant ho hauràs de descobrir»,... ! És clar que això ens va bé als humans per entendre'ns, però parlar als robots en aquests termes sembla com parlar-hi amb el nostre gos (de fet, ho fem). I també és clar que les explicacions d'aquesta mena (les anomenades intencionals) poden arribar a ser –o semblar– científiques. Bé li ho semblaven a Aristòtil, que a la seva Física explicava la caiguda dels greus en funció de la seva «voluntat» (és a dir, les pedres cauen perquè volen caure). Trobar una explicació no teleològica, més objectiva i menys voluntarista, va costar Déu i ajuda, com tothom sap.

Els lògics no poden dissimular que sospiten que tot això de robots que «volen», «creuen» i «decideixen», per molt formal que ho facin, no passa de broma (McCarthy parlava de com un termòstat «volia» –i s'hi esforçava– mantenir constant la temperatura) ni té gaire sentit. Per això formalment parlen, amb un tecnicisme fet per despistar, de llenguatge intencional. Però, com no em canso de dir, hi ha altres maneres de veure les coses, i no tot han de ser màquines de Turing (com Turing mateix explicava). Parlem-ne.

Tenir una **funció** predefinida i fer-la **calcular** a una màquina és

BUTLLETÍ DE  
L'ACIA  
NÚMERO 10-11  
PRIMAVERA-ESTIU DE 1997



relativament trivial. El problema general és més complex: suposeu que tenim un procés físic (o biològic, o evolutiu) del qual no sabem res: l'engegueu (en un estat que per definició serà l'inicial), deixeu fer, i en un moment donat (=estat final) l'atureu. La relació entre les condicions d'entrada i de sortida (o entre l'estat del món en començar i en acabar) defineix clarament una funció (en el sentit matemàtic). Que no sapiguem res del procés ni de la funció que aquest procés defineix no hi té res a veure. La qüestió ara és: com calcular aquesta funció? (És a dir: com predir l'estat final de l'evolució del sistema quan s'ha començat en un estat inicial conegut?) La resposta trivial és: reengegueu el procés (que en això consisteix precisament el càlcul). Un «calculador», per molt que sembli una altra cosa, no és altra cosa que aquell dispositiu –que cal que sigui més o menys controlable, perquè si no ja podem plegar– gràcies al qual s'executa aquest procés. Ara bé: què passa quan les condicions de reexecució ja no hi són? I d'altra banda: es pot sempre descriure –i construir– el procés?

Per variar, i per fer pensar en alternatives lleugerament heterodoxes que ens espolsin la son i la hipnosi habituals, us proposo llegir el resum en anglès d'una tesi en curs feta a mitges entre el Departament de LSI de la UPC (que com se sap acull informàtics més o menys teòrics) i el Grup de recerca de sistemes complexos (majoritàriament físics) de la mateixa UPC. El doctorand, **Jordi Delgado**, és un informàtic que volia saber el fonament del seu ofici i que per fer-ho va seguir (sense saber-ho) el consell de von Neumann: per saber informàtica cal estudiar Física (perquè, com deia ell –i també Turing–, el càlcul és un subapartat del món físic). D'aquí li surten tot un seguit de conceptes que més avall veureu i que resulten, per a un informàtic qualsevol, de familiaritat nul·la. Posats a sortir, hi surten fins i tot eixams d'abelles i formigues, soles i acompanyades. I no hi surten –però hi podrien haver sortit perfectament– fins i tot llimacs, que amb els seus centenars escassos de neurones deuen ser els éssers més estúpids de la creació (la qual cosa dóna una bona raó per començar-hi, com fan una colla de neuròlegs dels quals algun dia que ens vagui m'agradaria de parlar-vos). I la idea que s'hi configura és que la intel·ligència sembla més aviat un afer d'elements simples que, cooperant, generen complexitat i un difícilíssim equilibri. Dit altrament, que la intel·

ligència és un resultat **emergent** que sorgeix d'un mar de possibles caos quasi-impossiblement evitats. I també hi resulta evident que per estudiar-ho ens calen eines potents i no pas simplificacions o mecanismes explicatius inicialment comprensibles.

Delgado s'ha pres la molèstia de creure's la Lògica i la Teoria de la calculabilitat que entre tots li havíem inculcat durant la carrera i l'ha complementat amb nocions de Física (ben bé a l'inrevés del que va fer Feynman). Fixeu-vos que accepta que, com passa amb la religió catòlica, d'explicació teòrica del càlcul, si n'hi cap de vertadera, només n'hi ha una: la clàssicament acceptada de Turing (que, per raons que ara no fan al cas, entre tots hem simplificat fins als extrems ridículs d'un catecisme). Per què començar per aquí, tot i que potser no se'n sigui creient? Doncs perquè l'herència de Turing, tot i catequètica, està formalment i fortament estructurada, i perquè té molts anys i consens. Però Delgado també es planteja, com tants d'altres, per què tot ha de ser tan simple i monocolor; sobretot quan els nostres conceptes de càlcul s'extrapolen inadmissiblement en forma d'intel·ligència, majoritàriament vista (pels IA-aires) com un tipus suprem d'ordinador més o menys programable.

Poso uns quants exemples, alguns dels quals Delgado va deixant caure. Primer: sempre s'ha dit que la intel·ligència era social (i sociable) i, en canvi, les nostres màquines són, per disseny, éssers totalment solitaris. (I per molt que ara les distribuïm en xarxes i així es facin companyia i es comuniquin, la cosa fa olor d'*ersatz*, de succedani de la sociabilitat autèntica.) També s'ha dit que la intel·ligència és situada, reactiva, adaptativa, interactiva, evolutiva i bàsicament regulatòria (d'un «cos»), i sense símbols (tot això ja ho intuïen von Neumann i els «cibernètics» dels 1950 i ara ens ho descriuen els neurocientífics tipus Damàsió als 1990); i, així i tot, fem màquines aïllades, planificadores, decisores, voluntaristes i mentalistes, sense cap «cos» al qual referir-se (allò que en anglès es diu estar *embodied*) i plenes de símbols (és a dir, basades massivament en la representació del món extern, com si sense això no hi pogués haver res d'imaginable). És només això –o ha de ser necessàriament això– la intel·ligència? L'únic que sabem, de moment, és que els grans fundadors de la informàtica –Turing, von Neumann o Wiener (i el primer Minsky!)– no ho creien pas, de cap manera: pensaven en termes d'agents elementals, equilibrats

complexos, i de la intel·ligència com a qualitat complexa **emergent** (conceptes que ara tot just desenterrarem), i van haver de ser el «cop d'estat» dels matemàtics antiwienerians de Dartmouth el 1956 que acabés d'arrel (per això el van fer) amb tot el que no fos reduïble a aquesta simple equació:

Intel·ligència = Fer programes

(Programes, això sí, que facin coses «intelligents»)

i aquest corollari:

$IA \subseteq \text{Matemàtica.}$

Va haver de morir von Neumann, Rosenblatt (de disgustos), Wiener i la resta perquè una nova generació recuperés els tresors colgats dels mestres. És el que fem ara: desenterrar. En fi, ja veurem què passa i com acaba. Però, amb l'escepticisme que em dóna el coneixement d'episodis antics de la història de la ciència i amb la meua vellesa incipient, m'ensumo que la cosa evolucionarà cap a una situació en què ens direm, incrèduls: ¿te'n recordes, quan a la facultat ens explicaven que qualsevol cosa calculable era polinòmicament reduïble a un mecanisme d'una simplicitat i falta de realisme insultants que anomenàvem *màquina de Turing*? (Per cert, aquesta mera hipòtesi de treball de Turing i Church del 1936 ha esdevingut la *tesi de Turing*, batejada així amb evident ignorància del que deien tots dos i gairebé diria que amb desconsideració pel bon nom d'aquest intel·ligentíssim, sensible i malaguanyat homosexual anglès.)

En fi, esperem. Veurem com acabarà aquesta aventura apassionant (apassionant per als qui tinguin capacitat de passió, és clar). I mentrestant llegiu això que segueix<sup>2</sup>. Constatareu que és tan diferent del que ens diem habitualment els uns als altres per consolar-nos que potser alguna d'aquestes idees acabin essent veritat i tot. Heus-ho aquí.

<sup>2</sup> L'autor es refereix a un resum de la tesi de Jordi Delgado «On collective computation» que sortia a continuació d'aquest article al Butlletí de l'ACIA número 10-11. (n. de l'ed.)

## *La intel·ligència artificial: ciència? tècnica? marc teòric?*

**S**I mai ens fem les dues primeres preguntes, és clar que la resposta anirà des de «és una ciència» fins a «forma part de la tecnologia», passant per tots els matisos de l'entremig. I, atenció, encara que puguin semblar preguntes retòriques, fer-se-les té una enorme importància pràctica, perquè de la resposta que hi donem depèn la selecció de temes que investiguem, com els enfoquem, com ens hi posem -a l'hora de treballar-hi- i com n'interpretem els resultats. I això val tant per als professionals de la cosa com per als estudiants que hi accedeixen. Per això val més confiar-nos a un expert acreditat i històric impulsor de la IA (Minsky, si no m'enganya la memòria) que fa uns anys ens explicava, partint de la pròpia experiència, que la matèria es podia dividir en dues varietats : la IA *matemàtica* i la IA *informàtica* (ell en deia *neat AI* i *scruffy AI*, respectivament, termes que potser podíem traduir per *IA polida* i *IA basta*). La primera va de **teoremes**, i consisteix a demostrar que hi ha una notació que permet *descriure* -o potser fins i tot *predir*, però això ja fóra demanar massa- allò que es té entre mans, mentre que la segona va de **programes**, i consisteix a fer veure que Déu n'hi do les coses que pot fer un programador «manetes» si s'ho proposa. Aquestes dues escoles de «pensament» (per dir-ne alguna cosa), immiscibles, es fan ben evidents en qualsevol congrés de IA-aïres i, de manera més barata i accessible, la barreja és constatable experimentalment només obrint un revista com ara Artificial Intelligence (la seu catedralícia de la IA) i fent-hi un cop d'ull a l'índex.

De fet, la IA té algunes estranyeses ben pròpies. Per exemple, els

BUTLLETÍ DE  
L'ACIA  
NÚMERO 13  
HIVERN DE 1998

*polits* i els *bastos* no se solen conèixer ni tractar; en efecte, quan hi ha un congrés de IA, ni es veuen. Uns aprofiten per arribar dimecres, just quan els altres acaben de marxar. Així eviten haver de discutir a mata-degolla els posicionaments, evidentment absurds, de l'altre. De fet podrien coincidir en un mateix congrés, sempre que fos en mons paral·lels. Els *polits* (els *neats* d'en Minsky), amb els seus teoremes, omplen pissarres senceres i discuteixen sobre la posició dels punts i les comes, en un exercici que qualsevol *bast* consideraria inútil, perquè, suposant que fos comprensible, sempre quedaria la fatídica pregunta sense contestar: «i tot això per a què serveix?» o aquesta altra, que té un grau d'iteració més avançat: «i fóra de les condicions d'aplicació tan simples i idealitzades que fixa el teorema –és a dir, en condicions reals– a quina situació que fos mínimament realista es podria aplicar tota aquesta cosa abstracta?». Preguntes que ningú no contestarà mai perquè no s'arriben ni a fer. Com les que, en direcció contrària, podria fer un *polit* a un *bast*, després de quedar, potser, impressionat pel programa que se li presenta: «I tot això, en què supera, i per què, els programes que s'havien fet abans?» o bé aquesta, més penetrant: «Molt bé, el programa fa tot això, però: què és el que es proposava fer, i fins a quin punt se n'ha sortit o no, i com comparar-ho amb els programes rivals aproximadament equivalents?».

Si la IA ha de ser una tecnologia com Déu mana sembla adequat que hagi de ser cosa d'inventors, que com se sap són gent que descobreixen alguna cosa que abans no havia descobert ningú, i de dissenyadors, que són els qui es proposen, seguint unes especificacions prèvies, fer un producte que compleixi certs requisits i, si pot ser, optimitzi factors com ara el cost o l'estètica. La pregunta és: quin és el producte? i quines eren les especificacions? Sí, és clar, el producte havia de ser un sistema «intelligent», però en què ha de consistir exactament aquesta intel·ligència? I com sabrem si ho hem aconseguit? Tenint en compte l'antiquíssima relació que hi ha entre intel·ligència i sorpresa, la tendència que tots tenim és d'avaluar l'objectiu pretès (la intel·ligència) per l'efecte subjectiu que ens produeix de «resultat inesperat». Això ho fem tots sempre, o hi tendim, però, francament, caure-hi no és gens científic, ni presentable. Doncs, com mesurar la qualitat del «producte»? Ningú no sembla tenir-ne idees. Ara bé: aparentment, per seriosi-

tat, caldria que l'autor del programa hagués fet, prèviament, una declaració d'objectius, un repàs dels programes competidors i les seves prestacions, i, tot seguit, es posés a construir-ne el seu, n'observés el funcionament en situacions objectivament equiparables i finalment fes un balanç de fins on ha arribat ell i on havien arribat els precedents. Un cop fet això, els seus col·legues i el món en general li agrairan poder fer una avaluació externa i relativament objectiva dels mèrits del sistema proposat, i de fins a quin punt és «intelligent» o permet avançar cap a la comprensió o reproducció de la intel·ligència. Això és arriscat, però honorat. Ho fan, però, els *bastos* IA-aires programadors (els *scruffies* minskians)? Doncs generalment no, i les seves ponències als congressos s'assemblen més a una desfilada de galls ensenyant llüides crestes a clients potencials o competidors impressionables que no pas a una trobada d'enginyers o dels tecnòlegs que passen per ser.

I què passa mentrestant amb els *polits*, aquests insubornables matemàtics d'ofici o de vocació que mostren una tendència irrefrenable i hipnòtica cap al teorema, cap a la fonamentació teòrica i matematitzadora de qualsevol cosa –encara que sigui amb l'eina conceptual del calçador– i cap a la interpretació «lògica» (en el mal sentit de la paraula, el de la discretització, del simplisme i dels models abusius)? Doncs a aquests els fa molt contents veure que una notació i uns teoremes semblen descriure una realitat complicada. I aquí retrobem l'aversion al risc que ja havíem descobert en els *bastos*: les diferents metodologies, notacions i teoremes no es comparen mai, perquè de fet són incomparables (per disseny o per tradició intel·lectual, per gust estètic o simplement per no haver de comparar). Com en el cas anterior, l'assistent a congressos o el lector de revistes es queda amb l'habitual pam de nas preguntant-se frustrat: on som ara, i en què hem avançat (si és que hem avançat)? (I una altra pregunta, feta amb la veu fluixa: aquesta fòbia a prendre riscos, no deu pas ser més aviat covardia o deshonestedat intel·lectual?)

Sembla que la IA, tot i les seves pretensions de científicitat i de tecnologia seriosa, no passi mai d'aquest punt d'espectacle malabar, de «vejam qui la fa més grossa» per a gaudi del personal sensible (i dels militars americans o assimilats que, al cap i a la fi, són els qui acaben pagant la factura). No sé si l'efecte que la IA fa als llecs és de seriositat o no, però, vista des de dins, l'empresa deixa molt a

desitjar. I tot això pel que fa als treballadors *de camp*, és a dir, els qui proven cada dia d'obtenir resultats nous. Doncs: què no haurem de dir dels grans gurus oraculars, els sintetitzadors de la matèria, els qui tracen les línies de la recerca del futur i elaboren els marcs teòrics on encabir cada cosa? Aquí la qüestió de la científicitat es torna, massa sovint, dubtosa. Desgraciadament, en la nostra IA ha calat una tradició que altres ciències amb més pedigrí o experiència han aconseguit treure's de sobre. És una variant de posar el carro davant dels bous, i consisteix en això: en lloc d'elaborar una **teoria** –explicativa dels fets observats i predictiva o suggeridora de fets nous no observats encara– i, en acabat, elaborar opcionalment amb finalitats didàctiques una imatge o «metàfora» autoexplicativa que ajudi a comprendre la potser difícil o poc intuïtiva teoria, el que es fa en IA sovint, segurament massa, és partir d'una **metàfora**, i tot seguit, basar-hi les diverses teories que se li ocorren a cadascú. D'exemples no en falten: el màxim i primigeni és el de la metàfora segons la qual la ment (o la intel·ligència humana, que és el que al capdavant ens importa) és com si fos un ordinador. La intenció és clara: com que sabem que és això, ja tenim una manera d'entendre o imaginar-nos com deu ser allò. O aquesta altra: la intel·ligència és un *sistema complex* (o *autoorganitzat*, o el que es vulgui) com aquests que estudien els físics o els biòlegs: sí? doncs ja en sabem alguna cosa. (Les preguntes que un científic s'hauria de fer en aquest punt són: hem arribat a algun lloc, fent això? no hauríem més aviat de considerar la comparació com una simple imatge motivadora o com a mera hipòtesi de treball a l'espera de ser confirmada o invalidada?) La fallàcia de tot plegat és no haver entès que una metàfora, per molt brillant i il·luminadora que pugui ser, no és una teoria, ni la seva comunicació és, de cap manera, el substitut de l'explicació dels fets. Donar voltes a una metàfora no té res a veure amb manipular i extreure conseqüències d'una teoria. L'«explicació» metafòrica es queda allà on era el referent de la comparació (l'ordinador, el sistema complex), i no arriba més enllà. L'«explicació» oferta no és pas el punt final sinó el de partida, no és la tesi sinó la hipòtesi, no permet «entendre» la realitat sinó fer plausible i engrescadora una possible visió d'on arrencar la recerca. Naturalment, una bona explicació, en el sentit científic, no s'ha pas de quedar en un símil del qual pouar conseqüències.

Ha de poder permetre arribar, si és el cas, a la conclusió contrària, encara que sigui autoinvalidadora i (creativament) destructiva. Això no és masoquisme, és merament l'honradesa que cal suposar-li al científic. Per exemple, l'equivalència inicial ment/ordinador com a hipòtesi de treball ha de poder permetre arribar, un cop desenvolupada correctament, a la conclusió contrària: que, fet i fet, la intel·ligència potser no s'assembla gens a un ordinador (o a un sistema complex, o a una xarxa neuronal, o a un algoritme genètic, o al que sigui) com havíem suposat inicialment.

Per a la gent normal (i cal suposar que els IA-aïres ho són, normals) el perill de les metàfores és el seu poder «hipnòtic», la facilitat que ens donen de pensar que, pel fet que «veiem» una cosa, ja la comprem, i de confondre els seus components amb «fets» (que, doncs, essent fets, ja no caldrà prendre's la molèstia de comprovar). En ciència cal trobar *explicacions*, però aquestes han de ser causals i han de complir una sèrie de requisits notablement forts. De cap manera no han de ser «explicacions» en el sentit casolà del terme (de l'explicació com a «història» que permet intuir de què es parla, allò de «ara t'explicaré com va anar la cosa»). Aquestes, en ciència, no s'admeten: si mai n'apareix una en un text s'hi afegeix immediatament que es tracta d'una teoria o de consideracions «merament especulatives» (i un article que en contingui més enllà d'una certa mesura és sistemàticament rebutjat, com va poder comprovar el lògic Post als anys quaranta, quan va voler explicar les idees que tenia sobre la calculabilitat). No cal dir que en IA aquestes «teories» no mereixen pas el denigrant adjectiu d'«especulatives»: passen per teories *tout court*, com Déu mana, ningú no se'n fa pas creus i, de fet, omplen la major part del contingut teòric dels llibres i de les revistes de l'ofici. I no és que la idea sigui dolenta; al contrari, la ciència es basa en hipòtesis si fa no fa especulatives que ofereixen una explicació merament plausible dels fets (és a dir, el científic ofereix als seus col·legues una narració hipotètica de l'estil: «És possible imaginar que el que hagués passat realment fos tal cosa, perquè si fos així –i fóra raonable pensar-ho– això donaria compte dels fets observats»). A partir d'aquí el científic comença a treballar amb la hipòtesi, hi dóna voltes, la manipula formalment i n'extreu noves conseqüències no previstes –si poden ser sorprenents millor



(i encara millor si són inversemblants)– que tot seguit s’afanya a comprovar experimentalment. Si el fet predit s’observa realment, la teoria proposada queda (provisionalment) validada. Si no, la teoria es va refinant, i pot passar que, tot refinant-la, quedi tan lluny de la proposta inicial que tot plegat acabi sent poc comprensible, o fins i tot antiintuitiu, moment en què (si els fets ho validen) caldrà trobar una explicació més o menys metafòrica que ho faci menys incomprendible o, si més no, més tolerable. És aquí que apareix la metàfora, no pas abans.

Potser algú pensarà que el cas que descriu de suplantació de teories per simples metàfores més o menys potents (algunes, cal reconèixer-ho, amb capacitat «hipnòtica» d’arrossegar) és infreqüent. Doncs no. La IA actual en va tan plena que em pregunto si tota ella no és una enorme metàfora. Penseu només que es basa, fundacionalment i per vocació sistemàtica i quotidiana, en la «tesi» de Church segons la qual tota cosa que pugui realitzar operacions formals del màxim nivell és equipotent a un formalisme simple i intuïtiu (però bastant antinatural) convencionalment dit *màquina de Turing*. Aquesta convenció operativa (que Turing va prendre, i sembla que va ser l’últim a fer-ho, com el que era: una hipòtesi de treball, provisional) es basava en una imatge intuïtiva proposada per Turing mateix: la de la persona que calcula (*calculista* o, en anglès, *computer*, mot que com és sabut només s’aplicava en aquella època a persones, mai a màquines). I és d’aquesta metàfora, esdevinguda hipòtesi, que deriva tota la IA actual. Si algú no hi està d’acord o no creu que la imatge que dona sigui prou fèrtil haurà de lluitar molt per fer triomfar el seu plantejament (i, més prosaicament, ho tindrà força difícil per publicar en revistes de l’ofici, majoritàriament centrades en el paradigma hegemònic). Els recercadors que s’han estimat més partir d’altres bases conceptuals han acabat adonant-se de com n’és de forta la convenció metodològica dominant, com n’és de difícil que s’entenguin les alternatives que no es basen en la metàfora bàsica i fins a quin punt és «religiosa» (perquè demana «fe» i trontolla amb el dubte) l’adhesió a la imatge fornida per «la metàfora» per antonomàsia.

La ciència no és, però, religió. I el científic no demana a la ciència ni fe ni seguretat, sinó dubte metòdic en tot, grans dosis d’observació pacient (basada en la ferma creença, aquí sí, que no

hi ha res com la realitat) i capacitat d'autonegació i rectificació sempre que convingui. Mirant-me el panorama intel·lectual de la IA, que encara no sé si és ciència o tecnologia o si té més d'una cosa que una altra (quina?), em sembla arribar a la conclusió que la IA no és encara ni l'un ni l'altre, i que potser té un pèl més de religió (potser laica i tecnològicament punta, però religió) que no pas de ciència entesa convencionalment. Si aquest és el cas, potser caldrà consolar-se pensant que així els IA-aïres som més a prop de Déu. Que per pretensió no quedi.

Barcelona, 12 de febrer de 1997



## Redacció de fi de curs

**C**OM que el final de curs i el començament del següent és un bon moment per ponderar –com diuen que fan els filòsofs cada dia– on som, d'on venim i cap on anem, m'he permès en aquest número de fer una mena de balanç provisional de la cosa aquesta de la IA. La idea general que en trec és que, si els *IA-aires* professionals –els *intelligents artificials*, que en dic jo (pitjor fóra dir-nos [AI =] *artificialment intelligents*, com he sentit dir a Amèrica)– no som capaços de madurar sols en el nostre desconsol teòric, el temps –que és el millor enginyer– bé que ens farà aprendre i madurar. I això és independent de les teories, modes i envolats mentals que ens acostumem a fer nosaltres sols per passar el temps i sentir-nos realitzats. Les consideracions que em faig són aquestes. I només són una mostra.

Primera. La nostra civilització occidental, des dels grecs, ens ha anat preparant intel·lectualment per tenir i acceptar com a cosa natural tenir esclaus que ens ho facin tot i, com que tenir-los amb aparença humana no és gaire agradable o moralment justificable –per allò de l'explotació i els remordiments morals–, hem maldat fins que finalment aquests esclaus ja puguin ser totalment manipulables i que no se'ns assemblin gaire. La IA es pot veure com un intent de fer una mena de *nosaltres* perfectes i explotables sense que tinguem problemes de consciència i no ens agafin mals de ventre. En altres paraules, la IA neix de l'imperatiu, per al qual la nostra cultura ens ha preparat sòlidament, de fer *zombis* obedients, insensibles i que no cotitzin a la seguretat social (ni, és clar, protestin de res).

Segona. Per fer aquests zombis cal que sapiguem exactament què volem deixar per a nosaltres, i quin mort carregar als nostres sòsies. A això ens hi hem anat acostumant d'una manera més

BUTLLETÍ DE  
L'ACIA  
NÚMERO 16  
TARDOR DE 1999

o menys natural amb el temps. Ara ja no resulta estrany veure sistemes que fan dòcilment el que els diem. Ni que ho facin milions de vegades sense equivocar-se, ni queixar-se'n ni avorrir-s'hi. Cosa que resultaria inimaginable que ho fes un humà. Així, hem fet que els nostres ordinadors calculin i no parin de calcular, des de reserves de bitllets fins al sou que guanyem un cop descomptat el de l'inspector d'hisenda. I calcular, això ho sap qualsevol infant, és una de les tortures més inútils i inhumanes que es poden imposar a cap ésser vivent. Raó per la qual hem dissenyat els ordinadors, aquesta mena de resignats autòmats que fabriquem, com una varietat d'*idiots savants*, un gènere d'estaquirots que són, essencialment, *calculadors nats*.

Tercera. Un cop sabem o hem anat sabent què volem que facin els nostres zombis, per dissenyar-los i programar-los sempre fem el mateix. De fet, sembla que no puguem fer res d'altre: primer analitzem l'activitat en qüestió no sols com si fos un problema matemàtic, o plantejat com els agrada als matemàtics, sinó com si, a més, el problema admetés la solució que aquestes qüestions solen admetre, especialment quan el problema s'ha pensat bé perquè resulti resoluble, i quan els resoladors són alumnes llestos i aplicats que troben la solució que el professor havia previst. (Cosa que, com és sabut, és corrent que passi en classes de Matemàtiques però gens a la vida real). I tant esperem que els problemes es resolguin d'aquesta manera tan neta i pulcra, que ens quedem com ofesos quan el problema se'ns resisteix o la nostra solució resulta ridícula. Esperem que l'ordinador faci allò que li hem dit, i de cap manera som tolerants amb l'error, tot i que, paradoxalment, estem disposats a no sorprendre'ns si el resultat ens dóna alguna sorpresa si fa no fa inesperada. Ara, que un ordinador quan actua faci com nosaltres, és a dir, que *s'equivoqui* i, precisament com a conseqüència d'això mateix, *aprengui* a fer les coses bé sense que ningú li digui el què ni el com, això no sols no se li ocorre a ningú sinó que no ho admètriem ni bojos. Deixar que un ordinador *s'equivoqui*, quin disbarat! I ,en canvi, ben mirat, potser els informàtics no hauríem de fer màquines (insensibles i) infal·libles, sinó que més aviat hauríem de fer com fem quan eduquem un professional: ensenyar-li de què va la cosa, donar-li coneixements i seguretat, i esperar que *s'equivoqui* poc (o gens?) i que *n'aprengui* com

més aviat millor fins que finalment, en estat estacionari, la suma d'èxits superi prou abasegadorament la de fracassos. Si això és el que esperem i demanem d'un bon professional, com no demanar-ho d'una màquina? O com considerar-la *intelligent* o una *bona professional* si no té cap possibilitat d'explorar i admetre riscos (i, doncs, d'equivocar-se)? Ho faríem amb un humà?

Quarta. Algunes coses sí que les hem anat aprenent. Coses que potser no volíem saber. Per exemple: (1) Que, per molt intuïtiu i gratificant que seria si ho fos, el cervell humà no s'assembla gens a un ordinador. No és de cap manera com ens l'imaginem, sinó que és una cosa múltiple que, miraculosament, es posa d'acord amb ella mateixa per fer-nos el caritatiu favor de fer-nos creure que *veiem* i *pensem* coses (quan realment el que hi ha és un desencadenament quasicàòtic de processos escassament coordinats). I una altra cosa que hem après és aquesta: (2) Que potser ja no ens podem estar més de parlar de –atenció, que no va de broma!– *sentiments* (=estat de semi- o pre- alerta desencadenada per coses vistes, succs gàstrics o hormones diverses) i de *consciència* (=estat de visualització *unitària* de la situació, i de l'agent mateix, com si fos una sola cosa, estil *mapa* i *centre de comandament*). *Sentiments* i *consciència*, ningú no ho pot pas negar, *evolutivament* han estat una bona cosa per a nosaltres –i per això els tenim–, i potser –per què no?– fóra bo de ficar-les també en els nostres *zombis*. Però, això sí, amb un mecanisme de control: si no hi anessin així proveïts, vés a saber quina una ens farien. Perquè la civilització occidental, a més de fer-nos pensar en zombis (com a possibles, i àdhuc necessaris), i dissenyar-los i construir-los, també ens ha ensenyat que ens hem de *controlar* (controlar-nos i, especialment, controlar els qui no som «nosaltres») –i per tant, controlar, encara més, els petits Frankensteins que van resultant del nostre esforç d'informàtics. I també sabem, perquè no hem parat de dir-nos-ho, que els sentiments són justament una font de descontrol especialment temuda, a evitar de totes totes (cap religió ni filosofia no ha deixat passar l'oportunitat d'insistir-hi, des de l'infern de Zaratustra als exercicis de St. Ignasi). Així que, ¿com carai haurem de resoldre aquest dilema de posar-hi sentiments, si s'escau, a les màquines, perquè són *funcionalment útils*, i alhora no posar-n'hi, perquè són perillosos i amenacen amb l'horrible possibilitat de descontrol? (Se'ls hauria

de fer, a les màquines, classes de moral? o exercicis de ioga, per allò de l'autocontrol?)

Recapitulant: Si ens consideréssim enginyers als quals es fa l'encàrrec de fer màquines intel·ligents, el nostre desconcert no podria ser més gran, i el resultat fóra un disseny estrofolari i impossible, com d'*invent del TBO* (o d'*en Heath Robinson*, que és l'equivalent anglès del barceloní Franz de Copenhaguen). Hauríem de fer màquines *sensibles* però sense passar-se, que ho entenguessin tot però que fossin obedients com xaiets, que ens anticipessin en tot però ens deixessin sempre l'última paraula. Això –que és una mena d'equivalent asexual de la *poupée gonflable*– ja es veu que no lliga de cap manera, per moltes truites que somiem.

De moment, i ja que sembla que no ens acabem de decidir a posar sentiments als nostres zombis, podríem, això sí, dotar-los d'una certa *sensibilitat* per entendre els *nostres* sentiments (vull dir, els dels éssers humans, entre els-qual-espero-em-compto). En això ja s'hi treballa, veig: acabo de llegir un informe de Byron Reeves i Clifford Nass de Stanford sobre les increïbles reaccions emotives que produeixen certes respostes dels ordinadors sobre gent madura i preparada (i informada) com són els seus propis programadors: les reaccions, diguem-ho de seguida, no són gaire bones; si el que els diuen, o com els ho diuen, no els resulta prou convincent (als programadors), doncs, –com us ho diria?– el gel (i la fe) es trenca. I això no pot ser gaire bo. Conseqüència: caldrà dissenyar els ordinadors del futur no sols perquè calculin bé el resultat, sinó que ens hauran de fer creure (*persuadir* és el mot exacte) que el resultat és el que ha de ser i que fem bé de creure'ns-el i pujar-hi de peus. Per això no resulta estrany saber que a l'MIT s'escarrassen a hores d'ara a fer *diagnosticadors on-line d'estats d'ànim* –no rigueu, si us plau– de les persones que es miren els monitors (*frustration response detectors*, en diuen ells). La idea és estudiar-los la cara en tot moment per detectar-hi l'imperceptible gest de desconfiança, astorament o desconcert, que, pel que sembla, són universals i parametritzables (a partir de tres o quatre trets de la cara que la màquina registra contínuament), i corregir ràpidament la naturalesa de la frase amb què comunicar-se amb l'alterat interlocutor humà. Tot es fa, és clar, amb bona intenció: es tracta d'evitar malentesos o indisposicions. La qual cosa, d'altra banda, no és

pas nova del tot. ¿Qui no ha vist un recepcionista que, després de donar asèpticament, i sovint incomprensiblement, la informació que se li demana observa sorprès, veient la cara que té al davant, que el seu preguntador no ha entès ben bé res del que se li ha dit? Com tothom sap, davant d'una tal situació el recepcionista pot fer dues coses: tirar pel dret i qualificar l'interlocutor de ruc i passar al següent, o bé apiadar-se'n i tornar-li a explicar tot de cap i de nou, si és possible amb alguna petita variació orientadora i adornat amb el difícil –per segons qui– rictus muscular correntment anomenat somriure. No cal dir que m'alegro molt que amb les màquines provem de fer interlocutors amables. Això sol ja mereix l'esforç. I consti que lamento que l'ordinador pugui deixar recepcionistes i comunicadors a l'atur. Ho lamento, com sempre que passa que una persona més es queda sense feina –que, de passar, passa (que per això és la Informàtica, no ens enganyem!). Però en aquest cas, francament, potser no ho lamento tant com en d'altres. Posats a fer, no és estrany que la gent s'estimi més ser escanejat per la cara per una màquina amable que rebre miquels d'un humà ungit d'una estranya superioritat.

Que la intel·ligència té components desagradables però necessaris (calcular o avorrir-se infinitament, per exemple) ho hem anat descobrint els *IA-aires* amb gran pena i esforç. Que certs components de l'amabilitat són mecànics i fàcilment reproduïbles ho hem après ara de l'MIT i, abans, del comportament de publicitaris i polítics (almenys). Que la intel·ligència té dues components essencials que podríem dir *sentimental* i *heurística* (és a dir, que no existiria si no fos pels *sentiments* i per les *equivocacions*) ho sabem gràcies a Turing o Damásio. ¿No fóra hora potser, doncs, d'anar pensant a combinar aquests ingredients del còctel? De moment, tot esperant els primers ordinadors que s'equivoquin (una mica) i dels qui pateixin (també una mica), potser ja estem en situació de començar a concebre els primers ordinadors una mica *amables* o, encara millor, *simpàtics* (segur que els americans en diran *friendly computers*, però això ja està gastat i no és ben bé el mateix). Potser aquestes màquines no seran gaire o prou intel·ligents, i no mostraran la suprema simpatia d'equivocar-se una mica, de tant en tant, per no fer-nos sentir tan estúpids, o potser la seva amabilitat ens resultarà postissa o poc creïble. Però, sincerament, vistes les



coses que ens han arribat a fer creure gent aparentment simpàtica que fa com que ens estima molt, vol el millor per nosaltres i no s'equivoca mai, potser val més confiar-se als futurs ordinadors simpàtics. Seran més fiables –i la llarga, francament, més barats.

Barcelona, 20 d'agost del 1998

## *Sobre el principi (i la fi) de la Intel·ligència Artificial*

**Q**UAN parlo del *principi* de la Intel·ligència Artificial («IA» a partir d'ara) ho faig en els dos sentits de la paraula principi: en què es basa i com va començar. En canvi, l'expressió «*la fi* de la IA», és, com és obvi, autoexplicativa –si bé potser de les dues és, paradoxalment, la que demana més explicacions. La primera expressió ve de la pregunta que sovint em faig com a professional: «per què coi la intel·ligència artificial s'assembla tan poc a la natural (la nostra, vaja)?» I encara més: «per què hem triat com a suport per a la primera unes màquines que s'assemblen tan poc als nostres cervells?» Pel que fa a la segona expressió, la pregunta és: «s'acabarà, això de la IA?» I, «si s'acaba, com s'acabarà o què en quedarà o a què donarà lloc?»

El que segueix és el resultat de provar de respondre'm aquestes dues preguntes. Però de moment ja avanço les meves conclusions –temptatives, no cal dir-ho. Pel que fa a la primera: la IA no vol ni ha volgut mai fer realment màquines *intelligents* –ni molt menys mirar d'entendre què deu ser la *intelligència*–, per molt que els seus cultivadors ho hagin pretès amb estrany autoconvenciment i pesant insistència; a la IA sempre s'ha tractat, *només*, de *dissenyar* sistemes i màquines (cosa d'altra banda gens sorprenent: és el que fan cada dia els enginyers); *només* que *aquestes* màquines són especials, pel fet que fan o poden fer coses que *fins ara* s'havien qualificat d'«intelligents» –i com a tals ens havíem reservat els humans en exclusiva. Pel que fa a la segona: a mesura que les màquines intelligents ens aniran resultant més familiars i inte-

BUTLLETÍ DE  
L'ACIA  
NÚMERO 17  
HIVERN DE 1999

grades en el nostre entorn quotidià, s'hauran acabat tot d'una no sols els nostres dubtes filosòfics (bé, potser no del tot, perquè els filòsofs potser també tenen el dret de viure) sinó també la IA com a *camp específic* de la informàtica, i els seus mètodes i objectius deixaran de ser tan transcendents i discutits com ho són ara: si més no, ho seran tant com ho són els de la termodinàmica o la mecànica de fluids. És en aquest sentit que anticipo *la fi* de la IA com a disciplina científica i enginyerívola més o menys compacta, problemàtica i doctrinària i la seva integració parcial en el camp de les especulacions socials i filosòfiques (i fins i tot, em temo, la seva banalització més o menys televisiva). El romanent (és a dir, la IA com a teixit d'activitats professionals vagament afins) deixarà de ser unitari i es dissoldrà en la resta de les enginyeries, tradicionalment caracteritzades per una molt lliure facilitat d'anar canviant el camp específic d'actuació i els propis mètodes i tendències segons el moment i la circumstància.

Siguem francs: la IA no s'ha basat mai en cap d'aquestes preguntes: (1) «què és la intel·ligència?», i (2) «què és i com és la intel·ligència humana (i concretament el cervell)?» No és pas que alguns no s'ho hagin demanat. La primera qüestió, concretament, els filòsofs i els periodistes no paren de plantejar-se-la. La segona va ser, de fet, el punt de partida de la sensible i molt meritòria feina de Turing i Von Neumann. Ara, cal reconèixer-ho, avui cap de les dues qüestions no sols no interessin realment ningú, tret dels biòlegs, sinó que probablement més val que sigui i continuï així.

De fet, mai ningú del ram s'ha preguntat, abans de començar la feina: «què és la *intel·ligència*?» A tot estirar, això ho van fer –i encara parcialment– Galton al segle passat i Binet i Terman a l'actual, i el resultat són els *tests d'intel·ligència* que coneixem. Res més. La intel·ligència, com la vida, és relativament fàcil de *reconèixer*, però difícil o impossible de *definir*. Sabem que hi ha característiques que provenen –o s'hi correlacionen– del fet que alguna cosa té vida o intel·ligència: certes respostes i actuacions, reproductibilitat, comunicació i, sobretot, *adaptabilitat*. Ara, de la mateixa manera que la imprecisió definitòria de *vida* no ens impedeix combatre els virus i els prions –tot i que no sabem ben bé què són, si *vius* o no–, res no ens ha impedit mai fer màquines

que fessin coses *intelligents* (= normalment atribuïbles a alguna intel·ligència), independentment que les màquines que ho feien fossin pròpiament *intelligents* o no. I si no ens ha preocupat gaire fins ara, ens ha de preocupar en el futur?

Tampoc ningú no ha fet cap mena de cas a les possibles preguntes sobre què és i com és la intel·ligència humana, i concretament el cervell. No és que alguns no s'ho hagin demanat –Turing, McCulloch i Von Neumann de fet van començar fent-se aquesta pregunta– o que uns altres no pretenguin que això els importa poc o molt (els xarxaires neuronals d'avui). És que a ningú no li interessa realment el tema, i el que n'han sabut tampoc no els ha servit per a res. N'hi ha prou a repassar els models i productes que es proposen com a sistemes intel·ligents per veure que no s'assemblen gaire a cap cervell –si no és via *metàfora*– ni s'hi inspiren gens. De fet els dissenyadors de la cosa, en la seva gran majoria, no saben pas com funciona, el cervell, ni els ha calgut gens saber-ho. (Chomsky mateix deia fa pocs anys, a Girona, que difícilment podia pensar res més llunyà als seus interessos personals que saber què carai passa al seu cervell.) Fa poc uns biòlegs entusiastes de la UAB em van demanar que els parlés de la influència que havien tingut els models *biològics* del cervell sobre els esquemes de la IA. Pobres, quina desil·lusió; els vaig haver de dir que no n'hi havia hagut cap, d'influència, ni gota. La IA era cosa exclusiva de matemàtics i informàtics, que havien –havíem– pres com a guia del que era la intel·ligència merament *allò que els matemàtics* fan quan pensen o resolen un problema. Res més. Els resultats d'aquesta hipòtesi simplista no deixaven de ser espectaculars, però en cap cas es basaven en cap mena d'anàlisi de *què és* la intel·ligència –o de com és el cervell. I, tot i que darrerament estem assistint a una allau de noves i interessantíssimes revelacions sobre com funciona el nostre sorprenent i riquíssim òrgan cerebral, és dubtós que això ens desviï del camí traçat de ja fa tant de temps. De la mateixa manera que el coneixement cada vegada més precís del vol dels ocells o de la nostra pròpia locomoció no és gaire probable que afecti gens el projectes de nous avions o de nous cotxes.

El desconeixement, fins i tot la crassa ignorància, de què és la intel·ligència –la natural, vull dir, que fins ara era l'única *real*– no ens ha pas impedit fer sistemes *intelligents*. Igualment, hem

pogut construir avions, cotxes i pales mecàniques sense conèixer com volaven les aus, com caminàvem nosaltres o com movíem els braços. No sols no ho ha impedit, sinó que, un cop hem dissenyat, fabricat i emprat aquests ginys, ningú no s'ha sentit menystingut ni desplaçat. Ni sentim que ens hàgim convertit en ocells pel fet de volar (com a màxim, l'únic animal que recordem quan volem és la mare del pilot) ni ens considerem vexats pel fet que una màquina pugui anar més ràpida que nosaltres als mateixos llocs o que ja no ens necessitin per remoure terra o asfaltar un camí. Per què ens hauríem de sentir molestos si algú fa reserves de tren millor que l'estaquirots que ens posen davant del mostrador? Hi ha cap persona que es pugui molestar si un *agent* electrònic li recorda el que ha de fer o li aconsella el que ha de dir durant el dia?

Sincerament, l'únic que pot arribar a preocupar-nos dels nostres sistemes de IA –i d'això ja se n'ocupen els diaris i els filòsofs, no us hi amoïneu– és un conjunt de *dos petits detalls*. Primer, per alguna mena de complex d'inferioritat ens preocupa –i molt– que una màquina pugui posar en evidència l'escassa *intelligència* de què podem fer gala, amb esforç, la majoria de nosaltres. I segona: si els sistemes realment intel·ligents que farem hauran d'incorporar, com sembla, dues característiques fins ara exclusivament humanes com ara els *sentiments* i la *consciència*, on quedarem nosaltres? Quina cosa seran aquestes *màquines* que haurem creat? I com ens haurem de relacionar amb *coses* que tindran allò que fins ara era –deixant a part la *intelligència*– específicament *humà*?

Jo, francament, no en tinc la resposta, és clar, però la meua experiència o la meua intuïció em diu que d'una banda acabarem tenint-hi una relació semblant a la que tenim amb les màquines que ja tenim (per exemple, el nostre cotxe), és a dir, alhora *possessiva*, *utilitària* i *funcional* (sabem quan i com fer-les servir, ho fem, i prou). I, d'altra banda, pel que fa als nous aspectes (*sentiments* més *consciència*) segurament farem com ja fem amb les *màquines* vives i sensibles amb les quals convivim des de fa força temps: els *animals domèstics* (que també tenen uns quants *sentiments* i també, com a mínim, la *consciència* de ser ells mateixos). Senzillament, a cada casa hi haurà, a més del gos, un ordinador (o *robot*, si té mobilitat) que podrà parlar amb nosaltres i al qual podrem parlar. Haurem de reservar les nostres manyagues per a l'un (el viu) i les nostres

confidències per a l'altre (l'artificial), en un tot que, a la llarga serà harmònic i tothom acceptarà com a natural. (Ara: el que ja no es tan segur és que a la mateixa casa hi hagi criatures. I és més que probable que la casa només tingui un habitant humà.)

Vista la tendència actual, sembla com si tots haguéssim d'acabar vivint sols en paradisos virtuals solitaris, connectats, això sí, amb tota mena d'umbilics electrònics: imatges virtuals, veus telefòniques, ordinadors amics, animals domèstics (aquests ens obligaran almenys a sortir al carrer a passejar) i, potser, algun company/companya humà, si pot ser *adoptat* (preferentment, sembla, d'un color i país tan llunyà com es pugui) i en tot moment reemplaçable –tipus kleenex– no fos cas que el contacte *personal* –difícil i imprevisible, i per al qual estarem mancats de manual d'instruccions– irrités excessivament la nostra hipersensible minso de pell.

Això que dic no és ciència-ficció. Només cal mirar al voltant. Francament, la IA no crec que ens porti gran cosa més del que ja tenim o volem. Només desitjo que us vagi bé, que ens vagi bé a tots, en aquesta nova era. I si algú aconsegueix conèixer un ésser humà, viure-hi i fins i tot reproduir-se, que els vagi bé també als fills que en surtin. Almenys, que tinguin la sort que els pares no els confonguin amb un *gadget* més ni els treguin només a fer pipí al parc. I, sobretot, que hi conservem tots la salut (la *mental* és clar, perquè l'altra avui –si no vius al tercer món– ja és una cosa trivial).

Barcelona, 18 de maig de 1999



## *Agents: actors amb paper, personatges a la deriva, o éssers realment «intelligents»?*

**N**o cal ser cec per constatar-ho. Els anys 90, la moda és parlar d'*agents*. Va començar com una bona idea a l'hora de formalitzar sistemes d'IA complexos, amb unitats intercomunicades (per exemple, xarxes d'ordinadors) i de seguida tothom va considerar raonable el enfocament. No sols això, sinó que va passar allò que sovint passa amb els ous de Colom: tothom va dir «com no se'ns havia ocorregut abans una cosa tan òbvia!?!». La raó és clara: fins ara sempre havíem considerat allò que programàvem com si fos una entitat (ordinador, robot, el que sigui) que vivia en un món sense congèneres, un món sobre el qual sabia el que s'ha de saber i en el qual feia el que s'havia de fer, sense més por a cap interferència externa que no fos el temut error de paritat, de sintaxi o qualsevol altra fútil inconveniència, la responsabilitat de la qual sempre calia atribuir al programador, que n'era, com a suprem creador de la cosa, el culpable nat en cas que anessin mal dades. Ara, en canvi, admetíem que hi pogués haver altres programadors, ordinadors o robots que també intentessin fer coses actuant sobre el mateix univers que nosaltres havíem arribat a considerar propi. Què fer, doncs, quan dues entitats d'aquestes volien fer el mateix o actuar sobre un mateix objecte, o simplement s'interferien?

Per a una situació com aquesta ja teníem una eina lògica. Era la lògica de les creences, que permetia formalitzar no sols les nostres accions sinó també situar-les en un món que semblava ser –que

BUTLLETÍ DE  
L'ACIA  
NÚMERO 20  
HIVER-PRIMAVERA DE  
2000-2001



els sistema «creia» que era— d'una determinada manera. Era fàcil passar d'aquí a estendre la idea a altres objectes animats, fent suposicions sobre el que aquests «creien» i tenien intenció de fer actuar en aquella mateixa realitat que compartien amb nosaltres. Això semblava una gran idea, perquè per primera vegada havíem gosat admetre que no estàvem sols al món, que hi havia altres entitats dotades de voluntat (o, si més no, que actuaven de fet en el nostre terreny) i que reconèixer això ens permetria d'evitar interferències indesitjades (les que ens apartaven del nostre objectiu) o fins i tot aprofitar-les cooperativament (conjuminar les nostres accions amb les dels aliens per arribar a objectius que potser nosaltres sols no hauríem pogut atènyer). Que bé, oi? Doncs dit i fet. De sobte tot es va convertir en *agents* (com, paral·lelament, en informàtica tot s'anava convertint en *objectes*). Tant, que aviat vam tenir el típic postdoc americà-jove-brillant que va decidir canviar l'antiquada metodologia dels manuals d'IA a base de reduir-ho tot a una «ciència dels agents», si fa no fa, amb resultats de vegades il·luminadors però sovint agafats pels pèls, trivials o fins i tot empobridors. Parlo per exemple del manual clarament ambiciós i deliberadament trencador (*trend-setting*) titulat *AI: a modern approach*, concebut, pensat i escrit per Russell (el jove-brillant-promesa) i Norvig (el veterà-programador-de-LISP-amb-ganes-d'innovar-opotser-cansat-de-sempre-fer-el-mateix). Aquest llibre, publicat a mitjans 90, potser sí que era «modern» en aquell moment i potser sí que volia ser una síntesi de tot l'anterior amb les noves ulleres dels agents (la il·lustració de la portada així ho semblava indicar: hi sortien totes les glòries passades de la IA com si esperessin ser recatalogades per la gran i prometedora nova visió). Ara, de fet el que resultava de llegir-se'l amb calma era clarament que s'hi havia fet un intent volenterós d'unificar coses encara potser inunificables, amb una metodologia resultant que tenia més d'embastat i prematur que de consistent o consolidat. I és que potser hi ha més gana que menjar, i encara no ens ha arribat l'hora de les síntesis grandioses i il·luminadores.

Bé, sigui com sigui, la idea, potser encara no generalitzable, era passablement bona i ha anat tirant endavant. Els experts en xarxes han ideat mètodes més o menys ad hoc perquè les seves unitats es comuniquin i, si cal, s'interfereixin (en el bon sentit) a

cor-que-vols, que és del que es tracta; mentres els lògics, tronats descendents d'eximis i periclitats filòsofs que són, hi han aportat el seu impertorbable mètode de fer com si ho sabessin tot de tothom i de tot, ampliant, com ja és costum, el seu repertori de termes i signes amb un seguit d'operadors lògics que indicaven, ja no sols les proposicions rellevants a la realitat estudiada o descrita, sinó també els «coneixements», les «creences», els «objectius», els «desitjos» i les «intencions» (i fins i tot la «moral» o regles del joc permeses) del «seu» agent i de tots els altres, de manera que així tot es pogués coordinar satisfactòriament i tothom, o almenys ells, poguessin atènyer l'objectiu malgrat –o gràcies a– les interferències dels altres agents. (De tot això se'n diu *lògiques BDI* i ja en vaig fer un petit resum en un número passat d'aquest mateix butlletí.)<sup>3</sup>

Aquest plantejament, que respon al típic vici autista dels lògics, s'assembla bastant al rellotger cec de què parla Dawkins a propòsit de l'evolució biològica. Però és que ara sabem –o, més ben dit, saben els biòlegs– que l'evolució, a més de tenir i seguir una suposada lògica interna, ha estat també el resultat de camins abandonats, de subevolucions sense cap mena de lògica (ni interna ni cap altra) i, sobretot, de fenòmens *externs* que no hi tenien res a veure, des de la caiguda de meteorits que «passaven per allà» i provocaven extincions i efectes hivernacle a manta fins a oscil·lacions polars que tot d'una capgiren el clima i precipiten glaciacions i el retorn a les caveres.

I és que si volem parlar realment d'agents i d'interaccions ja no podem passar ni un moment més sense capbussar-nos d'una vegada en la dinàmica dels sistemes complexos i en l'anàlisi de què coi passa quan passen tantes coses *allhora* i *imprevisiblement* que qualsevol programador hi perdria l'esma. Però això vol dir que, no sols hem d'abandonar la vella i insostenible impertorbabilitat del matemàtic o programador que ho ha de preveure més o menys tot, sinó que també hem d'oblidar tota la *Matemàtica* que tradicionalment s'ha considerat que havien de saber els informàtics, que era bàsicament algebraica i discreta, i aprendre'n una de (per a nosaltres, totalment) nova. I potser, a més, haurem d'aprendre per primera vegada *Física* de debò, però no la normal sinó la que ara ocupa tothom, que és la dels fenòmens no lineals. Com que no sé si se m'entén, passo tot seguit a mirar d'explicar-ho breument.

<sup>3</sup> L'autor es refereix a l'article de la plana 95. (n. de l'ed.)

Subdivideixo el tema en dues parts diferenciades, començant per la física no lineal i acabant per una hipotèticament necessària nova matemàtica.

Primer. Suposem que el sistema que resulta d'interactuar entre si multitud d'agents és no sols complex sinó asíncron, continu i obert a l'exterior (aquest adjectiu és crucial i marca tota la diferència; *obert* equival a dir, si fa no fa: *disposat a tenir sorpreses*). El que en resulta és un sistema la dinàmica del qual és no sols complexa sinó difícilment estudiable si no és pels ara ja acreditats mètodes de la complexitat física, de la *dinàmica de sistemes complexos* (popularment coneguda com a «teoria del caos determinista» o «dels atractors estranys»). Es tracta de veure com el sistema obert evoluciona, de percebre si va en alguna direcció concreta, de si és estable o quasiestable o té algun atractor (encara que sigui fractal), de com és de sensible a petites variacions en les condicions inicials, de com reacciona davant de fenòmens aleatòriament imprevisibles (que hem d'acceptar que n'hi haurà, si volem ser realistes), de si d'alguna manera el sistema o alguns dels seus components s'autoorganitzen, i de si autoorganitzant-se donen lloc a algun fenomen *emergent*, amb propietats noves, a investigar, que probablement seran interessantíssimes i sorprenents. No cal recordar el comentari que feia Daniel Hillis a propòsit del comportament de la seva *Connection Machine*: deia que quedava allucinat de veure que l'activitat de la màquina de vegades es concentrava en un lloc, de vegades es dispersava de manera fractal o aparentment aleatòria, de sobte oscil·lava entre diferents pols o es tornava a concentrar en diversos punts, etc. (i això que la seva màquina era merament paral·lela, programada de manera determinista i funcionant com a sistema tancat, sense sorpreses, interferències ni «imprevistos» generats des de l'exterior). La dinàmica del càlcul, com es veu, pot arribar a ser molt complexa, i els mètodes tradicionals de la informàtica molt pobres per abordar amb èxit situacions d'aquest estil. La meua impressió és que el futur de l'anàlisi dels sistemes informàtics passen per la Física, i, més concretament, pels sistemes dinàmics, la mecànica estadística, la termodinàmica dels sistemes semiordenats, els fenòmens d'escala i de transició de fases, les situacions crítiques, el càlcul a la vora del caos, l'autoorganització de sistemes complexos (fets de components possiblement simples),

etc., etc. Si no n'esteu convençuts, consulteu la pàgina de web de l'Institut de Santa Fe on s'explica el que fan. Es dediquen a estudiar la dinàmica de evolució, dels ecosistemes i de mil coses més, entre les quals figuren de manera prominent coses com ara (i copio literalment) el càlcul *adaptatiu*, la simulació d'*agents adaptatius*, el tema «*càlcul, dinàmica i inferència*», els *fenòmens d'escala*, la *dinàmica de xarxes* o els *autòmats cel·lulars evolutius*, tot per mètodes de la física dels fenòmens no lineals i amb la idea bàsica que l'objecte estudiat és un –diguem-ne– *agent* (possiblement simple) que està dotat d'estratègies pròpies i que es mou *adaptativament* en un entorn *complex* (compost, per exemple, d'altres agents) i *obert* (en principi *imprevisible*). Us en quedariéu parats, de les coses que van trobant de mica en mica. El celebèrrim nòbel Feynman ho va plantejar en públic a l'any 1981 (càlcul = fenomen complex, informàtica = dinàmica de sistemes) i l'igualmente nòbel Gell-Man (l'inventor dels quarks) ens ho explicava fa poc a Barcelona. La qüestió és: els informàtics, ho saben, tot això? tenen idea del que els altres fan? (perquè val la pena que vigilin; potser quan se n'hauran adonat ja s'hauran quedat sense el seu estimat i exclusivista objecte d'estudi). Em ve al cap, a propòsit d'això, un estudi recent –fet per físics, i no precisament per informàtics– sobre el problema canònic 2SAT, que és sublineal, i el 3SAT, que és NP-complet, en què s'investigava el comportament del problema kSAT per a valors de paràmetre k entre 213; doncs bé: hi apareixien de manera natural i esperable típiques transicions de fase i algunes fluctuacions crítiques. No va ser el mateix Von Neumann que va dir l'any 1948, i li ho van publicar el 1955, que l'estudi dels ordinadors va lligat al del *soroll* i dels *errors* i que per això estudiar Informàtica vol dir haver d'estudiar Termodinàmica?

I segon. Si no volem anar per la ruta de la Física –encara menys de la no discreta i de la no lineal– i ens volem quedar en el terreny acollidor i familiar de l'àlgebra, de la recursió (inducció a partir d'ordinals) i de la localització de punts fixos, aleshores us recomano que us llegiu algun article recent (1997-99) de Peter Wegner (el més introductori, publicat a *Communications of the ACM*, o el més complet titulat *Coinductive models of finite computing agents*). És possible que la teoria encara no estigui totalment afinada, però ja s'hi poden entreveure algunes coses: a canvi de tenir un mo-

del tradicional de l'agent de càlcul («tradicional» = sistema finit tipus màquina de Turing), (1) la *interacció* sobretot, (2) l'*aleatorietat* (imprevisibilitat, diríem) dels inputs, que actuen com a autèntic «soroll», fan, en primer lloc, que el sistema es comporti d'una manera que fàcilment –això sembla bastant clar– superi en potencia el nostre sostre clàssic, la màquina de Turing, i, en segon lloc, que hàgim de capgirar totalment els nostres sistemes conceptuals d'anàlisi i verificació de programes. A partir d'ara, si aquesta línia encetada per Wegner s'imposa, no haurem de pensar els nostres programes en termes d'*inducció*, recursivament a partir d'un cas-base, sinó que haurem d'anar practicant la *coinducció* basant-nos en la suposició que no hi ha cap cas-base i d'una manera que l'autor bateja il·lustrativament com a procés de «desconstrucció» fins a trobar els punts fixos *màxims*, no pas el *mínim*. I no sols això. El que té Wegner d'interessant és que no sols ens obligarà a una disciplina gairebé inversa («desconstructiva» en comptes de recursiva/constructiva) a la que ens resulta familiar i «natural» (ho era, però?) sinó que ens obligarà a plantejar-nos si volem partir peres amb els matemàtics i anar pel nostre propi camí (o obligar-los a ells a seguir-nos), perquè per fer el que Wegner suggereix hem d'oblidar la Teoria de Conjunts dels matemàtics, que conté entre els seus axiomes el fins ara indiscutit de la *fonamentació* (*grounding*, en anglès) i entre els seus mètodes la inducció sobre un ordinal –opcions que no deixen de resultar naturals per a un matemàtic educat en Pontologia intuïtiva dels (seus) conjunts– i passar-nos a la nostra pròpia «Teoria (informàtica) de conjunts» (no vull dir que sigui informàtica, sinó tan matemàtica com l'altra però feta i pensada *ad usum informaticum*), en què el primer axioma dit passa a ser el de *no-fonamentació* (=«no hi ha cap conjunt de base», o bé «totes les cadenes de conjunts per inclusió son infinites») i el mètode demostratiu passa a ser la coinducció (en el si de «coalgebres»). Per cert, Wegner veu en l'aparentment innocu axioma de fonamentació Parrel de la inducció com a únic mètode possible de càlcul (fins ara) i com a explicació de per què carai cap màquina ni sistema formal podia superar mai el *límit de Turing*: la inducció ho feia tot automàticament comparable i *equipotent*, i l'única manera de superar aquesta estranya barrera (el *Turing tarpit* hi ha qui en diu) que el mateix Turing trobava absurda, provisional i poc natural

és justament deixant de banda definitivament la inducció com a mètode implícit en tot càlcul. Més concretament, prohibint-lo i fent exactament a l'inrevés.

Que potser és massa, tot plegat, per a nosaltres? Doncs prepareu-vos. El preu de no fer passos endavant, per dolorosos que puguin semblar, és estancament. I8 d'estancament, en tenim cabassos plens des de l'època que Turing i Von Neumann ens van deixar orfes. Ara, si gosem tirar pel dret el premi pot ser: primera, tenir (per fi!) un bon *model* del que passa al món real dels sistemes (d'*agents*) realment *interactius oberts*; i, segona, poder avançar en la programació-predicció realista del que se suposa que han de fer les nostres màquines, que han arribat a ser tan complexes que ja no les entenem. Potser no és gaire demanar que, simplement, ens posem a la seva alçada. Elles ens ho agrairan, i sobretot els nostres descendents. I, de passada, potser així els informàtics sabrem per fi què fan les màquines que fem veure que programem. Ara bé...

Potser tot serà bo per a tothom, però potser no tot serà necessàriament bo per a nosaltres, «intelligent artificials» altres informàtics. Veureu: és que, de sobte, l'efecte d'una notícia recent m'ha fet pensar que potser tot no han de ser flors i violes i que la nostra capacitat d'adaptació també té límits. I és que no voldria deixar passar l'ocasió de fer-me ressò d'un tema de moda les darreres setmanes de desembre (del 1999). No, no té res a veure amb l'anterior, però també representa un canvi de magnitud majúscul, que fa feredat només de pensar que ens pugui arribar de cop. Parlo de la descoberta del fet que els *nanotubs* (cilindres buits de carboni pur de diàmetres de l'ordre del *nm*) puguin superar clarament i destacadament –tant en mida com en propietats conductores i semiconductoras i en la possibilitat de fer xips planars– tot el que els nostres xips de *silici* ja definitivament no poden prometre. No us sembla pas que és una profunda ironia que els ordinadors del futur ja no siguin de *silici* sinó, com un vulgar exemple de nosaltres mateixos, de carboni? I no vull ni pensar com ens haurem de reciclar els informàtics, sobretot els que siguin dissenyadors de xips, per adaptar-nos sense traumes a la idea que tot el que sabem de circuits ja no val; repetirem així potser la desgraciada història dels experts dels 1950 en díodes i tríodes termoiònics quan en dos anys tot va quedar substituït per transistors: van haver de llençar

tots els seus manuals al foc.

Doncs, i ara tornant al tema que m'ocupava més amunt, i posant-me pessimista (de vegades també va bé mirar-s'ho pel cantó difícil): haurem potser de llençar al foc (o consignar en unes golfes) els llibres que explicaven que la màquina més omnipotent era la de *Turing* (tot i que per a Turing era una simple hipòtesi de partida) i que el sacrosant i potentíssim mecanisme d'*inducció/recursió*, que Hilbert el 1926 va fer magistralment assequible als simples mortals, no sigui el que hem de fer servir els informàtics per mirar de controlar els nostres incontrolables i complexos productes? I, horror dels horrors, haurem de fer màsters d'Anàlisi matemàtica i tornar a estudiar Física per entendre el que fem? No desitjo cap desgràcia per a ningú, però em temo que, si algú no atura el món, ens hi veurem abocats sense remei. Sort que, si a alguns ens agafa per resistir-nos-hi, potser passarà que ja ens hauran prejubilat per totalment superats, incompetents o per reducció de plantilla, o ens concediran la baixa per depressió, com solen fer amb els professors d'ESO. Això sí, en aquest cas tindrem tant de temps com voldrem per meditar ponderadament sobre la fragilitat humana tot jugant a la petanca amb altres informàtics també orfes. Així que ja ho sabeu, germans: o *carboni*, coinducció i dinàmica complexa, o petanca al parc.

Però no cal que ens ho mirem d'aquest color, i amb una mica de sensatesa no cal pas que reservem plaça al frenopàtic ni que passi res d'apocalíptic. Ja sé que els nostres plans d'estudis d'Informàtica i d'IA s'han mantingut persistentment i sòlidament ancorats i construïts sobre el sauló o la sorra, de consistència dubtosa, dels *paradigmes fundacionals* provinents dels anys 40 a 60, que els retocs més o menys modernitzadors i cosmètics que han sofert han estat més aviat per adaptar els hackers vocacionals que ens arriben a les necessitats de la trista, imitativa, seguidista i poc creativa i arriscada empresa que tenim, i que no hi ha res més difícil que canviar la mentalitat de persones que han après coses discutibles amb la seguretat solemne del *dogma* en què han acabat convertits els paradigmes fundacionals que deia (elaborats per genis intel·ligents, sensats i creadors que potser no ens mereixem d'haver tingut) i amb el convenciment difícilment mutable de tenir-les per veritats indubtables, indiscutibles i, pitjor, òbvies (quan no ho eren,

gens ni mica sinó tot el contrari, per als originadors dels conceptes en qüestió ni ho son tampoc per als no-informàtics, no-creients o simplement agnòstics). Tot això ja ho sé, i prou que ho sento. Però també sé que els joves, tot i malejats i deformats, i alguns fòssils que encara conserven la capacitat adaptativa que diuen que ens caracteritza als humans, se'n salvaran i podran pujar al tren que passa. Va ser Darwin qui va dir que «Les espècies que sobreviuen no són pas les més fortes o les més intel·ligents sinó les que s'adapten més bé al canvi». Doncs ara que en tenim un, de canvi (no pas petit), és l'ocasió de demostrar que, contra el que es pugui suposar, potser sí que som intel·ligents i capaços d'adaptar-nos i, per tant, de sobreviure.

Per cert, ja fa temps que hi ha gent (psicòlegs, biòlegs ara matemàtics i físics) que veuen com a característica bàsica i definidora de l'indefinible concepte d'«intel·ligència» el d'*adaptabilitat*. Podem aprofitar-ho per treure'n un corollari immediat: en comptes de preocupar-nos per fer *agents* «intel·ligents» el que cal és fer que els agents que construïm, que no cal ni tan sols que siguin complexos, i molt menys omniscients com un lògic o ormniprevisors com un programador, i que poden ser tan intel·ligents o sòmnes com una formiga, formin part d'un tot (un sistema dinàmic i/o evolutiu) que, aquest sí, sigui adaptable als canvis, previsibles o no, que hi pugui haver i que amb tota seguretat hi haurà. Parafraçant la dita anglesa que diu que l'única prova d'un púding és que a la gent que se'l menja li agradi, l'única prova de debò que el sistema que en resulta és o no realment adaptable és que sobrevisqui, i en les millors condicions possibles, als canvis, com més dràstics i inimaginables millor. I si ens agafa per substituir l'adjectiu *adaptable* per l'entelequial i difús «intel·ligent», doncs haurem fet el primer sistema realment intel·ligent. I, a més, estarà efectivament fet d'agents, que s'entrecreuran i interactuaran en un entorn *obert* (=sorpresiu, no-anticipable) i necessàriament no-definit, responent a esdeveniments inesperats i, possiblement, arribant a crear alguna cosa nova o alguna qualitat *emergent* com a imprevista conseqüència de la seva pròpia actuació o *autoorganització* espontània.

El millor del cas és que ja disposem d'eines analítiques que la Física i la Matemàtica ens han preparat per poder saber i seguir –potser fins i tot predir, encara que sigui aproximadament–



l'evolució i resultat del sistema. Dedicar-s'hi serà sens dubte una activitat més creativa i innovadora que la que ens hem limitat a fer en IA, que era merament fer clons programables de nosaltres mateixos (sense ni tan sols saber com nosaltres mateixos estàvem construïts), entenent per «nosaltres mateixos» una replica del pensador/raonador explícit i rigorós (i esbiaixat cap a la matemàtica) que som la majoria dels humans en alguns rars (de fet raríssims) moments de la nostra vida. Aquesta és una característica única nostra (només els humans la tenim), però en tot cas no forma part més que d'una petitíssima part de la nostra activitat vital i cerebral; n'hi ha prou de veure el que diuen els neuròlegs, repetidament: menys d'un 10% de l'activitat cerebral es dedica al que en diem «pensar»; la resta és l'activitat funcional pròpiament definitòria del cervell que, com a òrgan bàsicament regulador que és, es dedica, és clar, a «regular el cos» adaptant-lo a l'entorn de cada moment i situació i elaborant un complicat equilibri d'homeòstasis, emergències i funcionament crític al caire del caos que ell mateix s'autogenera necessàriament (tant, que disminueix més i més com més s'atansa el seu funcionament a l'electroencefalograma pla).

Abundant en el tema dels canvis que s'acosten, una altra cosa que podríem fer per aprofitar-los és canviar la famosa i tronada etiqueta dels 1950 «Intelligència Artificial» per una altra, com ara Sistemes adaptatius (o d'agents, si tant us agrada la paraula) o similar. Tindria l'avantatge que ens faria veïns dels físics i matemàtics (amb els seus «sistemes dinàmics») i dels biòlegs (amb els seus *sistemes evolutius*), tots dos grups amb una llarga tradició i experiència que ens falta als informàtics, que encara no hem passat de tenir-ho tot, incloent el més calent, a l'aigüera, i que potser ja fóra que ens comencéssim a espavilar. Així descobriríem algunes coses interessants que ja sap tothom i, veient-les a la nostra manera, potser contribuiríem a l'esforç col·lectiu i convergent de diferents grups de col·legues que ja fa temps que s'interessen de manera eficient per la *intelligència*, el nostre objecte de culte. De passada, podríem desterrar la a-científica, imprecisa, perillosa i equívoca paraula «intelligència» (com a mínim, deixariem de semblar-ne els propietaris o els summes sacerdots) i, tot renunciant a la patent, potser a partir d'ara ja no enganyariem ningú i, millor, no ens enganyariem ni nosaltres mateixos. I, a sobre, aprendríem (dels altres)

coses interessantíssimes i fins ara inimaginables. Us imagineu res millor com a raó per dinamitar (sensatament, no cal dir-ho) un pla d'estudis encarcarat? o com a candent tema de conversa per quan juguem a botxes al pati del frenopàtic? Si n'arribarem a veure, de coses...!



## Calcular i somriure o «El que la intel·ligència no és»

**L**A Informàtica –i la *Intel·ligència Artificial* (IA), que des del 1956 n'és una de les filles preferides– ha considerat tradicionalment que la *intel·ligència* consistia bàsicament en l'habilitat (sobrehumana) de fer càlculs molt complicats en molt poc temps. I d'una altra banda (completament diferent, i perdoneu-me la transició brutal entre un tema i un altre), sempre s'ha vist, i és veritat, que la capacitat de *somriure* és típicament humana.

D'aquesta manera, no és gens estrany que els ordinadors, que clarament són capaços de fer la primera cosa (*calcular* a cor-que-vols) són intel·ligents, proposició tan evident que ningú ha sentit mai la necessitat de demostrar. Doncs ara, com es veu en la convocatòria feta per la Caixa sobre «Emoció i coneixement» l'octubre del 1999, sembla que ens veiem impellits a afegir a aquesta intel·ligència, tan escandalosament mecànica, alguna cosa que la «humanitzi» i, com sap qualsevol dependent de comerç o alt executiu, què hi pot haver millor que començar aquesta operació a base d'aprendre a *somriure*, fer-ho de manera franca i freqüent, i practicar perquè quedi tan natural com sigui possible? Aquesta és una porta a la «humanització» dels ordinadors que, segur, farà cordial (*friendly*, vaja) el nostre contacte amb les fredes màquines, avui que en depenem tant i que necessitem la seva confiança per a tot.

Ara bé: segur que aquests dos components (capacitat de *calcular* i de *somriure*), per molt exclusivament humans que siguin, denoten alguna intel·ligència? Comencem pel primer: pel que es desprèn

BUTLLETÍ DE  
L'ACIA  
NÚMERO 21  
ESTIU DE 2000

d'una sèrie d'estudis recents –l'últim que m'ha arribat és dels *Proceedings* de la Royal Society de maig del 1999– la capacitat de «càlcul prodigiós» en humans que caracteritza els *idiots savants*, documentats des de fa tant de temps, està *anticorrelacionada* amb la resta de capacitats intel·lectuals humanes (les que correntment ens farien descriure una persona com a «intelligent», en qualsevol dels múltiples sentits del terme). Això no és cap sorpresa: cap dels *idiots savants* descrits pels documents (avui el més conegut del qual és potser el personatge de Dustin Hoffman al film *Rain Man*) han estat mai considerats com a persones pròpiament intel·ligents. I aquí no estarà de més recordar que el que hem estat fent fins ara els informàtics és justament això: *idiots savants* mecànics. De la qual cosa es desprèn –vaja, jo la'n desprec– que és més que dubtós que les màquines que hem fet siguin catalogables com a «intelligents», almenys no pas més que ens veiem amb cor de fer-ho amb els nostres sorprenents i inesperats «megacalculadors» humans (humans, sí, però històricament i popularment tractats com a cretins o subhumans).

I ara continuem pel segon. Somriure és un fenomen, l'explicació i el mecanisme del qual són encara poc clars, que caracteritza *únicament* l'especie humana, i per això considerem que fer somriures «humanitza» les nostres relacions. (La supressió d'aquest recurs és àmpliament explotat pels cineastes que volen retratar un futur deshumanitzat o simplement fer-hi sortir zombis.) Ara bé, sens perjudici del que acabo de dir, multitud d'estudis corroboren allò que molts hem constatat: somriure és una capacitat que no denota necessàriament intel·ligència. Deixant de banda els somriures postissos o calculats, tan desgraciadament freqüents, tothom sap que és molt i molt freqüent poder constatar que persones amb gravíssimes deficiències mentals, des de les afectades per la síndrome de Down fins a les víctimes de greus paràlisis cerebrals, somriuen contínuament. Com en la situació anterior, i sense de cap manera voler comparar els casos, pocs qualificarien els pacients sotmesos a cap d'aquestes afliccions d'«intelligents» en el sentit normal, socialment corrent del terme. (N'hi ha excepcions, però, com en tot.)

La conclusió que jo trec de tot això, provisionalment, és que el que hem fet, el que ha fet la IA, per molt que la intenció original fos

de construir màquines intel·ligents, no ha passat mai de desenvolupar dues característiques típicament humanes fins al límit (bé, de fet només ha desenvolupat la primera, perquè això de fer somriure els ordinadors encara està per veure). D'aquí a la «intelligència», sigui el que sigui aquesta cosa, hi va un tros. L'afer em recorda no sols la capacitat dels humans –ben coneguda dels enginyers– de sobrevalorar les seves pròpies capacitats davant de situacions desconegudes sinó també la metàfora proposada per un conegut científic sobre el miratge que suposa «avançar» científicament en un tema: deia aquest que, al ximpanzé que s'enfila en un arbre atret per la lluna, li pot semblar que amb aquesta ascensió hi arriba una mica, o s'hi atansa, ignorant que per arribar a la lluna no s'hi val a pujar a arbres sinó, entre altres coses, a calcular equacions diferencials, estudiar l'electrònica de les comunicacions i la química dels propellents dels coets, saber organitzar un col·lectiu altament qualificat d'especialistes i una mica de sort.

Amb l'ascensió a l'arbre del càlcul massiu i de la *friendliness* segur que estem més a prop de la lluna (de la intelligència, vull dir). Però no trobeu (i és una sospita raonable) que per abordar d'una vegada aquest tema no hauríem de fer una mena de NASA, o bé tenir la paciència d'esperar algun segle més?

En fi, en la meua ignorància i desorientació supines em proposo contribuir al desconcert general –críticament i enriquidorament, això sí– a base d'incloure, si els amables editors d'aquest Butlletí m'ho permeten, dos articles meus en què parlo de *xarxes neuronals* i d'*ordinadors emotius*, temes que com es pot veure no tenen aparentment res a veure. Així i tot, espero que el lector pacient que se'ls empassi acabi dient, amb l'ai al cor, «Houston, tenim problemes», sospitant que no tot és tan senzill i lineal com pronostiquen els llibres de text sobre la IA, i que de l'estudi i comprensió de la intelligència tot just estem a les beceroles. I que consti que no cito cap dels flipants articles que enguany mateix generen els neuròlegs sobre aquesta estranyíssima capsa, de contorns alhora boteruts i arrodonits, que ens proporciona migranyes i ens serveix per culminar el cos, aguantar el nas i separar les orelles (com sentenciava un vell proverbi sicilià), altrament dit *cap* o *testa* (o *olla* o *coco*, segons terminologies alternatives). El nostre cap és, per cert, un adminicle que, a part la funció estètica que sol acomplir

(com a targeta de presentació, síntesi de la personalitat, «reflex de l'ànima» i aquestes coses), té com a nucli explicatiu una «pasta humida i tèbia» –en paraules de Turing– que n'ocupa el litre-i-mig central i que Déu-n'hi-do les insòlites capacitats que mostra i la de sorpreses que ens està donant, ara que per primera vegada hem gosat de remoure aquesta inversemblant «pasta pensant» cerebral i penetrar-hi.

Algun lector que no em conegui se sorprendrà potser que, en un article suposat científic, parli de cops d'estat: concretament, un de molt sonat que va tenir lloc el 1956 perpetrat per uns científics contra uns altres, en la més ortodoxa tradició dels cops d'estat a què ens tenien acostumat fins llavors els polítics i militars. Si l'esmento no és sols per un deure de rigor històric sinó perquè si no, no es pot entendre gran cosa del que va passar. També parlo de temes aparentment repugnants –per a un cultivador de les ciències «dures»– com ara de (a) la identificació dels humans amb certs objectes (al paràgraf que he titulat «Del significat de les màquines», que més aviat pertany potser a la psicologia), de (b) l'adaptació social a un fenomen que es presenta com a nou i desconcertant (a «Les màquines demanen adaptació», que fóra cosa de sociòlegs), o de (c) la crítica a la moda (informàtica) dels *agents* (tema clarament informàtic, o directament frívol), subjectes que clarament cap matemàtic «ortodox» consideraria altra cosa que mera falta de rigor, vaguetats dignes de commiseració. (De fet, part dels problemes que comento van ser originats per aquests matemàtics que, en el seu zel de rigor, van oblidar que, com deia von Neumann, no es pot fer cap anàlisi rigorosa si les bases són dubtoses.)

Per al qui tingui curiositat de saber per què he triat dos temes tan desaparats per parlar del mateix i intentar fer entrar el clau per la cabota us confessaré que els dos em van ser imposats com a amigable motiu de conversa en sengles programes de Catalunya Ràdio, emesos l'estiu del 99 cada dia havent dinat (per facilitar la migdiada? per adormir les pedres? perquè ningú mai més no s'interessés per la Ciència?... és broma: el plantejament era intel·ligent i els temes força interessants i ben triats). Els programes prenen la forma de tertúlies multidisciplinàries conduïdes per l'Adolf Tobeña, de la UAB, metge d'ampli espectre que s'interessa

ahora per la intel·ligència, els nous coneixements i el que hi tenim a dir els tecnòlegs. Potser sense saber-ho estàvem, més de mig segle més tard, reproduint l'esquema de les converses que reunien a la riba esquerra del riu Charles gent de Harvard (com el metge Cannon) i de l'MIT (com el matemàtic Wiener), tots dos units i apassionats per la ciència i la democràcia (contra l'obscurantisme i el totalitarisme, respectivament) i moguts sentimentalment per la indignació que els va causar el bombardeig indiscriminat de civils a Granollers i Barcelona, en una de les primeres cafrades massives del segon terç del segle XX? Cafrades que no sé ben bé si demostren que l'espècie humana és intel·ligent o simplement una continuació no gaire millorada dels mamífers carnissers. Bípeda, això sí, i amb un cap verticalitzat.

El cap és, per cert, la capsa aquella de la qual deia més amunt que, a part servir per penjar-hi el barret, tenia la pretensió de ser suport d'un «enigma envoltat de misteri» (en frase cèlebre) anomenat «intel·ligència», una de les característiques de la qual és la de pretendre que *ella mateixa* es pot comprendre i reproduir, com si fos la cosa més natural del món que una entitat es pogués *autodescriure* o (en una altra frase coneguda) que els ratolins de laboratori haguessin de tenir la capacitat de comprendre la relativitat.





## Ordinadors emotius

BUTLLETÍ DE  
L'ACIA  
NÚMERO 21  
ESTIU DE 2000

**A**QUEST text segueix, per raons no lògiques sinó anecdòtiques que aquí es faria feixuc d'explicar, un article meu que es diu **Xarxes neuronals, cervell i ordinadors**, que trobareu en un altre lloc d'aquest Butlletí.

En aquest, com en l'altre, hi trobareu iteradament exposades certes cabòries que em caracteritzen com a informàtic madur (vull dir, vell) que les ha vistes de tots els colors. Com en l'altre, en l'estil del qual està escrit (tot i que el tema no pot ser aparentment més divers), vull combatre un vell i nefast costum que tenim els informàtics: el de basar-nos en idees simples (exemple: *intelligència = ordinador*), convèncer-nos nosaltres mateixos que aquestes simplificacions tòpiques són uns formalismes abstractes dels més sofisticats de la Matemàtica del segle XX (i del XXI) i, per si tot això fos poc, parlar oracularment, amb un posat magistral fàcilment confusible amb la simple arrogància. Entenc que tota aquesta confusió hagi passat perquè els fundadors, Turing i von Neumann (que deliberadament havien combinat Matemàtica i Biologia i que, per cert, van morir just abans del «cop d'estat» del 1956, de què ja he parlat en un altre lloc) tenien una cultura, visió i categoria intel·lectual altíssimes i van situar i definir el problema de manera admirable perquè nosaltres, que n'hem heretat la fama, vivim de les rendes i en fem barata en forma de simplificacions abusives –que es fa difícil pensar que Turing o von Neumann haguessin acceptat (de fet tots dos abans de morir s'hi estaven oposant fortament). De tot això, la pitjor cosa que ens ha pogut passar és que els científics que ens envolten –potser creient que les nostres simplificacions eren continuació natural de la profunditat dels fundadors– han acceptat acríticament les nostres sovint inacceptables proposicions, comen-

çant pels mateixos matemàtics, passant pels lingüistes i psicòlegs i acabant pels biòlegs. El que em demano és: per quins set sous els informàtics –que ja no tenim Turing ni von Neumann ni n’hem heretat el geni, la visió ni la profunditat– hem d’estar més a prop de contestar la pregunta «*Què és la intel·ligència?*» que no pas els neurobiòlegs que estudien el cervell? Com que se m’escapa, no hi entraré, però sí que voldria tocar un tema que s’encavalca entre la Informàtica i la Biologia, que és el que anuncia el títol d’aquest article. Hi pot haver *ordinadors emotius*? Cal? El «sentiment», forma part de la intel·ligència? I la «consciència» (sigui el que sigui)? I, finalment: tot això que fan/faran els ordinadors (per cert, què?), és realment tan nou com sembla? Quin sentit té o pot arribar a tenir? I com ens ho prendrem nosaltres, pobres humans? Heus aquí algunes reflexions que jo m’he fet per a consum propi, per si serveixen a algú.

### *Sabem adaptar-nos a les màquines?*

Parlar d'*ordinadors emotius* s’ha convertit en una moda científica últimament. A l’MIT de Boston o a l’IAS de Princeton se’n parla. Fins i tot se n’ha publicat un llibre, titulat *Càlcul afectiu*, l’autora del qual, Rosalind Picard (precisament de l’MIT) va venir a Barcelona a parlar, l’u d’octubre del 99, d’«ordinadors emotius». La pregunta és: per a què volem aquesta mena d’ordinadors? Per a què podem voler ordinadors que tinguin emocions? Els necessitem?

Doncs, ves, la paraula «necessitar» és molt difícil de definir. No sé de ningú, als nostres països, que realment sàpiga què «necessita» (o què «vol») i què no. Per això ho plantejaré al revés: ja que els ordinadors *ja els tenim*, la pregunta seria: bé, i ara, què en fem? Sabent les possibilitats que donen, per a quin fi (nostre) podríem fer-les servir? L’ideal últim fóra, és clar, que nosaltres acabéssim una mica més feliços o humans que no érem al començament.

Començo, doncs, per fer-me una pregunta senzilleta i més modesta: com podem aprofitar les múltiples *màquines* de tota mena –des de pales mecàniques a rentaplats– de què disposem? I fins a quin punt elles són ajudants nostres o nosaltres en som esclaus? Doncs això, francament, encara no ho tenim gaire clar. I, sobretot, no l’hi tenim amb les tan recents i atabalants màquines que

anomenem *ordinadors* (i que, com tothom sap, canvien de prestacions, potència i característiques tan ràpidament que cap persona normal –especialitzada o no, tant hi fa– no les pot seguir sense marejar-se, i menys pretendre entendre-les). Des d'aquest punt de vista l'*ordinador* contrasta, potser, amb l'*automòbil* que és una altre invent que, tot i ser més que centenari, encara no acabem d'entendre. I això que en aquest cas es podria defensar que sí, que ja comencem a tenir una idea una mica clara (no pas gaire) de les seves possibilitats i dels seus límits (tot i que, com passa amb els ordinadors, no paren de canviar, i sempre *a més*): per exemple, ja comencem a saber i acceptar que els cotxes són bàsicament eines per traslladar-se d'un lloc a un altre, i no sols (a) per presumir de poder, virilitat o no se sap ben bé què, (b) per trencar-s'hi la carcanada –la nostra i la d'algú que, pobret, passava per allí–, (c) per estar corrent sempre en competició amb –o contra– algú (real o imaginari, o amb prou feines entrevist a través dels vidres) o (d) per vés a saber quina altra cosa. I això ens passa amb una màquina que es va inventar fa més de cent anys i a la qual tot just ara comencem a posar al seu lloc –i amb penes i treballs dolorosíssims i injustificables, com demostren els esgarrifosos nombres de morts i asseguts en cadires de rodes (per una raó que les mateixes víctimes no sabrien pas explicar-nos). Doncs si aquest necessari procés d'ajustament a les mides, possibilitats i perills d'aquesta eina, el *cotxe*, que hem vist sempre i que ens resulta aparentment tan familiar i fàcil de dominar ha estat tan penós, què podrem dir d'una eina com és l'*ordinador* que pràcticament acaba d'aparèixer, que tot just ara comencem a entendre (de fet, potser només ens ho sembla, que l'entendem), que té tantes possibilitats que no ens en podem imaginar ni l'u per cent, que fàcilment fa més la sensació de tenir-nos dominats ell a nosaltres que no pas a l'inrevés? I si hi ha tants conductors auto-dependents, com no hi ha d'haver ja un nombre inimaginable d'addictes a aquesta maquineta que se'ns ha ficat a casa (bé, la TV ja ho havia fet abans) i que ens promet –i només és un començar– l'accés pràcticament instantani a tot arreu del món i a tota persona imaginable (i fins i tot a mons i persones «virtuals», que no existeixen)? Si la potència dels cotxes ha matat més persones que totes les guerres, què no haurem de fer per defensar-nos i no deixar-nos xuclar per una màquina tan

inconcebiblement potent ara mateix (i no parlo del futur, perquè no me'l puc ni imaginar), enlluernadora i addictiva? I això que acabo de fer només és una consideració trivial, a l'abast de qualsevol usuari, que no pretén de cap manera espantar ningú.

De fet, jo sóc optimista i crec que hi tenim molt més a guanyar que a perdre amb aquestes màquines. Ara, dit això, cal no enganyar-se: (1) la nostra adaptació, individual i social, als ordinadors serà tan o més traumàtica –i a un ritme tan ràpid que molts es quedaran pel camí– que la que hem tingut i estem tenint amb els cotxes, i (2) la societat que en sortirà de tot plegat (i nosaltres com a individus) serà totalment irrecognoscible de dècada en dècada: produirà traumes –sobretot a qui no s'hi adapti– molt més grans que els que rebria el senyor del 1870 que aparegués de sobte aquí amb la seva levita i la seva ignorància de la llum elèctrica, de l'automòbil, dels avions, de la TV, de la potència mortífera que hem generat o dels telèfons mòbils que hem inventat per dir milions de bajanades des de qualsevol lloc.

### «Emotivitat» en/amb ordinadors?

La qüestió dels sentiments relacionats amb els *ordinadors* ens afecta per dues vies, molt diferents, que cal distingir:

1. Ha arribat un moment en què els tècnics que ens dediquem a dissenyar ordinadors i fer-los servir ens hem començat a plantejar si els *sentiments* poden ser realment útils de cara als ordinadors; vull dir, per al *seu propi* funcionament, independentment que nosaltres hi siguem o no. Els animals tenen «sentiments» (o cosa similar) i nosaltres també. No és casualitat, això. L'evolució ens ha preparat i ens ha portat aquí. Els sentiments no són una cosa gratuïta sinó adaptativa, que l'evolució ha produït o fomentat perquè complien unes certes funcions. Per això és legítim pensar que potser aquestes mateixes funcions que compleixen els sentiments en els animals i en les persones també es podrien incorporar a un ordinador i fer-lo funcionar millor. Com es veu, és una qüestió purament *tècnica*, encara que sembli absurd parlar dels sentiments com d'una «qüestió tècnica». Però és així: els ordinadors potser funcionarien millor –resoldrien millor

alguns problemes, diríem– si els dotéssim de sentiments. No deixa de tenir la seva lògica. Ara, una altra cosa, ben diferent, és plantejar-se aquesta pregunta:

2. Què fer amb els ordinadors de cara a nosaltres? (que nosaltres sí que en tenim, de sentiments). Els ordinadors poden arribar a ser –vaja, ho són– totalment insensibles, però nosaltres podem demanar-los que es portin amb nosaltres d'una manera més o menys «sensible». Aquí ja no es tracta de fer funcionar els ordinadors millor o pitjor amb els sentiments, sinó de nosaltres, que sí que funcionem millor –amb l'ordinador o amb el que sigui– amb sentiments, ja sigui perquè ens hi sentim acompanyats, o compresos, o perquè allò que nosaltres busquem (perquè *buscar* alguna cosa és el que fem, quan treballem amb ordinadors) l'ordinador hi sigui sensible, i ens ho doni, o fins i tot ens ho anticipi. Aquí no queda res descartat: podria ser que la màquina canviés la seva «actitud» amb nosaltres pel fet, per exemple, de trobar-nos la veu trista o de veure'ns nerviosos, inquiets, insatsfets amb els resultats, etc.

La primera i la segona són dues coses completament diferents. La fi del segle XX té la particularitat que per primera vegada han coincidit els dos aspectes en el temps, i per això el curs 1999-2000 s'està caracteritzant per parlar tant de *sentiments* en els ordinadors, encara que l'expressió sembli un contrasentit o una provocació. En tot cas, la moda ve de dos orígens diferents: el primer, purament tècnic: el bon rendiment dels ordinadors com a màquines; l'altre és més aviat nostre i de caire «psicoterapèutic»: donades les estranyes i potser amenaçadores eines que estem creant a un ritme tal que sembla que, més que inventar-les nosaltres a elles, siguin elles les que ens diuen com les hem de concebre, poder deixar de veure-les com a entitats forasteres i trobar-nos-hi més a gust, menys amenaçats i més amb pau.

El *problema* de la relació entre els ordinadors i nosaltres en realitat el creem nosaltres: de fet som *nosaltres*, el problema. Tenim la tendència a projectar-nos sobre tota mena de coses i de vegades ens sembla que hi veiem «problemes», problemes que realment no existeixen. Però nosaltres els «veiem» com si hi fossin i ho fossin de debò. Amb les màquines «intelligents» això passa molt sovint.

Com que poden semblar «humanes», és fàcil que ens hi projectem i acabem veient-hi el que no hi ha. La màquina acaba essent una excusa per discutir entre nosaltres sobre temes que no tenen res a veure amb la màquina sinó amb nosaltres, les nostres pors o les nostres esperances.

Hem acceptat que l'enginyeria, concretament la Intel·ligència Artificial, ens hagi proporcionat estris o màquines que tenen més memòria, millor càlcul i velocitat, i més capacitat d'interrelacionar i organitzar material que nosaltres. Això ho hem acceptat, i tenim màquines utilíssimes per resoldre bé aquestes qüestions. Manequem diners, els traiem d'allà on sigui sense que se'n perdi en cap moment la pista, fem reserves de bitllets des de qualsevol indret, obtenim informació, juguem a escacs –i aprenem a jugar-hi millor–, planifiquem decisions exageradament complicades o aparentment insolubles i, fins i tot fem «guerres intel·ligents». Tot això ho hem acceptat de bon grat: els usuaris ho veuen bé i ja ho tenen ben assimilat. La qual cosa porta a qualsevol persona a plantejar-se el següent: no es podria pensar que afegint al que ja tenim un petit detall les màquines que tenim millorarien? Aquest «petit detall» foren els «sentiments». Introduir mecanismes que processessin o expressessin emocions no sembla pas que hagués de ser un gran problema d'enginyeria i, en canvi, si es fes humanitzaria molt aquestes màquines. La pregunta és, doncs: per què no es fa?

I la resposta és, en part, que de totes les aplicacions tan útils dels ordinadors només algunes, relativament poques, es podrien qualificar de pròpiament «intel·ligents». En realitat, la majoria només són o càlcul matemàtic massiu –brutal, fins i tot– o aplicacions pràctiques sofisticades de la informàtica avançada. I de tant una cosa com de l'altra amb prou feines se'n pot dir, sincerament, «Intel·ligència». Les màquines citades, tan reeixides i satisfactòries, només vénen d'una certa tradició dintre d'aquesta ciència: la d'intentar «reproduir» la intel·ligència humana i, per fer-ho, construir aparells que calculin, que «pensin» en el sentit de tenir capacitat de raciocini i decisió. Són, això sí, grans ajudes; però... «intel·ligents»? Perquè tinguin intel·ligència segurament cal alguna cosa que de moment se'ns escapa. Caldria, almenys afegir-hi algunes altres condicions sense les quals estrictament no hauríem de fer passar cap sistema com a «intel·ligent». Una de bastant clara és que el

coneixement que tingui la màquina li permeti afrontar els problemes de manera similar o equivalent a com els afronta l'ésser humà (trobant-hi solucions tan simples com nosaltres quan resolem, per exemple, el problema de caminar sense topar amb obstacles). I això encara estem molt lluny d'aconseguir-ho. No sabem com organitzar el coneixement de cap màquina d'una manera adequada perquè es comporti de la manera que hem dit. Des d'aquest punt de vista, incorporar les emocions representa incloure un tipus d'informació que abans no es tenia en compte, però el procés de càlcul de la màquina continua essent el mateix.

Dit això, que és el teló de fons teòric i tècnic on estem situats en aquest moment, afegiria el següent: el problema –diguem-ne «gros»– que se'ns presenta és que tenim unes màquines a les quals nosaltres hem demanat *algunes coses* aparentment simples però amb moltes pretensions. (Exemples: validar càlculs o raonaments llargs i complicats, plantejar i seguir estratègies, jugar a jocs –i guanyar-nos!–, «parlar» amb nosaltres, acompanyar-nos en les nostres decisions, etc.) Algunes les hem fetes i n'hem dit «intelligents» però com que demanàvem coses que nosaltres no érem capaços de fer –*i les feien!*– teníem la temptació de dir que *nosaltres* havíem fet màquines *intelligents*, fent més exhibició d'afartament que no pas de ganes de menjar. Aquesta primerenca i pretensiosa visió de la nostra potència reproductiva és el que en podríem dir Intel·ligència Artificial «tradicional» o «antiga» (o «primitiva», si no ningú se'n sent ofès). És la que els experts veiem com el *màxim* a què es pot arribar de moment, de tal manera que percebem que intentar continuar per aquí sense canviar radicalment els mètodes és com topar contra una paret, i que gosar trasbalsar aquests plantejaments tradicionals ens portarà a un fracàs, que alguns ja presenten.

Així doncs, alguns creuen (creiem) que ens cal una Intel·ligència Artificial «nova», i que ens convé encetar-la ràpidament, perquè si no ja no podem pretendre més temps que el que construïm sigui realment «intelligent». En què ha de consistir, aquesta cosa *nova*? Doncs jo no ho sé. No sé si serà una nova etapa en l'evolució del que tenim (allò que dèiem més amunt: s'hi afegeixen, per exemple, sentiments –gran novetat, i benvinguda, i probablement necessària– però deixem la resta com estava), o implicarà algun canvi real (i



llavors potser ja no podrem deixar la resta «com estava») o serà realment una *ruptura* respecte de tot el que anat fent fins ara.

### *Les màquines «ens» envaeixen*

El que és segur –i ja ens en comencem a adonar– és que per fer algunes de les coses que demanem a aquestes màquines ja *entrem* en un terreny que *fins ara* era *exclusivament* nostre (vull dir, *dels humans*). Algunes de les característiques demanades sí que trepitgen efectivament el camp i les propietats de la intel·ligència (entesa com a *intel·ligència humana*, que fins ara era l'única que existia, o si més no el model de tota altra). De fet, siguem sincers, la intel·ligència no sabem què és. Ja fa més d'un segle que uns psicòlegs van fer una escala per mesurar-la (la de Stanford-Binet). Però no hem pogut anar gaire més enllà: sabem reconèixer –intuïtivament– la intel·ligència (o, més ben dit, alguns dels seus atributs), però no la sabem caracteritzar ni, menys, definir. (En això ens passa com als biòlegs amb la *vida*.) En aquests casos queda bé de dir que la intel·ligència és un fenomen multifactorial (o *n*-dimensional), però –a part el joc de paraules– aviat quedem en evidència quan la Neurociència cada dia que passa ens revela un detall més de la seva complexíssima estructura (parlo de la humana) i dels seus meravellosos engranatges materials, gens conformes a cap de les idíl·liques teories –i, menys que cap, a la d'«el cervell és com un ordinador», de perdurable i nefasta influència– que ens havíem inventat per sortir del pas. Repeteixo: de què pugui ser –o en què pugui consistir– la intel·ligència, francament no en tenim gaire idea. Ara bé, algunes llambregades sí que percebem. (Potser les hauríem pogut tenir nosaltres sols, reflexionant sobre què és la intel·ligència, però l'existència o pretensió de màquines intel·ligents ens ha obligat a pensar-hi i ens ho ha fet potser més evident o urgent.)

Una és aquesta: una característica òbvia de la intel·ligència (humana) és que no tolera gens bé les repeticions, i menys quan un s'adona que està perdut en un cercle viciós. Posem-ne un cas: quan a algú se li diu «fes això, i fes-ho quaranta vegades» arriba un moment, segur, que abans de les quaranta vegades, la persona en qüestió se sent ruc i sospita que li han aixecat la camisa. És més que

probable que a les vint vegades s'ho replantegi i es digui: «escolta, fa vint vegades que estic fent la mateixa cosa: no fóra pas qüestió de fer-ho d'una altra manera –més efectiva–, o de canviar de tema (i jurar no tornar a deixar-m'hi ensarronar mai més), o de fer veure que ho faig i imaginar-me'n un resultat plausible, o d'engegar n'orris a qui m'ho ha fet fer, o de buscar un mètode més ràpid o efectiu de sortir de l'embolic sense que es noti,... ?» Això és la intel·ligència, i segurament la seva manifestació més característica. És la definició favorita del matemàtic Hofstadter, que ell resumeix en BOOL (acrònim de *branching out of loops*, és a dir, «evitar les repeticions que no van enlloc»). Potser sí que –deixant la conyeta del BOOL– realment és això la intel·ligència. Perquè fer una cosa 40.000 vegades, això ho fan les màquines i ho fan les formigues, i no sé si es pot considerar gaire intel·ligent. Una altra manera de dir el mateix: la intel·ligència consisteix a canviar les regles del joc quan es detecta que el procés és massa repetitiu o s'ha tornat «mecànic». Els tècnics presumptuosament d'això en diem «fer un pas al següent metanivell», i correspon al que va fer Copèrnic quan va decidir que, perquè li quadressin les coses i els planetes, s'havia de deixar d'epícles i andergues, per molt aristotèliques i beneïdes que fossin, i canviar de perspectiva: posar-se al lloc del Sol, veure la Terra des d'allí i deixar-se anar. Això, francament, dubto que les màquines ho puguin fer mai. Ja els podem dir, ja: «quan tinguis un problema o t'avorreixis massa, passa a un metanivell», perquè elles no saben –com els ho podríem fer «imaginar»?– a quin nivell han de passar, o els «convé» fer-ho. Doncs bé, una situació com aquesta (la de tenir màquines d'aquestes) la trobem a faltar, i ens fa l'efecte que encara som molt lluny d'arribar-hi. I ens convé molt tenir una màquina com aquesta, realment intel·ligent, perquè quan es fa palesa o necessària la intel·ligència (o «expertesa») d'un ésser humà és justament imprevisible el nivell al qual ha de saltar. En circumstàncies extremes, una persona és capaç de fer allò que qualifiquem de comportament heroic, imprevist, sorprenent, genial, estúpid o qualsevol altre (bo o dolent, adaptatiu o rebutjable i, fins i tot, assassí o genocida), sovint sospitat o imaginable, però, en rigor, sempre imprevisible. Poso un parell d'exemples de despropòsits, alhora catalans i hospitalaris: un sistema expert de diagnòstic mèdic situat al servei d'urgències d'un hospital –que fins aquell

moment havia funcionat a satisfacció de tothom– intenta mantenir estable la pressió sanguínia del pacient fins que «no s’adona» –o no li diu a ningú– que aquest acaba de morir: en comptes de notificar «pacient mort, deixo de controlar constants vitals» i canviar de tema es posa a fer pujar la pressió frenèticament; un altre cas: de sobte es declara un foc a Urgències, i el sistema continua fent la seva feina sense ni sospitar que quan es cala foc el que cal fer és oblidar-se de la pressió –o de qualsevol altra relativa fotesa– del malalt i treure’l comes-ajudeu-me cap a un lloc segur –però, és clar, això és un metanivell prèviament «inimaginat» pel sistema, que d’altra banda no té prou capacitat d’«iniciativa» per sortir-se’n de manera mínimament airosa.

I tot això que dic és bastant obvi: no cal ser cap tècnic per adonar-se’n. Una altra característica de la intel·ligència –a part aquest del «canvi de nivell» que acabo de descriure– és aquesta: la intel·ligència consisteix bàsicament, també, en una «autosituació». L’èsser intel·ligent –humà o no (per evitar l’ambigüitat home-màquina avui es tendeix a fer servir el tecnicisme *agent*)–, com que se suposa que actua entre d’altres, immers en una constel·lació d’(altres) agents (cadascun amb el seu propi pla d’actuació), ha de saber o suposar i predir què faran previsiblement els altres agents i «situar-s’hi» adequadament (la paraula tècnica és *autosituació*). Els agents així construïts, encara força rudimentaris, els americans els anomenen «autòmats autosituats» i es poden considerar, segons com, a mig camí de la intel·ligència. Perquè cal tenir en compte que per a nosaltres la intel·ligència és (també? sobretot?) «posar-nos al lloc que toca». Quan ens trobem davant d’una persona (perdó, *agent*) nova, ens demanem: «vejam, aquesta persona, què deu saber?», fem unes quantes deduccions i suposicions i tot seguit actuem en funció d’allò que imaginem que aquest agent sap i li és familiar. Naturalment, si l’agent en qüestió resulta ser un extraterrestre, els punts de contacte, pressuposicions compartides o coneixements comuns són aproximadament, o exactament, zero: caldrà explicar-li-ho de seguida i penosament *tot*, incloent moltes obvietats que mai hauríem pensat d’haver d’explicar a ningú. Si, a l’altre extrem, la persona és un veí i, per tant, hi compartim coneixements, suposicions, hàbits, etc., molt probablement no li haurem d’explicar pràcticament res. En qualsevol cosa, la comuni-

cació humana, entre agents, tendeix sempre a ser mínima. L'únic que hem de fer és plantejar-nos el següent: «aquest senyor sap o suposa o està pensant  $x$ , i jo voldria fer-li veure  $x'$  o  $y$ ». Doncs bé, la comunicació, que és el procés pel qual provo d'aconseguir el meu objectiu, consisteix a transmetre la informació necessària per arribar-hi, i en el cas de veïns o persones pròximes la informació pot ser mínima, i ho sol ser (fins i tot amb un simple gest pot haver-n'hi prou). Si es tracta del veí (és a dir, si compartim el context o jo he fet l'esforç de situar-me en el seu context de coneixements o fer-ne suposicions raonables) nimietats òbvies tal com recordar-li la ciutat on som, el temps que fa, l'hora del dia, la relació que hi ha entre nosaltres, etc., tot i ser potser molt importants pel que li vull comunicar, són perfectament supèrflues i estalviables. Jo sé allò que hi ha implícit en el que li dic, jo sé que ell ho sap, hi ha una sèrie de trivialitats que compartim (com ara: avui fa calor, o no, ens vam conèixer tal dia, ell coneix això o sap allò d'aquell, etc.) i jo sé que les compartim, etc. Això fa la comunicació senzilla i fluida. En canvi, si el meu interlocutor és extraterrestre segurament hi hauré d'entrar en prolixes i penoses explicacions prèvies sobre fets dels qual ni tan sols tenia consciència que fossin obvis, tals com que hi ha gravetat, i pluja, i certes característiques rellevants de les relacions humanes, l'hora que és o el lloc que ocupem i la importància que això pot tenir, etc. Doncs bé, tot això és, o *també és*, intel·ligència, i les màquines intel·ligents que vulguem construir, les haurem de fer capaces d'*autosituat-se*, i això vol dir que haurem de calibrar quin «paquet» inicial de coneixements els hi haurem de subministrar. Per exemple: els hem d'explicar que existeix la pluja? o que els dracs són animals –d'una mena cultural i estranya, per cert? És rellevant que ho sàpiguen tenint en compte el que els hem de dir tot seguit? I, si ho és, els ho hem de fer saber explícitament? I com ho fem? Amb una il·lustració, o axiomàticament, o amb una definició (contràriament al nostre cas, en què els dracs els vam conèixer a través d'històries)? O potser els haurem d'«explicar contes», als agents autosituats, perquè s'«autosituïn» davant dels dracs i dels seus múltiples usos i significats culturals, simbòlics, patriòtics, etc? O dir-los que som a Europa? (caldrà?). O que les coses, si no s'aguanten, cauen –i sovint es trenquen, cosa que generalment se suposa que cal evitar (i que nosaltres potser vam

aprendre a base de bufetades)–, etc.

### *La moda dels «agents»*

Tot això no té res a veure amb el fet o la possibilitat que les màquines intel·ligents (perdó, dels *agents*) hagin de ser *autònomes* ni, encara menys, que hagin de tenir conductes oportunistes, d'auto-preservació, o de perseguir el guany o l'avantatge propi davant d'altres agents. Jo, francament, això no crec que estiguem pas ni remotament en el punt de plantejar-nos-ho. A partir del que acabo de dir de l'«autosituació» –que no passa de ser un tipus de conducta *reactiva*– no voldria extrapol·lar ni pensar que de moment puguem dissenyar res que vagi gaire més enllà: això de tenir *autonomia*, o «motivacions» o tipus de conductes pròpies em sembla que ens cau, de moment, força lluny. Jo només parlava d'una cosa tan simple, i tan complicada de reproduir en una màquina, com és la *contextualització* (que la màquina sàpiga en quina situació es troba, i què pot fer i què no, i què se n'espera, etc.). Això és elemental, i només cal pensar en els ridículs espantosos als quals podem arribar els humans quan no sabem quin és el nostre lloc, o no haver mesurat davant de qui estàvem, o quin registre havíem de fer servir davant de què o de qui, quin nivell de vocabulari (ric, o no) i tipus (científic, vulgar, groller) fer servir segons com hàgim calibrat l'interlocutor, quines normes rituals o d'etiqueta havíem de seguir, etc. (Penseu en els esforços d'adaptació que hem de fer en els viatges o per anar a un país nou o per parlar o conviure amb persones d'una altra cultura). Doncs resulta que aquest esforç –que en si mateix no pressuposa intel·ligència, però que prepara el bescanvi entre continguts i éssers intel·ligents– és no sols bàsic sinó imprescindible per fer fluida la comunicació humana. I per fer-ho no cal, ni de bon tros, que aquesta màquina autosituada hagi de tenir ni *autonomia* ni tan sols allò que diem *sentiments*.

I ja que parlem d'«autonomia» i «sentiments», fóra bo recordar l'esfera de cada concepte i les seves relacions. Els *sentiments* no cal que depenguin de l'autonomia o de finalitats pròpies (es pot ser un perfecte esclau i tenir-ne, de sentiments) ni tan sols de res més que d'un mer conjunt de *regles* que el programador pot haver introduït en la màquina (i això ja ho estan fent avui mateix, a

Stanford) per fer-la més «simpàtica», ja sigui adoptant actituds amables o donant explicacions més senzilles –quan sigui obvi que les prèvies no s’han entès o s’han malinterpretat– o, en general, mostrant una certa *empatia* cap a l’usuari, segons l’estat d’ànim que hi detecta (cosa que, d’altra banda, no és tan difícil de mecanitzar com sembla: de fet, ja s’ha està aconseguint, com he dit, a Stanford). De les característiques que he intentat donar successivament del que ha de tenir un ordinador veritablement intel·ligent, això últim –i, francament, només això últim (la simpatia o empatia aparent)– és el que tenim a l’abast de moment. Però, sincerament, d’una màquina que només fes això, per molt que des de fora ens ho pogués semblar, jo no diria pas que poguéssim qualificar-la de màquina *sensible*. Entre els humans, qualsevol persona no gaire intel·ligent, o gens, pot ser ensenyada perquè somrigui tothora i faci veure que es preocupa d’allò més del client que té al davant, però tots sabem que d’aquesta amabilitat clientelar a la intel·ligència pròpiament dita hi ha un bon tros.

Si parlem d’*ordinadors emotius*, la primera qüestió és: és fàcil, de fer-ne? La meua resposta és, com s’ha vist molt matisada: hi ha molts components en l’emotivitat. Els enumero, de més fàcil a més difícil: (1) obtenció més o menys mecànica d’empatia, (2) capacitat més o menys automàtica –i encertada– d’«autosituació» i, finalment, (3) avaluació «global» (diríem que «sentida» –i aquí ja parlaríem per primera vegada d’«emotivitat» de la màquina). De la part no pròpiament reactiva, és a dir, dels motius o de l’*autonomia* de la màquina, no en parlo perquè crec que surt fora del tema de l’emotivitat per endinsar-se en un altre de ben diferent (i que no el pressuposa: podem concebre màquines autònomes, amb motivacions tan predefinides i precises com es vulgui i que no tinguin cap mena de sentiment –i no cal que em remeti a exemples humans que ho il·lustrarien ben bé).

La segona qüestió és: com estem de lluny, de tenir ordinadors més o menys emotius i/o autònoms (en algun dels graus esmentats)? Podem fer-ne a petita escala, que s’hi comportin? Tenim ja la metodologia per resoldre-ho? La veritat és que no, francament. Aquesta és, segur, una opinió personal, però la baso en una situació que em va tocar viure. Jo, que sóc una persona que em dedico a aquesta feina i que em caracteritzo entre altres coses

per un accentuat sentit del ridícul, vaig tenir l'ocasió fa uns anys de preparar, per encàrrec d'un congrés sobre la matèria, un curs –«tutorial», se'n diu, d'això– sobre l'estat de la qüestió en màquines emotives i autònomes, part de moda dins del tema de moda, els *agents intel·ligents*. Jo ho vaig enfocar des del punt de vista tècnic, de les eines (lògiques) de què es disposa per tractar aquest tema i resoldre'l, unes tècniques anomenades *lògiques BDI* (*BDI Logic*, que vol dir 'lògica de creences, desigs i intencions'), però vaig aprofitar per recollir tota mena d'informació sobre tot el que s'estava fent en aquest tema. Doncs vaig veure que el que s'està fent és de tot punt elemental i ridícul (i no ho trobo ridícul perquè faci riure sinó per la desproporció entre el que es diu que es té i el que es té realment), tot i que la gent que en veu els resultats pugui quedar-ne molt sorpresa i impressionada i, de ridícul, no trobar-l'hi pas gens. El que es fa és, dit pelat i nu, agafar un ordinador o robot (diguem-ne «agent» per concessió a la moda) que potser caminarà o buscarà o vés a saber què farà o amb qui interactuarà, però que jo programaré com un ordinador dels de tota la vida: «fes això», «fes allò», «ara passa aquest encàrrec o missatge a l'agent *p*», «si passa *x* fes *y*» i trivialitats diverses del tipus conegut (i abusat) dit *sistema de regles*. O sigui, que la programació és com la de tota la vida, la que va inspirar el concepte mateix de «programació» al seu autor, Alan Turing, el 1936: la metàfora del capatàs, que va dient a tothom: «tu, fes això!», «atenció, si passés tal fes qual», etc. L'únic que aquí ho «enriqueim» –i d'aquí ve el meu sentit del ridícul a l'hora d'explicar-ho– amb «opinions» (que se suposa que té la màquina sobre l'estat del món i les intencions dels altres agents), «desitjos», «intencions», «objectius», etc. Per cada conjunt d'aquestes coses diverses es prova de muntar una estratègia per obtenir l'objectiu proposat tenint en compte una sèrie d'hipòtesis, plausibles però força simplistes, sobre com es comportarà el món i els altres agents mentrestant, i finalment avaluem el rendiment de tot plegat i si de cas ho corregim o diem que n'hem «aprens» i que a la pròxima les coses aniran més rodades. No és pas que jo trobi particularment que aquest no sigui un enfocament correcte, si bé fortament ingenu i rudimentari (però per algun lloc cal començar!); no: el que em fa sentir ridícul és que tot sigui –o es cregui que hagi de ser– tan racional i explícit, com si el programador fos un personatge omnis-

cient (i «prescient»), un matemàtic (amb les fórmules i solucions a punt), un –literalment– «programador» (etimològicament, en grec, ‘un que ho escriu abans’), com si no hi pogués haver sorpreses (precisament nascudes de la mateixa actuació «programada» de l’agent o de la rica i complexa interacció entre agents), obstacles insospitats, canvis de pla –com el pas a metanivell «incendi» en un servei d’urgències, recordeu?–, situacions del tot imprevisibles o imprevisibles –que és justament on els éssers intel·ligents demostren la seva intel·ligència–, etc. Doncs tot això es preveu (programar és això), s’avalua (i d’això en direm abusivament «sentir» o «satisfer» els «desitjos»), etc. I m’estalvio de parlar d’altres vergonyes (personals, meves) com l’intent de definir, repertoriar i categoritzar les «pulsions» (o «instints») bàsics que «han» de motivar una màquina.

No nego que potser els agents acabaran essent «emotius», «intel·ligents» (per convenir una paraula) i fins i tot «motivats». El que em costa moltíssim de creure és que siguin una mena de versió clònica mecànica d’un matemàtic o d’un filòsof tothora imperatorable, intemporal i racional al límit. No sols dubto que sigui possible. També dubto que sigui desitjable. I tot plegat em deixa un gust amargant de simplificació abusiva d’objectius i de mitjans, de motius i de conductes. I, francament, com a persona molt tendencialment racional i racionalista no crec pas en absolut (com tampoc creia Russell i molts altres lògics) que la intel·ligència sigui necessàriament reduïble a l’aspecte racional ni que fer-ho sigui bo per ningú. Penseu que dir a una maquineta, com sembla que no tenim inconvenient a dir, «si tens el motiu  $x$  –que efectivament tens, perquè quatre instruccions més amunt ha quedat prou clar– i la situació en què et trobes és aproximadament com la  $y$ , aleshores el que has de fer és això i allò», i així una pila més de benèvolos consells per atènyer finalment l’objectiu que he/ha/hem/s’ha proposat d’aconseguir el sistema enmig d’un món complex, imprevisible i –malgrat tot– simple, prescrit i lineal. A mi personalment em fa l’efecte d’un escamot hitlerià; al cap i a la fi, les SA i les SS també estaven programades així –i no hi calien ordinadors–: «si et trobes pel carrer un agent enemic o dubtós o impresentable o desplaent i mostra por o culpabilitat o es deixa estovar, doncs estova’l o detén-lo o actua-hi discrecionalment». D’això, ja en sabem. I per a això no cal/calía intel·ligència. No ho presentem doncs com un avenç



de la tècnica i en diguem *lògica BDI* i ho presentem en congressos. Programar persones perquè actuïn segons objectius prefixats –que assoleixin flexiblement, «intelligentment»– ho hem fet un munt de vegades en un munt de llocs i circumstàncies, massa. Però per això no necessitem ordinadors. Desgraciadament tenim molts candidats humans –«intelligents» i molt probablement també «sensibles»– que ens poden fer la feina molt ben feta i bé de preu.

*Doncs potser que ens aclaríssim, abans*

Voldria acabar amb tres lots de consideracions barates sobre la qüestió:

1. No sé si la gent normal s'adona gaire –o gens– que les màquines intelligents no són en absolut res de nou. Fer-ne només ha estat possible a partir de l'any 1946, però la possibilitat, el concepte i la necessitat (tenir esclaus mecànics que facin la feina per nosaltres) ens ve de l'època dels grecs i, ja molt precisa i clarivident, des de mitjan segle XVII. No sé pas de què ens estranyem: les màquines les tenim descrites als llibres (Descartes, Hobbes, Pascal, LaMettrie en parlen, i Lull, Leibniz o Turing en fan un model abstracte), i l'únic privilegi nou que tenim ara és veure-les funcionar i, tot veient-les, sorprendre'ns del fet que el vell somni hagi estat finalment demostrat possible. El més sensat que podríem fer és, em sembla, deixar-ho aquí. Aprofitar les extraordinàries prestacions d'aquestes màquines –en velocitat i flexibilitat (el seu contingut d'«intelligència» és bastant més discutible)–, no extrapolar salvatgement a partir d'aquestes andròmines ràpidament obsolescents –i menys encara veure-hi més intelligència que no pas, simplement, sorpresa. I, si volem anar més enllà i tenir màquines que no sols ens ajudin sinó que ens facin companyia i «s'emocionin» i ens «estimïn», més val que ens ho prenguem per la banda tranquil·la: aquestes màquines ja existeixen. S'anomenen *persones*. Tenen el desavantatge que sovint s'equivoquen, s'enrabien sense motiu i són generalment de conducta difícil, interferent i incompreensible. Per a casos de disfunció greu, escassetat o inconvenients diversos tenen l'avantatge –però– de tenir succedanis, d'inconvenients i prestacions

més limitades –però èxit creixent– anomenats *animals domèstics*.

2. Em fa l'efecte que el que fem amb les màquines que construïm és no sols esclaus mecànics, que no s'espantllen mai ni s'equivoquen ni cotitzen ni paguen hipoteques. No. Em sembla que, en el fons, volem anar més enllà. És com si volguéssim construir robotets que fossin exactament com voldrien que fossin els nostres fills. Per exemple, jo vull tenir un «fill» que sigui metge, arquitecte, advocat, enginyer o el que es vulgui. En principi, la cosa sembla fàcil: li ensenyo tot el que ha de saber o que a mi m'hauria agradat que m'ensenyessin, l'«educó» (etimològicament, 'faig passar per l'adreçador'), procuro que sigui respectuós amb els meus hàbits i objectius, i que faci tot el que s'ha de fer, i bé. El problema és que els fills fan enrabiatar de valent, trenquen gerros, ens fan passar adolescències ferotges, esguerren els nostres millors plans que havíem fet per ells i acaben mals estudiants, mals metges/.../enginyers –això quan han acabat estudiant el que volíem–, s'equivoquen en les decisions que prenen, molt sovint fracassen (o ho així ho veiem) i fins i tot les relacions paternofiliales es poden capgirar o anar en orris. La solució (idealment)? Doncs construeixo una rèplica de mi mateix, l'agafo i li explico que ha de ser un metge/.../enginyer i li dono tot d'instruccions pertinents perquè ho sigui. Això sí, mentre no hi arribi no ha de trencar cap gerro ni fer-nos empipar ni renunciar a res nostre. I quan hi hagi arribat, ha de ser de l'especialitat que a nosaltres ens agradava, ha de ser professionalment perfecte, ens ha de fer quedar bé i en la seva vida professional –a l'altra, si n'hi ha, la possibilitat és optativa o tolerable– no hi ha d'haver lloc –de cap manera– per a la paraula *error* ni, molt menys, per a la paraula *fracàs*. Ja sé que cap persona tova (vul dir, de carn i ossos) acceptaria com a fill ningú que no se li assemblés –i menys un manyoc de fils recoberts de metall. Però si el fill de debò fos –es comportés– com una perfecta màquina d'aquestes que acabo de descriure, que aprengué sense molestar, trencar gerros, agafar rebequeries incomprensibles o discutir per bajanades, juraria que molts pares hi firmarien a ulls clucs. El problema és que això ja s'ha fet moltes vegades (més d'un *infant prodigi* ens ho podria explicar des de dintre) i que, a part d'haver arruïnat

la vida de la pobra víctima, el sistema no ha funcionat: el fill ha acabat trencant gerros compulsivament o en una comuna *hippy* fent sistemàticament tot el contrari del que li havien dit.

3. La gent tradicionalment carrega els neulers a la ciència i la tecnologia amb el papus que totes dues acabaran «programant» la societat, deshumanitzant l'home i desintegrant la cohesió i la solidaritat social. Doncs bé, a propòsit del que deia en un anterior paràgraf sobre els novíssims agents BDI, hi ha una altra cosa de què la gent normal no s'adona gaire –o gens–, i és que la capacitat humana de «programar» altres éssers humans per executar qualsevol mena de cosa –per injustificable o execrable que sigui– és no sols antiquíssima sinó que deu ser una de les habilitats sociopoliticomilitars més perfectes, perfeccionades i experimentades de què es disposa avui dia. I si algú no s'ho creu és que és vocacionalment autista o no té televisor a casa. I consti que no ve d'avui ni la TV hi té res a veure: la creació de clònics i autòmats socials ens ve d'una llarguíssima tradició de mil·lenis. L'única diferència és que, per imperatius d'una tecnologia incipient, tradicionalment han hagut de ser humans –i, doncs, raonablement intel·ligents– i, en canvi, ara, per primera vegada ja ens plantegem de prescindir de la part més tova i humida d'aquestes màquines, com si les preferíssim dures, resistents a tot, de pell llisa i metallitzada, i amb una sobrepotència afegida (no fos cas que se'n necessités). Vaja, com amb els cotxes.

Per això –i amb això acabo– quan parlem dels *ordinadors* i de la *intelligència*, i de si els primers en tenen/tindran/tindrien (una mica/bastant/molt) de la segona, sempre m'assalta el dubte de si realment estem parlant del que sembla que diem o bé estem parlant de nosaltres mateixos, i que *el problema* no és pas cap màquina –en tot cas la màquina en qüestió no passa de ser l'excusa de les nostres expansions– sinó *nosaltres*, exactament nosaltres, amb les nostres expectatives, limitacions, autoimatges (si fa no fa deformes), pors, exageracions, afany de notorietat, timidesa, supèrbia, etc. Per fer aquesta reflexió no m'hi falten referents. En qualsevol moment sento brunzir motos o cotxes que, sobrerevolucionats, passats en decibels o en km/h, fan curses contra vés-a-saber què o qui (o potser combaten contra el no-res?) per demostrar alguna cosa (a

algú?): tot plegat diu més del propietari i de les seves característiques i finalitats que no pas de l'eina –en aquest cas irrellevant– que s'ha triat per implementar-les. I el mateix que passa amb el cotxe centenari passa i passarà amb l'ordinador recent (cada vegada més «recent»): serà a cada moment allò que nosaltres voldrem que sigui. De fet, hem arribat en un punt en què les característiques d'una màquina que nosaltres construïm (intelligència, velocitat, sentiments, potència, o el que calgui) són merament instrumentals i no tenen cap importància. Avui la màquina –qualsevol màquina– (com la societat mateixa que anem configurant a poc a poc) és un mer *mitjà d'expressió* de l'home actual, fragmentat en individus. Només ens cal esperar que l'home del futur tendeixi cap a la bonhomia i la justícia i no cap a l'estupidesa universal, perquè qualsevol meta que es proposi –la que sigui– quedarà potenciada i multiplicada, precisament gràcies a les màquines, per un coeficient d'un milió, com a mínim. I segons quina sigui la característica triada, la cosa pot acabar en un paradís o en un infern. Per la banda pessimista la constatació de l'estupidesa o crueltat extrema amb què podem actuar (en certs camps ja en portem una certa pràctica de segles) ens pot fer témer el pitjor. Però per l'altra –la dels optimistes com jo (que creiem que per a cada situació sempre n'hi pot haver una de pitjor)– la constatació que, segons com i quan, els humans som miraculosament capaços dels més inesperats altruïsmes ens (em) fa tenir grans esperances. (Una altra qüestió, aquesta ben diferent, és el fet –individualment preocupant– que qualsevol adaptació que se'ns demanarà a cadascú de nosaltres, per gradual i fàcil que se'ns posi, serà tan diabòlicament exigent i ràpida que per a l'individu normal –el que amb el temps es fa vell, vaja– acabarà resultant-li terrorífica... i sort que un any ens morirem i deixarem d'haver-nos de posar al dia: com deia Einstein, la ciència progressa i afortunadament els vells es moren).

Per cert, a les dues situacions que he caracteritzat com a extremes, el «paradís» i l'«infern» que podríem arribar a tenir (tot i que la més plausible és una tercera, pura o mixta, inesperada i imprevisible), sortosament la Ciència-ficció ja ens hi ha preparat intel·lectualment i anímicament. Fins i tot ens ha descrit diverses versions de les dues, i n'ha calculat algunes combinacions lineals. Potser cal apreciar l'avantatge, si és que ho és, que tindrem: per

entendre aquests «cels» i aquests «infern» ja no caldrà estudiar religió: i si acaba essent un infern, serà un infern laic, amb aire condicionat i gratificacions instantànies. A mi particularment no em sembla pas que ens esperi cap mena d'alternativa «cel o infern», encara que siguin en modalitat laica, ni que calgui imaginar-s'ho real ni pròxim (ni de cap manera), i menys encara que ens hàgim de posar a tremolar pel que vindrà. Confesso que la meua és una hipòtesi optimista i, crec, *probable* (tot i que reconec que va a contrapèl de l'apocalipsi *ad usum*). Això sí, passi el que passi, no caldrà traslladar-se a cap «més enllà» ni passarà amb les nostres «ànimes» (de fet no caldrà ni que ens morim abans) sinó amb els nostres cossos, o el que en quedi. Jo més aviat m'inclino per suposar que aviat no tindrà sentit parlar de cels ni d'infern, i que el mateix repertori aquest dels símls religiosos s'autoextingirà per falta de correspondència entre les imatges que viurem i les que hem gastat fins ara. Si, malgrat tot, persistíssim a buscar-hi analogies religioses, o simplement apocalíptiques, m'inclino per pensar que el que ens espera es tractarà més aviat d'un «purgatori», si fa no fa com el que vivim en aquest tombant de segle (si bé jo confio que molt més passador, just i equilibrat). El que és segur, però, és que el que passarà no s'assemblarà gens a res que ningú hagi pogut anticipar ni imaginar (només cal saber una mica d'Història per adonar-se de les innombrables ficades de pota que hem fet amb tradicional persistència a l'hora de «predir» el «futur»: el «predit» s'ha semblat al «succeït» com un ou a una castanya).

I si hi ha algú que té tendència o voluntat d'espantar-se abans d'hora, que s'esperí. Les màquines que construïm solen tenir la gentilesa de desenvolupar les seves habilitats pel seu compte mentre la nostra limitadíssima capacitat de comprensió i adaptació —a allò que, paradoxalment, nosaltres mateixos hem fet i dissenyat!— es va lentament afinant i aproximant (a costa, potser, d'alguns accidents pel camí). No hi fa res que per entendre els nostres propis «engendres» incorrem en unes quantes forassenyades extrapolacions o triguem algun segle a posar-nos a to. Com deia un illustre pensador, no té sentit amoïnar-se per saber com passarem un pont si encara falten multitud de revolts i de milles (o segles) per arribar-hi... i això en el cas que n'hi hagi, de pont (que potser només ens ho imaginem)!

## *Xarxes neuronals, cervells i ordinadors*

**P**EL seu origen, aquest text correspon a la intervenció que l'Associació d'Estudiants de Neurociències de la Universitat Autònoma de Barcelona em va invitar amablement a fer el 14 de maig del 1999 sobre la «**Inspiració biològica dels models de la Intel·ligència Artificial**». Desgraciadament, com els vaig confessar obertament, la Intel·ligència Artificial com a disciplina ha tirat pel dret i no s'ha interessat mai, *gens*, per la **biologia** de res –ni tan sols del cervell. Ha estat sempre un afer exclusiu de matemàtics i informàtics unidimensionals, deliberadament autoexcloents i autolimitatius. El tema de la conferència era, per tant, vacu. Jo em vaig limitar, aquell maig, a explicar el que havia passat –que té unes antigues i molt potents arrels filosòfiques– i per què tot va anar com va anar. En aquest text que teniu al davant, en canvi, escrit amb la distància d'alguns mesos (i en el qual els lectors d'aquest Butlletí reconeixeran –i fins i tot veuran repetides *verbatim*– algunes de les meves dèries particulars i iteratives), no he volgut tornar a ser tan decebedor per als pobres biòlegs, que tenen tot el dret –com els informàtics, almenys– a demanar-se «què és la intel·ligència», ni tampoc he volgut tornar a donar la lliçó magistral de torn sobre la qüestió, costum que els informàtics exercim amb una certa freqüència (i arrogància) sense que, que jo sàpiga, ningú ens en hagi donat cap revelació, metaconeixement o permís que ho justifiqui. I tot i que la història continua essent la mateixa que els vaig explicar aquella tarda, ara ja no em veuria amb cor de tornar a donar a uns estudiants de Biologia entusiastes que, a propòsit del tema de la conferència (sobre la «inspiració biològica»), em

BUTLLETÍ DE  
L'ACIA  
NÚMERO 21  
ESTIU DE 2000

tornessin a preguntar «què?» la seca –si bé tristament certa– contesta «absolutament res!». Per això, i per quedar-me més tranquil de consciència, he volgut refer aquí l’argumentació d’aquell dia centrant-la en una comparació entre la Intel·ligència Artificial, que com se sap està basada pròpiament –fundacionalment, de fet– en el rebuig (explícit, sense contemplacions) del *cervell* com a model de la *intelligència* –i l’acceptació (cega, incondicional) de l’ordinador com a model de quasi tot–, i les anomenades *xarxes neuronals*, que ja fa un temps que són moda, que fa l’efecte que hagin aparegut recentment i que, com el seu nom sembla indicar, no s’avergonyeixen d’inspirar-se –no pas tant com sembla, però– en el cervell. Que s’entengui bé el que pretenc fer. Vull «fer dissabte» dels mals endreços acumulats, empolegats i encarcerats que els informàtics hem anat acumulat acríticament, redimensionar i posar al seu lloc els tòpics que fem circular i ens creiem –i que (i això és més greu) induïm a creure a d’altres científics (biòlegs, lingüistes, psicòlegs, etc.), que a una natural falta de criteri informàtic afegeixen una admiració immerescuda cap a nosaltres– i contribuir a abandonar aquest posat d’oracle que no sé per quins set sous ens ha concedit graciosament la comunitat científica, una admirativa comunitat que inexplicablement ha anat seguint les nostres intuïcions al peu de la lletra encara que fossin equivocades, delirants o simplement inaplicables (quan se’ls ha fet passar del camp de la simplificació matemàtica al seu propi).

### *Les «xarxes neuronals»*

Les **xarxes neuronals** són des de mitjans 1980 un tema de conversa entre científics no estrictament informàtics: des dels matemàtics als psicòlegs i lingüistes passant pels físics i biòlegs. Doncs ara resulta que amb les que els enginyers i els informàtics han creat es poden fer coses sorprenents –i se’n fan– com ara regular i optimitzar el trànsit a les autopistes o als aeroports, ajudar al diagnòstic mèdic en casos complicats –i al mig d’una intervenció quirúrgica!–, seleccionar parelles –és a dir, fer d’«agència matrimonial»–, fer simulacions climàtiques, elaborar estratègies militars, i moltes altres coses. Per això, com que ara tenen aplicacions importants, ja no en podem prescindir. I per això és bo de saber d’on vénen.

La història és aquesta. Les xarxes neuronals van aparèixer, sense aquest nom tan explícit, als anys 1940. Un biòleg, Warren McCulloch, inspirat pel cervell i les seves neurones, va aconseguir de concebre un sistema de *calcular* que fos més realista biològicament que els dels matemàtics, un model alternatiu –però equivalent– al concebut anys abans per Alan Turing, que és essencialment un esquema formal d'*ordinador* (màquina els primers exemplars de la qual precisament es començaven a construir aquells anys). Aquest últim sistema de càlcul és estrictament *lògic* i gens basat en cap similitud amb el cervell ni amb les neurones. McCulloch va formular el seu sistema «neuronal» només de manera teòrica, però als anys 1960 Rosenblatt va materialitzar el concepte en forma d'aparell (el 'Perceptron') i hi va obtenir uns primers resultats, prou bons i encoratjadors, si bé inopinadament un col·lega, Marvin Minsky, va denigrar tots els esforços de Rosenblatt en un llibre, la qual cosa va deixar pràcticament morta aquesta via de recerca. Després, als anys 1980, certes troballes d'orígens diversos van permetre una aplicació més sofisticada de les *xarxes neuronals*, com ara ja se les anomenava directament.

La diferència potser més vistent entre aquests dos conceptes, els *ordinadors* com els entenem habitualment per una banda i les *xarxes neuronals* per una altra és que en aquestes la «memòria» (les dades, els «records») i el processament cognitiu estan integrats, mentre que als ordinadors les dades i el processador estan separats. Una altra oposició entre ordinadors i xarxes neuronals és que els primers treballen amb *símbols*, mentre que les segones no (o ho fan a un nivell «subsimbòlic», amb unes entitats no simbòliques difícils de precisar que anomenem correntment *estats*). Hi ha més diferències, però com que ja hem esmentat les bàsiques, de moment deixem-ho aquí.

A tota aquesta història que tot just he esbossat, que és correcta i és la tradicionalment explicada, cal afegir-hi una dimensió humana, perquè la Ciència no és senzillament un conjunt de resultats o d'instruments sinó una Història en què intervenen persones normals. Per això vull complementar la versió estàndard de la història que acabo de donar explicant l'embolcall humà que hi va haver al voltant. La història en qüestió comença realment l'any 1936 amb l'anglès Alan Turing. Persona cultíssima, documentada, seriosa,



que no va merèixer la vida que li van donar l'*establishment* del seu país i època, Turing, a l'any 1936, en un d'aquests articles que fan època, va intentar entendre i explicar-se què vol dir *calcular*. Calcular i, de retruc, *raonar*. Perquè la idea de reduir «raonar» a «calcular» era vella ja en ple segle XVIII; de fet ja havia nascut força abans: Llull ja l'havia suggerit (i sabia ben bé què deia), i sense anar tan enrere Descartes, Pascal, Hobbes i Leibniz havien afirmat a mitjan segle XVII que essencialment *pensar és calcular*. Així, no ha de ser pas cap sorpresa que el 1936, l'any de Turing, es digués i es veiés majoritàriament així. De fet aquesta fecunda simplificació és la base de tots els ordinadors i de tota la informàtica d'avui. Hi ha una pila de coses que fem els informàtics que no s'entendrien sense aquest postulat bàsic, que torno a recordar: «*pensar*» és essencialment «*calcular*». Això tan simple, que situa els ordinadors i la informàtica estrictament en el camp del raonament rigorós regit per la Lògica, ha pogut acabar essent una trampa mortal (... però de moment deixem aquesta qüestió per més endavant). Turing va partir d'aquesta vella idea (raonar → calcular) i d'aquí es va plantejar lògicament «què és calcular?» La resposta, formal i rigorosa, és alhora un esquema senzill i comprensible. Tècnicament és el formalisme matemàtic conegut com a «màquina de Turing» que avui s'explica als primers cursos de moltes carreres. Com que Turing vivia a Anglaterra, lluny de Princeton, que és on s'havien plantejat i es discutien «seriosament» aquestes qüestions tècniques, la seva «màquina» no va ser gaire coneguda, encara que ara pugui semblar el contrari, fins que John von Neumann (personatge importantíssim sense el qual la meitat de les coses, en matemàtica i enginyeria, no serien avui com són), es va llegir amb molta atenció l'article de Turing, va copsar ràpidament la seva gran importància, en va voler conèixer personalment l'autor –al seu Institut d'Estudis Avançats de Princeton, al qual ell havia atret prèviament Gödel i Einstein, pels boscos del qual van estar estones passejant el modest Turing i el brillant exdeixeble de Hilbert– i en va fer tanta propaganda com va poder, anunciant que el mecanisme de Turing, amb capacitat de càlcul (i, per implicació, de «raonament») prefigurava un nou tipus de màquina que, si mai es construïa en una versió «real» (a dissenyar la qual von Neumann va col·laborar decisivament després de la guerra), prefiguraria la

realitat de la primera «màquina intel·ligent» i d'una nova societat alliberada de les rutines intel·lectuals. (De fet, ho va acabar essent: és l'*ordinador* d'avui.)

Tanta era la influència científica i social de von Neumann que el 1943, en plena guerra, l'amo dels més grans magatzems de Nova York, en Macy, va convocar un congrés amb científics de primera fila, psicòlegs, lògics, filòsofs, biòlegs, etc. per esbrinar i definir què era la intel·ligència i si es podia reproduir. Entre els participants hi havia el biòleg McCulloch i el lògic Pitts, que van escriure una ponència, després publicada i famosa, que es deia «Sobre un càlcul immanent en la lògica de les idees» (i observeu aquí per enèsima vegada l'associació de «idees», «pensar» i «càlcul») en què s'intentava convertir la «màquina de Turing», la qual es prenia com a descripció estàndard indiscutible de «què es pensar», en alguna cosa menys formal i matemàtica i més realista (per a un biòleg), és a dir, més semblant al cervell, ja que en darrera instància és aquest qui «raona» dintre nostre i que ens resulta més familiar i natural que no pas un esquema abstracte d'instruccions lògiques executades mecànicament. Van convertir el mecanisme de Turing en un conjunt d'unitats (que per raons òbvies van anomenar metafòricament –i abusivament– «neurones») interconnectades entre elles formant una «xarxa» (dita així perquè en recordava una) que, vist en conjunt semblava realment el model (simplificat, a escala) del *cervell*. Tot i que la descripció era formal, semblava clar que, un cop engegat, el mecanisme «cerebral» descrit es comportaria com un autèntic cervell i, per tant, podria fer tot el que nosaltres (i la màquina de Turing) érem capaços de fer. En va quedar tothom prou content, i van deixar el problema per l'endemà de la guerra. I llavors és quan tornem a trobar una sèrie de gent –sobretot metges, biòlegs i algun matemàtic despistat i escadusser– preocupant-se una altre cop per saber «què és la intel·ligència». Els trobem reunits en un congrés l'any 1948, una altra vegada organitzat per un filantrop ric, en què l'intel·ligentíssim, culte, simpàtic, important, totpoderós i omnipresent matemàtic i lògic von Neumann (que ja hem presentat) va fer una dissertació profunda sobre el tema que encara avui ens deixa meravellats i no acabem d'entendre (i de la qual ha sorgit mitja informàtica de la que tenim, des dels *autòmats finits* fins als *autòmats cel·lulars*, als *autòmats autoreproductors* –a mig

camí de la vida– i a una primera estimació del cost físic del *càlcul* com a disminució de l'entropia o desordre).

### *El cop d'estat*

Del climàctic 1948 fins a l'anticlimàctic 1956 –any prou important, com veurem– l'activitat conjunta de lògics i biòlegs entusiasmats no s'atura, en una activitat popularment coneguda amb el neologisme (de Norbert Wiener) *Cibernètica*. Semblava que tot d'una tindriem alhora màquines intel·ligents, pròtesis per a amnèsics o lesionats cerebrals, robots parlants i treballadors industrials que deixarien els obrers humans lliures per fer el que volguessin, amb el plus afegit de la comprensió –finalment– de la pregunta eterna «què és la intel·ligència». Doncs no: tot l'anterior se'n va anar en orris quan una colla de matemàtics, amb una antipatia declarada –i mai gens dissimulada– per tot el que fossin cervells reals i matèria viva així com per metges i biòlegs o gent que hi parlés (com el seu col·lega Wiener, que s'hi relacionava sovint), van declarar la guerra a l'enfocament diguem-ne «cibernètic» i a qualsevol altre que no fos estrictament formal i matemàtic com el de Turing (actitud amb la qual segurament Turing hauria estat profundament en desacord... però Turing ja s'havia mort, aparentment suïcidat). Ho van fer l'estiu del 1956, a partir del convenciment que saber què era la intel·ligència i reproduir-la no tenia res a veure amb conèixer les propietats o el funcionament del nostre cervell. Al contrari, la intel·ligència –la del futur, la que podrem un dia tenir directament a l'abast– ja la teníem davant nostre: era la màquina del dia: l'*ordinador*. (Potser el fet no era del tot casual, ja que aquell any IBM havia estat regalant ordinadors a les universitats americanes més importants.) Vist així, la intel·ligència consistia essencialment a agafar un ordinador i programar-lo, amb programes que fessin coses «intel·ligents», i després estudiar què havien fet i com.

Això, que en principi pot ser un bon enfocament per atansar-se a la intel·ligència i entendre-la, va ser clarament, tal com es va produir, un *cop d'estat*, les conseqüències nefastes del qual encara patim. Ara sembla que inventem coses noves, però no és ben bé així. Per exemple, les xarxes neuronals no s'inventen pas l'any 1980: ja estaven inventades (surten a l'article del 1943). El que fem

ara és refer-nos del cop d'estat del 56 i d'una dictadura que ha durat més de 30 anys i que ha consistit a dir que «la intel·ligència són els ordinadors» i, per tant, qualsevol cosa intel·ligent és factible en un ordinador. I si no, no es tracta d'intel·ligència. És a dir, «intel·ligent» és tot allò que és equiparable al producte de l'activitat d'un ordinador. En particular, *el cervell humà és un ordinador* (una desafortunada i sorprenent inversió del prudent i gens dogmàtic enfocament inicial dels sensats von Neumann i companys). Afortunadament aquesta dictadura, que ha fet tornar bojós entre altres els lingüistes i els psicòlegs –que totalment bocabadats intentaven trobar «el programa» de cada cosa– s'esta trencant, i per molts llocs diferents. Dic que aquesta dictadura va ser nefasta perquè va aturar molts projectes de recerca interessants senzillament perquè molts científics van quedar enlluernats davant del «paradigma informàtic» del cervell, o de la psicologia, de la lingüística, de tot. I no sols va ser una dictadura intel·lectual: molts que treballaven en les línies originals van veure com se'ls assecaven els fons de finançament i se'ls rebutjaven articles en revistes científiques simplement perquè no corresponien a la nova idea hegemònica. Així, Michael Arbib –per citar un biòleg, per cert enamorat de Catalunya, que el 1960 ja tenia prou prestigi– confessava que va haver de reconvertir-se, de fer articles –i quasi creure-se'ls– sobre tot això que la intel·ligència i qualsevol cosa intel·ligent és *com si fos* un ordinador, i que el cervell hi té tant a veure, amb la qüestió de la intel·ligència, com hi pugui tenir el pàncrees. (Aquestes misèries no demostren altra cosa que la Ciència és una activitat humana com qualsevol altra, i que els científics poden ser tan mesquins com tota l'altra gent que passa pel carrer.)

Si la dictadura s'ha trencat per fi ha estat, d'una banda, perquè hem vist que en certs problemes de gran envergadura –com ara interpretar escenes complexes o frases del llenguatge– si continuàvem prenent que resoldríem la qüestió a base d'ordinadors era clar que xocàvem contra una paret: senzillament, no tenia sentit haver de fer un ordinador superpotent per fer unes coses tan senzilles com algunes que fan els nens d'un any (reconèixer cares o veus, o gestos o matisos de la parla). Clarament, era inadmissible que per fer coses que fa sense dificultat un infant, que té neurones encara per connectar, tendres i amb poca pràctica de funcionament,

haguéssim de construir megaordinadors Cray o similars només a l'abast dels governs. Era insensat i absurd –i humiliant– que un nadó derrotés un pressupost de recerca cada cop més descontrolat.

Quan hem aconseguit sortir d'aquesta closca asfixiant del «paradigma informàtic», moltes idees s'han anat recuperant després d'haver estat tapades o proscrites –o d'haver estat senzillament «impensables». Idees que en condicions normals segurament haurien anat progressant de manera natural i ens haurien aportat intuïcions útils o alternatives. Una és la de les xarxes neuronals, tan antigues i sortosament recuperades, però n'hi ha d'altres: el càlcul analògic, la programació evolutiva o genètica, etc. Com a informàtic em sento corresponsable, per la part que em toca, d'aquesta limitació efectiva de la llibertat científica gràcies a la qual, desgraciadament, tots nosaltres –els informàtics els primers– n'hem sortit perjudicats (perquè hem deixat de tenir idees, obrir camins, sentir-nos sense traves per imaginar models alternatius o heterodoxos, etc., condició bàsica per al progrés de la Ciència).

Hi ha força coses a dir sobre la *intelligència*. Realment, la intel·ligència és moltes coses alhora. Quan abordem un problema que no sabem com enfocar ho fem de la manera més elemental possible: de la mateixa manera en què hem pogut resoldre, més o menys, altres coses que si fa no fa ja ens havien neguitejat abans. I això es el que fan normalment els matemàtics, que són com ja he dit els primers que van posar mans a l'obra el 1956, l'any del cop d'estat. La primera aproximació, que –deixant antipaties personals a part– no deixa de ser raonable per a un matemàtic de professió, va ser fer-se aquesta pregunta: «en què consisteix la intel·ligència en el seu estat més pur?» I contestar-se-la així: «Doncs consisteix en *allò que fem els matemàtics*». Aquesta definició operativa –i corporativa– no sols permetia desencallar la qüestió sinó que en suposava una altra de ja contestada –afirmativament–: a tota activitat d'un matemàtic se li ha de suposar intel·ligència (!?). Aquesta resposta no es cap caricatura dels matemàtics sinó que són paraules textuais del matemàtic McCarthy aquell famós estiu del 56, que ho formulava dient que la intel·ligència és el que nosaltres considerarem intel·ligent i farem fer als nostres ordinadors, i el que ells facin serà per definició intel·ligent. Això, que és històric i que jo, que sóc del gremi, transcriu d'aquesta forma potser un pèl ridícula, és

–trobo, personalment– una bona manera, perfectament racional i sensata de començar a avançar en terra verge (un cop atacat i descartat l'enfocament biològic i «cerebral» en què creien els il·lustres predecessors als quals es volia superar o suplantar). Una vegada fet aquest pas, només calia plantejar el procés de pensar intel·ligentment, tal com un matemàtic fa teoremes. I això requereix dues coses: coneixements –coses que es poden expressar de manera concreta i si pot ser simbòlica– i Lògica, que és el que permet manipular rigorosament aquests símbols en què hem codificat el coneixement. Així doncs: Intel·ligència = Símbols + Lògica.

Doncs bé, la gran sorpresa és que aquesta simplificació iniciàtica, proposada el 1956, va tenir un gran èxit el 1976, després de 20 anys d'anar-la practicant en tots els camps i des de tots els punts de vista possibles. Al 76 es feien programes que sorprenien tothom, i molt. Abans d'explicar què vol dir això de «tenir èxit» en el camp de la «Intel·ligència Artificial» (el nom que els colpistes del 56 van donar a la cosa) s'ha d'advertir una cosa. *Intel·ligència Artificial* no vol dir gaire res: vol dir, essencialment, capacitat de crear *sorpresa* en l'espectador humà que constata que una màquina és capaç de fer una cosa que ell mai s'havia esperat de veure fer a una màquina. Amb sentit humorístic diríem que la intel·ligència exhibida per una màquina és proporcional a la quantitat de «*ohs*» que emeten els espectadors –la qual cosa no certifica que la màquina sigui intel·ligent, només que a la gent li ho sembla– perquè se suposa que les màquines no pensen ni són intel·ligents, gairebé per definició, ni actuen com nosaltres i, per tant, quan alguna màquina fa res semblant a nosaltres o que ens ho recorda o fa res atípic (respecte del que s'espera d'una màquina) ningú no pot evitar –amb plaer o amb espant, això ja dependrà de cada un– de considerar la màquina, valgui la paradoxa, «intel·ligent», d'atribuir-li allò que en diem *intel·ligència*. (Tot i que, si calgués descriure pròpiament la *intel·ligència* i caracteritzar-la en poques paraules, potser l'hauríem de basar essencialment en la capacitat d'*adaptació* de l'individu, més que no pas en la de preveure i calcular).

En resum, en aquesta primera etapa de la «Intel·ligència Artificial» els matemàtics, que eren els qui havien agafat la paella pel mànec, havien definit «intel·ligència» com a sinònim de «activitat de matemàtic». I si avui els matemàtics treballen normalment amb

ordinadors, doncs –el corollari és claríssim– programem tant com puguem, fem fer la feina pesada als ordinadors i els resultats, si són prou «intelligents» (o si més no «sorprenents»), donaran lloc a –seran, de fet– «la intel·ligència». Alguns matemàtics, però (i penso amb McCarthy especialment), ja van veure de seguida que això que els ordinadors siguin imitacions de matemàtics no és veritat, i que el que hem dit que fan el matemàtics només ho fan quan *escriuen* els seus resultats (els seus «teoremes»), no pas quan els conceben i elaboren. En els moments crucials de *crear* teoremes el que fan servir és la intuïció, el sentit comú. I McCarthy va anticipar que per arribar realment a la intel·ligència havíem de programar en les màquines també aquest «sentit comú». (Com? McCarthy reconeix que no en tenia ni idea, però que era per aquí que calia enfilar.) Això no va ser així de seguida, ni és així ara.

### *El triomf de l'infant David contra el Cray Goliat*

No era per aquí que jo volia anar, però (la qüestió del sentit comú encara ara la tenim oberta). Vull tornar al que deia de l'any 76, en què alguns ordinadors, fins i tot sense ni mica de sentit comú, eren capaços de sorprendre tothom, i molt, fent coses que ningú s'hauria esperat d'ells. La idea havia estat en la majoria de casos recopilar els coneixements d'un expert en una matèria donada –química, medicina, geologia o informàtica–, veure i sistematitzar quines regles aplicaven quan es trobaven amb un cas concret, i programar un ordinador de manera que es comportés de la manera més aproximada a com ho faria el personatge en qüestió. D'aquests programes se'n va dir «sistemes experts», se'n va fer una gran propaganda i se'n van vendre –i se'n venen– molts. Només calia triar un tema, interrogar fins a l'avorriment un bon expert (humà) en el tema, treure'n tota la informació possible, explicitar-la, codificar-la i tot seguit programar un ordinador prou ràpid perquè fes la cosa, i exhibir-lo. Exemples: un ordinador es dedica –amb èxit del tot acceptable– al diagnòstic mèdic (a intentar suplantar els metges, diguem-ho clarament, tot i que els metges que en quedaven bocabadats no s'adonaven d'aquest petit detall que els amenaçava potencialment el propi futur). Un altre exemple: un ordinador es dedicava a fer psicoteràpia, amb una interacció amb

el pacient que aparentment suplantava el psicoanalista i resultava prou plausible (tot i que el creador del programa, Weizenbaum, es feia creus de com un programa tan summament simple com el que havia creat podia despertar en els espectadors tal crèdula sensació de realisme). Fins i tot se'n va fer un –apropiadament batejat Parry– que actuava com un humà autènticament paranoic.

Però tot això, bromes apart, representava per als informàtics un punt d'impàs. Podien imitar experts, però no podien ni somiar a fer, ni tan sols disposant de megamàquines i megapressupostos, coses elementals com algunes que fan els nens d'un any d'edat: veure-hi, distingir-hi, entendre, parlar (intelligentment, no merament sintetitzant paraules), adonar-se d'una situació i adaptar-s'hi, etc. És en aquest moment de crisi, cap al 1980, en què per intentar fer coses que fa un infant cada vegada s'hi han de posar més màquines i més grans –i més increïbles–, quan (re)apareixen les xarxes neuronals de què parlàvem al començament. Però és que el fenomen de les «xarxes neuronals» no és només el redescobriment d'una heretgia reprimida durant anys pel triomfalisme «informàtic» (=fer més i més programes, i així tindrem –«sorgirà»– la intel·ligència) sinó que les xarxes neuronals també tenen unes característiques que les fan diferents dels programes d'ordinadors habituals i que, un cop vistes, no deixen de fer-nos pensar que d'informàtica potser no n'hi ha una sola, o que la informàtica no té per què ser sempre allò que hem pensat que havia de ser. Per exemple, les xarxes neuronals no s'han de programar: només s'han de configurar, i a partir d'aquí deixar-les funcionar i mirar que fan –i resulta que el que fan es aprendre, reconèixer, abstroure, etc.– i tot sense que ningú els digui o descrigui –o tingui la més lleugera idea– sobre el que han de fer, o com. En resum, no els cal «programa»: s'autodefineixen soles en el seu propi funcionament; per exemple, poden reconèixer cares després d'haver-ne vist algunes mostres, o reconèixer situacions –que cap programador no sap ni pot anticipar– després de viure'n unes quantes. Una altra característica de les xarxes neuronals: la memòria. Bé, resulta que les xarxes neuronals no en tenen, de memòria. En un ordinador, tu dius: «em cal la dada  $x$ , que tinc a la posició de memòria  $y$ ». Doncs bé, les xarxes neuronals tenen una «memòria» no localitzada, «distribuïda» pertot arreu (de fet és tota la xarxa la que «recorda») i accessible no per posició o adreça



(saber on és la cosa que es busca) sinó per associació («no sé ben bé què busco, però s'assembla o recorda vagament z»).

Més característiques inherents a les xarxes neuronals: no cal augmentar la potència de càlcul a base de construir «xarxes», és a dir, fer grans juxtaposicions intercomunicades d'ordinadors més o menys potents perquè s'ajudin els uns als altres en els seus càlculs i se'n comuniquin els resultats. No: la xarxa neuronal ja està feta, ella mateixa és un mecanisme cooperatiu que no cal construir, i quan treballa (quan fa els seus «càlculs») la cooperació entre les seves unitats (abusivament anomenades, per raons històriques, «neurones») es fa de manera natural i espontània, que raonablement no hi ha cap necessitat de programar. Encara una altra característica de les xarxes neuronals, aquesta potser una mica més sorprenent: els ordinadors són màquines totalment intolerants a l'error. Tu t'equivoques en un «bit» i la màquina s'atura esperant que tu t'adonis de com ets de ruc i maldestre i facis el que puguis per refer-te'n, refer-ho i donar-li-ho ben donat. Doncs bé, les xarxes neuronals no: si una unitat («neurona», n'hem dit) falla, la xarxa no et deixa penjat ni és la fi del món. És més, poden arribar a fallar més de la quarta parta de les «neurones» d'una xarxa sense que els sistema es col·lapsi. Funcionarà segurament malament, però funcionarà –a empentes i redolons si cal i amb resultats aproximats–: quina diferència amb els irritants ordinadors i la seva insistència a rebutjar-nos mitjançant missatges semidesxifrables com ara *syntax error*, *system shutdown*, *unattendable request* i similars crides a deixar la nostra pudorosa tendència a l'eufemisme i les bones maneres (ens ve a dir, críticament, que allò que passa no és el seu problema –de fet res no ho és–, que som uns ases, que ens espavilem, que li ho posem tot ben posat i que, si en som capaços, tornem a passar-lo a veure, que aquesta vegada potser ens atindrà amb la seva infinita paciència; el que no ens prohibeix, per fortuna, és descarregar el nostre pap amb renecs: per això deia allò de la suspensió momentània dels eufemismes).

Per cert, en aquest favor que ens fan les xarxes neuronals de no deixar-nos penjats de seguida, fenomen tècnicament anomenat *degradació gràcil*, és fàcil de veure-hi, com així ho van fer de seguida molts, un gran paral·lelisme entre aquestes xarxes «neuronals» i el nostre cervell, que també presenta degradació gràcil (en el nos-

tre cas anomenat «vellesa» o simple deteriorament mental): per exemple, tant en malalts d'Alzheimer com en xarxes neuronals el llinar de «neurones no operatives» que cal típicament perquè el sistema deixi de funcionar és del 25-30% (quan s'arriba a aquest percentatge de neurones «mortes» el cervell o xarxa deixa de «calcular» correctament, ja ni per aproximació). Són sorpreses com aquestes que van fer pensar molts biòlegs, una mica prematurament, que estàvem a punt d'entendre el cervell, i que ho estàvem aconseguint sense haver-lo d'obrir, només construint «xarxes neuronals», abstractes i indolores. Però no: no havíem descobert res sobre el cervell; simplement havíem construït un mecanisme, tan formal i abstracte com l'ordinador, que tenia certes característiques a partir de les quals el més sensat era no extrapolar res (i menys cap al cervell real o cap la intel·ligència): fer-ho –temptació difícil de resistir, però– hauria estat repetir una bestiesa com l'anterior de configurar el nostre cap com si fos un ordinador. I quan se surt d'un dictadura el pitjor que es pot fer és entrar en una altra.

Tot això que aporten les xarxes neuronals, sense ser extraordinari –ni ser les xarxes neuronals els únics «nous» enfocaments que aporten sorpreses– és bo perquè trenca el monolitisme imposat per trenta anys d'insistència monocorde, aquesta dèria pseudoinformàtica consistent a dir i repetir que «l'ordinador és el model», i que la intel·ligència resulta del que fan els programes –més ben dit, del que fan els programadors–. (I viceversa: que la intel·ligència només pot sorgir de mecanismes semblants a un ordinador, que és el parany en què molts lingüistes i psicòlegs han caigut tots aquests anys de manera nefasta, com hem dit, i que amb tota seguretat els ha retallat altres camins molt més productius d'enfocar i explicar-se els seus propis i respectius problemes).

### *Un 10% que ho omple tot, com els gasos*

Tornem a la intel·ligència. Si som capaços de bandejar definitivament la potent –però moderadament útil i en darrer terme destructiva– metàfora «intel·ligència = ordinador» i no caiem en nous enlluernaments possibles com ara «cervell = xarxa neuronal», potser podrem entendre, a poc a poc i amb paciència, què és la intel·ligència. Perquè la «intel·ligència», com ja he dit, és moltes

coses. Començo per una, que il·lustro amb una observació àmpliament documentada i coneguda, recentment recordada per un neuròleg portuguès establert a Minnesota anomenat António Damàσιο en un llibre de divulgació molt interessant (titulat «L'error de Descartes», que recomano a tothom). La tesi de Damàσιο és que el cervell és essencialment un òrgan regulador, consagrat a supervisar i revisar constantment funcions «corporals» (del cos al qual va unit, és clar) tals com el control del sistema nerviós, els fluxos hormonal, etc. I diu que un petit percentatge del seu temps, no pas gaire més enllà del 10%, el cervell el dedica a omplir-lo amb unes activitats «buides», no directament funcionals ni productives, que en diem «pensar». Ara bé, aquesta operació de pensar, que sol ser feta en termes molt sintètics –«simbòlics», diríem– no és realment buida; és auxiliar i anticipatòria, i està sempre lligada a qüestions funcionals o al seu possible desenvolupament, optimització i adaptació. Doncs bé, és aquest <10% de capacitat cerebral tan especial –i fàcilment reduïble a símbols i manipulacions formals– el que hem magnificat, el que hem convertit en el personatge d'un matemàtic –esdevingut símbol suprem del famós paradigma de la «intelligència artificial»– i el que hem assimilat a un ordinador. (Per cert, per casualitat, perquè l'ordinador és un invent del 1946, i si haguéssim hagut de buscar-hi una metàfora preinformàtica a això de pensar potser hauríem sentenciat que el cervell és una màquina de comptabilitat, com no gaire abans, als anys 1920, un famós savi l'havia assimilat a una centraleta automàtica de telèfons, i si fos ara diríem sense gaire vergonya que el cervell és com l'Internet.). Doncs bé, hem convertit aquest espai residual del cervell en sinònim d'intelligència, l'hem intentat crear o reproduir a base de programes d'ordinador i hem arribat a pensar que no hi pot haver cap altra mena d'intelligència fora d'aquesta. Ara bé: és això la intelligència? No hi res més en la intelligència fora d'això? (Dues altres preguntes foren: és mínimament seriós, tot això? i també: és mínimament científic basar la nostra recerca en un joc de metàfores?)

Doncs potser sí que hi ha alguna cosa més en la intelligència fora de la inefable capacitat de calcular i de la irrefrenable tendència a fer-ho fins i tot quan res no ens hi portaria ni el context hi és mínimament l'adequat. Per exemple: hi ha l'*emoció* (tothom sap

què és, no cal pas definir-ho): on queda, això? I hi ha la *consciència*. Atenció, que no estic pas fent sentimentalisme ni teologia. Aquestes dues coses són atributs bàsics de *tota* intel·ligència. (Damàsió ens explica magistralment que no sols en són el testimoni sinó el *motor* bàsic, imprescindible: sense emocions com a motivació –motor, literalment– no hi ha intel·ligència.) I si totes dues funcions hi són és perquè són innovacions funcionalment adaptatives i per això l'evolució n'ha afavorit l'aparició i el desenvolupament. Dit científicament amb una paràbola, l'*emoció* és l'equivalent d'un servei de seguretat que fa diverses rondes durant la vigilància d'un edifici: 40 vegades per segon el nostre cervell fa una síntesi de les informacions que li arriben via els seus diferents «monitors» (els sentits). Que no hi ha res que cridi l'atenció? doncs tranquils i a continuar amb el que fèiem. Que hi ha algun canvi o cosa digna d'atenció? doncs alerta. L'alerta pot donar lloc a un de tres tipus de processos: *anar-hi* (aproximació, atac, desig, etc), *separar-se'n* (allunyament, fugida, por/temor/terror, etc.) o *no moure's* (interpretable segons els casos com a precaució, curiositat, terror paralitzant o el que sigui). Tot això, que sol anar acompanyat de «sensacions» internes molt viscudes, reactivables en qualsevol moment («recordables») i complexes, són les *emocions*. Avui el seu estudi cau plenament dintre el camp de la ciència (i de la Intel·ligència Artificial, vegeu l'assaig sobre *ordinadors emotius*). La seva utilitat adaptativa és evident, i la seva complexitat en els humans sorprenent (el resultat d'un conglomerat d'estímuls emotius pot ser tan diferent segons les persones, els moments, l'educació, la cultura, etc., que per això es van inventar un dia els psicòlegs).

I la *consciència*? Espero que ningú em negarà la utilitat adaptativa de reconèixer-me jo mateix com si fos una mateixa cosa, un «individu», al llarg del temps. Un individu és per definició un ens viu que es reconeix com a tal, que no s'agredeix sinó que més aviat mira de preservar-se el més sencer que pot, una entitat amb vida pròpia (mai més ben dit) que amb el temps es transforma i alhora manté uns invariants (físics i conductuals). En resum, una cosa autodenominada «jo mateix» a qui passen les més variades i inesperades coses sense per això deixar de ser el mateix «jo mateix» de partida i sense perdre en cap moment la capacitat de distingir-se (tot ell, i en particular el propi cos) de «l'altre», encara

que aquest altre sigui la mare o el veí. En el cas humà, la seu física d'aquestes habilitats (veure-hi, sentir i recordar) és l'hipocamp, un bocí petit i central del cervell que centralitza i unifica aquestes coses en forma de «memòria autobiogràfica», una mena de nucli o síntesi del «jo mateix» en què hem dit que consisteix resumidament aquest intangible fins ara fora de l'abast dels científics anomenat «consciència».

I ara em faig aquesta estúpida i aparentment redundant pregunta: per què necessitem fer màquines intel·ligents? (O dit d'una altra manera: no en tenim ja prou amb nosaltres?)

M'explico: suposem que volem traslladar-nos d'un lloc a un altre i amb les cames no en tenim prou. Doncs bé, inventem el cotxe. Però oi que el cotxe no ens suplantarà a nosaltres ni les nostres cames? Només l'hem inventat perquè ens ajudi a desplaçar-nos. Doncs ho fa, ens hi desplaçem i tots contents. He esmentat aquest cas tan trivial per aclarir que una cosa semblant sembla que hauria de passar també amb les màquines intel·ligents (però a l'article *Ordinadors emotius* plantejo els meus dubtes). Necessitem fer ràpid i fer bé això que ell nostre cervell només fa durant menys del seu 10% superflu? (Exemples: calcular interminablement les xifres decimals del perfil d'un biga de pont, recordar nimietats com on carai vaig assignar el seient del client  $x$  a l'avió de Bangkok de les 4, com es feia per saber si una molècula orgànica conté ferro o no, quins símptomes anaven associats a l'hepatitis B segons el toix de manual que em van fer estudiar a tercer de carrera, etc.) Doncs inventem un ordinador, i li fem fer totes aquestes coses. No us amoïneu, que no es queixarà pas ni se'ns rebel·larà, tot i que li fem fer milions de passades pels mateixos càlculs. Ara. Vol dir això que l'ordinadors ens ha «suplantat»? (Doncs potser sí, i, si fos així, millor per nosaltres: menys feina, o menys estúpida!) Només podríem sentir-nos suplantats si creguéssim que (a) nosaltres només servim per a això (i no! encara que segons quines feines que ens encarreguen ho podrien fer creure), i (b) que la intel·ligència només és això que fan els ordinadors: calcular, raonar en termes simbòlics sobre temes prèviament estructurats i pre-programables, amb una limitada o nul·la capacitat d'aprenentatge, i sense possibilitat d'error (i, si en produís cap, amb una limitadíssima capacitat d'autocorrecció). Perquè, a propòsit d'això últim, si no permetéssim a un ens pre-

sumptament intel·ligent cometre errors i mirar –pacièntment, si calde corregir-los, com carai aprendria res, aquest ens? Turing deia que sense error no hi pot haver res que es pugui qualificar pròpiament d'«intel·ligència». Potser tindriem calculadors perfectes, reservadors de vols sense *overbookings*, però darrere la màquina només hi podríem veure un *idiot savant* (personatge exemples del qual són els popularitzats per nombrosos escriptors, casos clínics o actors de films, com ara Dustin Hoffman a *Rain Man*). Dels molts *idiots savants* que hi ha hagut a la història, alguns detalladament documentats, mai ningú n'ha pogut considerar intel·ligent cap ni en un remot sentit del terme. Caldria veure un ordinador com el que és, un *idiot savant* tancat en un armari, desconnectat d'una realitat de la qual només sap l'existència i característiques pel que nosaltres li'n diem. I si un ordinador és això, també és un zombi, capaç d'anar incansablement repetint una mateixa «cosa». Una «cosa» que no té ni pot tenir en rigor cap sentit per a ell (precisament, entre altres raons, per la seva falta de connexió amb un cos i una realitat pròpiament viscuda). I ningú, que se sàpiga, s'ha pogut identificar mai humanament amb un zombi, per molt reservador de vols competent que pugui ser i per moltes frases amables que els seu programador li hagi incorporat per fer-nos contents. (Una altra cosa és tenir por que aquests zombis un dia ens suplantin de cop a tots plegats i ens facin viure sota terra, però això és l'argument de *Terminator* i com a mínim ens hi caldria en Schwarzenegger per fer-ho mitjanament creïble).

*Del significat de les màquines (i de com incomprendiblement maldem per «identificar-nos-hi» (!?))*

Voldria afegir tres últimes coses, i acabo:

1. Avui estem fent ordinadors –ordinadors, no pas xarxes neuronals– que són efectivament «reconeixedors d'estats d'ànim» (dels nostres, no pas dels seus). Tal com vaig explicar en un número anterior d'aquest Butlletí, a Stanford un grup de becaris (mal pagats, que són els que fan millor la feina) han construït un sistema molt senzill (una càmera acoblada a un sistema que reconeix certes característiques visibles en les cares dels usuaris

que delaten el seu estat d'ànim), de manera que si nota que l'usuari-interlocutor fa mala cara vol dir que aquest no ha entès bé una frase de l'ordinador, o que s'ha pres malament una resposta, o que hi ha alguna cosa que el preocupa, etc., i llavors l'ordinador de seguida s'afanya a afegir frases florides o amables del tipus «oh! quin greu», «potser no l'he entès bé», «vol que li ho detalli o justifiqui?». Potser amb això introduïm cortesia en el tracte, però no ens confonguem: no hem «humanitzat» l'ordinador... almenys no pas més del que «humanitzem» un venedor d'enciclopèdies quan li donem la consigna que no deixi mai d'ensenyar les dents blanques, o no pas més que l'hostessa d'un aeroport impossible s'«humanitza» fent-nos un somriure postís quan veu que, si no es torna simpàtica de cop, es pot trobar amb un motí de passatgers inaplacables.

2. Darrerament un equip ha localitzat (en rates) quines neurones cerebrals tenien capacitat motriu directa sobre membres del cos, per –a partir d'aquí– desenvolupar sistemes per moure braços o cames directament sense gaires mecanismes intermedis. Això està molt bé, i va en la direcció que deia: ajudar-nos de la intel·ligència mecànica per suplantar funcions que nosaltres hem perdut o tenim problemes a desenvolupar.
3. Hi haurà un dia que, finalment, els ordinadors acabaran parlant amb nosaltres directament i correctament. Ara bé, és fàcil predir que tindrem amb aquestes màquines les mateixes sensacions que tenim amb persones (humanes) amb les quals parlem correctament en la nostra llengua però amb les quals compartim poc per no dir gens, amb les quals no ens podem «sentir identificats» humanament de cap manera i amb les quals, per molt que parlem de referents aparentment comuns, «sentim» que no hi tenim res a veure ni hi compartim res.

Aquest tercer punt em duu al punt final del que volia dir, i que lliga amb el que deia al començament. La pregunta clau és: És possible «identificar-nos» amb una «màquina»? (Atenció a les dues paraules entre cometes.) Això té tres graus:

- a. Quan estem davant de persones (encara que siguin falses; per exemple, personatges de ficció d'una pel·lícula: Darth Vader

o Indiana Jones, posem per cas) nosaltres esperem de l'altre, suposat persona i suposat intel·ligent, un grau d'identificació. Ens sentim davant d'un amic, conegut, col·lega o simplement persona o semblant, podem vagament predir o anticipar quins comportaments i sentiments podrà arribar a tenir (per analogia amb els nostres), i som conscients de la relació i de les coses que compartim. Això ens permet fer bromes, usar la ironia, recordar fets passats o referenciables, etc., si el grau de proximitat és relativament fort. En canvi, si ens acaben de presentar, hi ha moltes coses que hauréu d'explicar o donar arriscadament per sabudes, esperant que no hi hagi malentesos o la cosa acabi malament. En el cas extrem que la persona amb la qual parlem (suposem que en la mateixa llengua i sense ambigüitats sintàctiques) és un extraterrestre o quasi, la nostra sensació d'estranyament (*alienació* se'n diu d'això en llatí) és aproximadament la mateixa que experimentaríem amb una màquina a la qual no se li hagués fet fer l'esforç de documentar-se mínimament sobre nosaltres perquè nosaltres hi puguem practicar un mínim d'identificació. Això no ha de ser tan difícil com sembla. Si les companyies venedores d'enciclopèdies (o d'andròmines diverses) han aconseguit fer-ho fins a «humanitzar» la persona-màquina de vendre que se'ns presenta a casa per la porta, per què no ho ha de ser amb màquines-màquines de debò (encara que cap d'aquestes pugui fer-se càrrec del que representa arribar a casa i veure la família o la TV després d'un dia esgotador ni s'ho pugui remotament «imaginar»)?

- b. En casos en què volem o necessitem «identificar-nos» plenament amb algú sense el recurs del llenguatge humà articulat, tampoc no hi ha problema. Amb això hi tenim una llarga experiència. D'aquestes màquines que no parlen i amb les quals ens identifiquem en diem «gat» o «gos», segons el cas: són fàcils de fabricar, naturals i ecològiques (i pertanyen a l'elàstica categoria d'«animals domèstics»). Que arribem a tenir una versió artificial d'aquestes màquines que fan coses diguem-ne «intelligents», ens escolten i ens obeeixen i amb les quals efectivament ens identifiquem és no sols fàcil d'imaginar sinó que ja ha passat. Gràcies als nostres empedreïts inventors i estranys coterraquius



anomenats genèricament «japonesos» ja en tenim un tast. I fins i tot sabem com pateixen els nostres fills quan se'ls mor una mena de rellotge de butxaca per falta d'atenció, o quant estem disposats a pagar per un «gos» mecànic de forma aproximadament paral·lepipèdica que és capaç de fer la meitat de les coses que fa un gos normal (i observeu que cap dels compradors d'aquest estrany «animal» sembla haver considerat que els gossos de veritat són força més econòmics i abundants, per cert).

- c. Fins aquí, doncs, la resposta a la pregunta «podrem arribar-nos a identificar amb éssers intel·ligents artificials?», que ve a ser equivalent a la qüestió «podrem arribar a fer éssers intel·ligents artificials i entendre'ns-hi?», sembla positiva. La capacitat que tenim els humans d'identificar-nos amb coses, encara que siguin iguanes antropòfagues o ossets de peluix sembla il·limitada. En canvi, hi ha un cas de màquines intel·ligents amb les quals **no** ens podem pròpiament «identificar» **mai**. Són aquelles a qui hem encarregat una tasca en la qual no tolerem el més mínim error. No ens podem identificar mai amb una *màquina* que hem construït perquè ens ajudi a fer coses complexes ràpidament, eficientment i explícitament (perquè en depenem molt i un error pot ser fatal i inexcusable) *sense errors*. (Per cert, les *màquines* per definició bàsicament han estat sempre això: estris que fan coses *sense errors*, i precisament per fer aquests estris existeixen els *enginyers*). Doncs bé, suposem que de sobte una màquina d'aquestes s'equivoca i en comet un, d'error. Podrem usar-la i confiar-hi, però el que no podem fer mai és identificar-nos-hi. Aquests aparells, que ens havien d'ajudar a fer *x* i perquè ho fessin els havíem donat tota la nostra confiança, equivocant-se cometen l'equivalent d'una *traïció*. Mai permetrem que un cotxe intel·ligent (d'aquests que ens prometen que anirà sol) ens atropelli o ens estimbi o que un sistema de vol ens mati, ni tan sols que ens perdi un enllaç o una maleta, per molt intel·ligent que sigui (o justament en la mesura que nosaltres els hem atribuït la condició d'«intel·ligents»). Els enginyers, que són els que fan «màquines», tenen prohibit (els despatxen, vaja si ho fan) fer-les i que fallin; no és concebible que un enginyer construeixi un cotxe «intel·ligent» que de tant en tant pugui atropellar una

persona per així aprendre que no s'han d'atropellar persones, o que faci imprudències, temeritats o simples maldestreses per aprendre com esquivar obstacles, etc.

En això últim, no cal dir-ho, les màquines intel·ligents no s'assemblen pas gens a nosaltres. (De fet, cal?) Nosaltres continuarem equivocant-nos, tirant gerros a terra o fent coses mal fetes. És això el que fem quan som infants –que és quan ens programem socialment (encara que dit així soni malament)– i és així com l'encertem quan ho fem: perquè prèviament ens hem equivocat i sabem el gust del fracàs, i fins i tot el sabor amargant de les pèrdues irrecuperables. Aprenem equivocant-nos: per això som intel·ligents, com deia Turing.

I també som diferents, els humans, perquè tenim *motius*. De vegades això no s'entén, i per això hi ha els novel·listes. Quantes coses hem fet sense voler-les realment, simplement perquè volíem impressionar, o venjar-nos, o oblidar! Hi ha el cas d'un escriptor anglès victorià que no va parar d'escriure novel·les fins que va morir sa mare; de cop, va deixar d'escriure: era clar que escrivia perquè així li deia o retreia coses –o a ell així li semblava– que no gosava dir-li directament. En aquest cas la vàlua literària de l'autor no era gran cosa i sense l'episodi personal d'aquest senyor potser la literatura no hauria perdut cap gran obra. Però... vés a saber: potser tot l'opus shakesperià era un diàleg de l'autor amb son pare (per un malentès que hi va tenir un dia de petit o per la ràbia que li va fer qui sap què). Precisament, això que ens caracteritza en exclusiva és el que anomenem *creativitat*. Es tracta de la propietat (humana) més boja, imprevisible, incontrolable i irracional (o a-racional). Encara que, vist a posteriori, un s'ho pugui explicar tot. Per exemple, Einstein visualitzava molt bé les coses de petit perquè no se'n sortia de cap altra manera: però qui podia preveure que d'aquesta frustració en sortiria la Relativitat? En català hi ha un refrany que ho explica: «quan li han vist el cul tothom sap que és femella»; el problema és encertar-la quan no es tenen dades ni tan sols se sap quin és el problema. En això, els humans som imbatibles. Com quan construïm tota la nostra vida d'una certa manera perquè un dia vam perdre, per exemple, un trineu (possibilitat remota que a hores d'ara deu resultar familiar a tots els que coneixen un tal Kane inventat per un tal Welles).



## *La «Intel·ligència Artificial»: el nom no fa la cosa; però aquí –potser– el nom és el problema*

**A**MB la Intel·ligència Artificial (IA) hi ha un primer problema que, francament, no m'estranyaria gens que fos el pare de tots els altres. Es tracta, per començar, de l'*objecte* i del *nom* mateix de la cosa. No dubto que ens puguem dedicar a l'estudi d'una matèria sense saber ni remotament de què va el tema en qüestió. ¿No rau a la base de tota la Física el *temps*, que és una cosa que, com deia St. Agustí, tots sabem què és fins que ens ho pregunten? I no treballa pas la Física amb estranyíssims i antiintuitius conceptes quàntics que amb prou feines entenen alguns, molt pocs, dels que hi treballen? No; encara que pugui sonar estrany, en Ciència, com en la lògica o la Matemàtica, no cal saber de què es parla; i menys fer com els professors de Filosofia de Batxillerat o com els assistents a qualsevol tertúlia, aquells que sempre diuen: «primer, cal precisar de què parlem, cal definir els conceptes, les paraules». Si fos així, la Ciència no hauria aixecat mai el vol ni hauríem anat gens lluny en la tecnologia ni en la comprensió de les coses, i no és pas gaire estrany, em sembla, que ni els professors de Filosofia ni els contertulians d'enlloc s'hagin mai distingit com a científics. Quan l'Astronomia i l'Astrologia eren una mateixa disciplina practicada pels mateixos oficients, o quan la Química es deia Alquímia i pretenia una transmutació tant o més difícil que aquella que ordinàriament se'ns atribueix als catalans (és a dir, no pas treure pans de les pedres, que això es veu que ho fem cada dia, sinó or –o peles– d'allà on no en raja

BUTLLETÍ DE  
L'ACIA  
NÚMERO 24  
ESTIU DE 2001

–que això també ho fem, però ens costa una mica més–), cap d’aquestes ciències no van avançar gaire. Va haver de ser el meritori cop de cap d’alguns practicants el que va separar l’Astronomia de l’Astrologia (aquí hi ha Tycho Brahe, sobretot) i la Química de l’Alquímia (i aquí Boyle, que és el que ho va tenir més clar) i el que va alliberar així les dues matèries i les ha convertides en ciències i tecnologies útils (per saber coses, en primer lloc, i, en acabat, per aplicar-les). Ara, vol dir pas això que els astrònoms o els químics hagin hagut de començar definint el concepte d’*astre* o de *substància*? No, de cap manera! Si haguessin hagut d’iniciar-se per aquí encara ara no haurien pas encetat el bloc de notes ni començat les observacions sistemàtiques i desprejudiciades de la realitat objectiva. I encara ara és probable que demanar-li a un astrònom o a un químic respectivament «què és un astre?» o «què és una substància?» li produeixi una enorme i sorpresa perplexitat, possiblement abocada al qüestionament vital, indissoluble sense la visita regular a cal psiquiatre. Per concloure la metàfora: ¿què ens fa pensar que anem bé si definim un camp com el de la IA com l’estudi de la *intelligència* i tot seguit, com si sabéssim què és o de què parlem, posar-nos a imitar-la? I observeu que no dic pas a *estudiar-la* perquè això fóra pretensions d’allò més: tothom sap que els IAaires no *estudiem* la *intelligència*, ni en tenim ni idea ni realment ens interessa; de fet la major part de les vegades no passem de fer o proposar *sistemes intelligents* –o que ho semblin–, és a dir, que superin passablement un engany sobre la base de les expectatives i capacitat de sorpresa de l’espectador. Si hem de fer cas de com es van enlairar les dues ciències que he esmentat (l’Astronomia i la Química) –i n’hi moltes altres, no us creguéssiu, fins i tot potser totes–, el primer que hauríem de fer, en comptes d’estudiar i/o definir una vaporositat inaprehensible com aquesta de la *intelligència*, és *separar-nos* de les corresponents ciències o pseudociències veïnes, les que siguin, i quedar-nos-en alliberats per sempre. I amb les (pseudo)ciències incloc també les tecnologies i pseudotecnologies basades en l’*épatement* del personal. O potser no, potser no hi ha ningú de qui ens hàgim de separar –tret dels mags assidus als congressos que sempre miren d’enlluernar-los amb el seu darrer producte sense poder avaluar què tenen de meritori perquè en cap moment ens han revelat què volien fer o

superar i en què se n'han sortit, i en què no-. Potser simplement caldria que ens definíssim, simplement, com a *informàtics* especialitzats en camps diguem-ne *avançats*, vull dir, disposats a esmolar la nostra eina, l'ordinador, en tot allò que calgui i capaços de no espantar-nos davant de terrenys fins ara reservats al *pensament*, a la planificació si fa no fa *racional* i a les activitats que diríem *mentals*. No espantar-nos, simplement. El que no hauríem de fer de cap manera és veure'ns envaïts d'*hubris*, de triomfalisme, pel fet que alguna cosa que ha sortit de les nostres mans pugui deixar admirat i bavejant algun públic, sempre tan predisposat als miracles. No es tracta de fer de petits doctors Frankenstein ni de *crear* res, que tot això és cosa d'*artistes*, i ja se sap que els artistes són gent especial, impredecible i irreproducible –tot i que més d'un IAaire hi cau, en aquesta irresistible temptació de fer veure que *crea*, tan gratificant i gratuïta. Es tracta més aviat, em sembla, de posar tot el nostre coneixement i cor a fer possibles tasques, cada cop més complexes, que els ordinadors poden executar *ja avui* (amb un *avui* que es desplaça en el temps, per definició, i és, per tant, sempre relatiu) i que, potser, *fins ara*, només podien fer persones molt *intelligents*, vull dir capaces i/o dedicades i/o especialitzades i/o expertes i/o etc. (i en aquest etcètera incloc les *unidimensionals*, *encaparrades*, *obsessives*, *professional(itzade)s* –en sentit no necessàriament afalagador– i fins i tot les *brillants* i les *creatives*, sigui el que sigui el que vulguin dir aquestes paraules). I si convinguéssim, per a descans de tothom i repòs del nostre ego, que som simples *mecànics* d'ordinador que hem provat de fer un màster en coses informàtiques complicades tals com la complexitat –dispenseu-m'hi la redundància– de les coses complicades que tothom es veu obligat a encarar algun dia, o el tractament del llenguatge *natural*, o les manipulacions i planificacions altament enrevesades i dels sistemes complexos de sistemes, d'algorismes o de regles o del que sigui, sempre que *el que sigui* es pugui confondre en primera aproximació amb el que la gent en diu, en *vulgata*, *raonament* o *intelligència* o altres enlluernadores vaguetats del mateix estil. Això, pel que fa a la *professió*, sobre la qual, en resum, el que us proposo és, *ipso dicto*, que els/alguns informàtics es dediquin esforçadament a encarar eficaçment i útilment temes complexos, *intelligents* –vull dir, que la gent percebi com a *que toca o voreja allò que en humans ens fa*

*parlar d'intel·ligència*– i s'ofereixin, humilment, per aplicar-los a la *resolució* efectiva i eficient de *problemes* humans ja preexistents o (auto)insinuats, no pas a la *creació* personal de res, ni a l'elaboració de sistemes diguem-ne *intelligents*. Si ho són o ho semblen, intelligents, ja ens ho dirà algú –i amb això ja omplirem la nostra quota d'ego– però sempre haurem d'anar amb la prevenció de no creure'ns el que ens diuen i explicar, encara que ens costi, que el que hem fet no és pas més *creatiu* que el que fa un comptable quan organitza un caos impossible, un enginyer quan calcula un bonic i físicament improbable pont de disseny o un físic quan integra unes equacions que no caben ni a la pissarra.

Ara, pel que fa no a la professió de *IAaire* sinó a la *paraula* i matèria mateixa d'estudi i de treball, la maleïda i insuportablement atractiva *intelligència*, ¿no us sembla que fóra força més pràctic prohibir(-nos)-la? El problema de la paraula *intelligència* és que va ser inventada en una altra època, molt i molt reculada (a la Prehistòria pel cap baix, o poc se n'hi deu faltar), en què qualsevol persona que fes alguna cosa (positiva, en principi) no precisament ni estrictament una proesa física sinó simplement una cosa més aviat *mental*, visual i màgica, sobretot *efectiva*, ràpida i inesperada, i fora d'*el normal*, del comú o de la mitjana –i , això sí, que deixés parats, *sorpresos*, els col·legues– se la qualificava tot seguit amb una paraula màgica: es deia que era *intelligent*. Igual que haurien pogut dir que tenia *personalitat*, *creativitat* o qualsevol altra cosa (i, per cert, us imagineu que ara haguéssim –o ens haguessin encarregat de fer–, els informàtics, màquines amb *personalitat* –sigui el que sigui això– o potser *creatives* –i , apa, espavil! –?). El problema és que es deia, i sempre s'ha dit des d'aleshores, que una persona *intelligent* ho és perquè ha pensat o reaccionat amb rapidesa, perquè és brillant en l'ús de les paraules, perquè és capaç de fer coses que nosaltres no ens vèiem amb cor de fer o creïem impossibles, perquè tenia una agilitat, flexibilitat, rapidesa o vitalitat envejables, perquè era seductor, perquè feia màgia o tenia al·lucinacions (considerades com un *mitja de comunicació* privilegiat de les èpoques reculades), perquè tenia un *savoir faire* o una habilitat –o els tenia en un grau– insòlit o perquè feia coses que no podien altre que sorprendre i deixar corprès, esmaperdut o simplement bocabadat. Ah! I a tot això s'ha *afegit*, des que tenim màquines –i ens hi hem

acostumat– l'impossiblement rigorós criteri següent: (intelligent =) «que no és esperable d'una màquina». Si totes aquestes situacions tan i tan diferents, i que mobilitzen tan diverses i contradictòries capacitats humanes (i, a sobre, hi barregem les de les màquines que tenim [en un moment donat]), són designades amb una sola paraula, *intelligència*, ja es veu de seguida que hi tenim organitzat un bon sidral, una situació d'escapatòria més que improbable i, com a *base* d'un *programa* científic, una proposició merament al·lucinatòria. Quins programes (informàtics) farem, que siguin *intelligents*? Com sabrem que ho són? Com ho mesurarem? Com sabrem si això ajuda, imita o suplanta la nostra, d'*intelligència*, si no sols no la sabem definir sinó que cada cop que l'hem volgut merament *mesurar* (via escales insultantment simples com ara l'estàndard de Binet-Stanford) no ens n'hem sortit i, pitjor, hem començat, psicòlegs i altres contertulians amateurs, a discutir –i encara discutim sense parar des dels anys 1920– sobre *què mesura* una prova d'intelligència i, de retruc, sobre *què* carai deu ser això de la intelligència? ¿Com sortirem d'aquest entortolligament autoperpetrat i, per defició i planteig, irremissiblement irresoluble? Com va dir un intel·ligent filòsof del segle XX, quan un problema fa massa temps que s'arrossega i no fem altra cosa que anar-hi donant voltes sense avançar i de manera cada vegada més *profunda* (és a dir, obsessiva i inútil) és que ha arribat l'hora de reconèixer que el problema és purament *intel·lectual* (vull dir imaginari, sense arrel ni contacte amb la realitat *real*) i/o no té solució i/o el més probable és que estigui *mal plantejat*. En qualsevol dels tres casos s'imposa doncs clarament replantejar-lo de cap a peus (o, simplement, oblidar-se'n). ¿No deu ser pas aquest el cas de la IA, en què pretenem basar-ho tot en un concepte i en un mot tan vagues i polivalents (¿què deu tenir a veure la intelligència del seductor-que-les-enamora-totes amb la del que-caça-cèrvols-amb-gràcia o amb la del que reacciona amb desimboltura davant d'un imprevist o amb la del que, al contrari, preveu i planifica cada pas d'una situació previsible però merament hipotètica per si mai arriba el cas?) que en comptes d'ajudar-nos a resoldre de manera senzilla, unificada i potent una sèrie de problemes semblantment aparentats l'únic que fa és justament el contrari, desconcertar-nos encara més, complicar-nos la vida amb conceptes de Filosofia de



segona mà, batxilleresca, ja llargament oblidats i mai prou ben païts ni metabolitzats sobre els quals, pobres de nosaltres, ens fan definir i actuar, i que –a sobre– ens fan aplicar? (La qual cosa no vol pas dir que, de vegades i de manera corollària i tangencial, subproduïm alguns programes informàtics *avançats* que resolen més o menys eficaçment i satisfactòria algun problema concret sense que sapiguem ben bé per què ha funcionat i sense que això ens illumini gaire sobre el que estem fent.)

Voldria il·lustrar-ne els efectes poc útils, i potser nefastos, de centrar-ho tot al voltant d'una paraula màgica, que, a sobre, és ininterpretable i ambigua, amb una paràbola: imagineu-vos que cap a l'any 3500 aC a algú se li hagués ocorregut que el que nosaltres fem quan caminem es pot descriure amb la paraula *locomoció*. Bé, ja tenim la paraula. Ara suposeu que a algú altre se li ocorre que aquest atribut humà és manifestament millorable, imitable i/o reproduïble. Bé, anant per aquest camí probablement ara tindríem una tona de tractats sobre la locomoció, incloent equacions integrals i teoremes i monografies sobre l'articulació, l'abast, la sincronització de cames i juntures, etc. El que és segur és que gràcies a cap d'aquests savis no tindríem *la roda*. La sort que vam tenir és que primer va venir el senyor de la roda i va resoldre el problema –per cert mai no plantejat conscientment com a tal– de manera tan i tan satisfactòria que mai ningú no va gosar, ni tan sols se li va ocórrer, de plantejar la locomoció com a *problema* ni la seva solució general simple (la roda) com a *objecte d'estudi*. Ara: Si mai algú hagués fet aital cosa no dubteu ni un moment que se n'hauria dit *locomoció artificial* i fins i tot n'hi hauria una Associació Catalana –l'ACLA, naturalment– que editaria un Butlletí on un sòsia meu probablement provaria de pensar fins a quin punt és absurda l'*LA* i la dèria d'amoïnar-nos per si les màquines locomotores arribaran mai a suplantarnos. (Per cert, els automòbils ¿ens han suplantat/vençut? –recordeu que en depenem per a tot i que les nostres ciutats i formes de vida i de treball s'han capgirat durant el segle passat per obra i gràcia d'aquest objecte, que essencialment és una roda (quàdruple) glorificada i en forma de closca materna. I els submarins ¿han derrotat els peixos en això de *nedar* –que és la seva forma particular de locomoció–, com es demanava, intrigat, Dijkstra?) Doncs bé, en aquest món,

paral·lel, de l'LA –i tan paral·lelament inútil com el de la IA, per què no dir-ho– probablement hi hauria algun provocador compulsiu aproximadament dit Ton Sales que es disculparia perquè al número anterior no va poder assistir a la cita periòdica d'autoflagel·lació professional gràcies –és un dir– a una grip inacabable, intensa i empipadora (de la mena 'collonera», vaja, sigui dit amb perdó) paradoxalment agafada a Holanda a partir d'un canadenc que venia de l'Índia. Ja veieu que si no tinguéssim la IA (o la «LA») per entretenir-nos, amb la globalització potser no ens distrauríem però de gripes en tindríem per triar. Com de dèries més o menys inútils.



## *El 2001 no va ser el d'en Kubrick*

**N**o, el 2001 no serà pas recordat com l'any del «2001» o de Hal sinó com l'any de Bin Laden. L'única pregunta que es pot contestar, potser, és: podia ser-ho? És possible encara un Hal? Ho serà mai?

El sotassinat té una sèrie d'idees fixes, que es van repetint recursivament i de manera sostenible (és a dir, com tota entitat matemàtica, té el seu «invariant» propi o, com tothom, va a la seva bola). Per això és inevitable que, tot mirant de repassar l'estat actual (a gener de 2002) de la Intel·ligència Artificial (IA) –què n'esperàvem, què ens ha donat i cap on pot semblar que anirà– els seus prejudicis ja exposats i coneguts sorgeixin espontàniament, autoaplicats al tema.

Un dels meus més estimats punts fixos i autorecyclables consisteix a dir –i ho dic sovint– que la IA no té entitat científica com a tal, sinó que només és la punta de l'iceberg de la Informàtica, que només apareix quan les aplicacions toquen temes d'«intel·ligència» (en el sentit d'assemblar-se o recordar la humana) o bé són tan complexes que demanen pensar-hi molt a fons, i trobar-hi tractaments nous i imaginatius. Aquestes dues circumstàncies, el tipus i la intensitat de la temàtica, ens inclinen a parlar d'aplicacions «intelligentes» i de reunir-les en un hipotètic camp –suposadament coherent i estudiable– convencionalment batejat «IA». Així ha passat sempre, i el flux de la història ha fet que allò que s'havia plantejat com un tema complex d'IA (reconeixement de formes, aprenentatge estadístic, «objectes», «agents», «frames», estructures complexes de control o comunicació entre mòduls, etc.), un cop resolt, s'hagi reintegrat a la Informàtica normal com a producte o concepte corrent, eficient i desproblematitzat. Vist així, la IA

BUTLLETÍ DE  
L'ACIA  
NÚMERO 26  
HIVERN-PRIMAVERA DE  
2001-2002

no és altra cosa que un catàleg més o menys estructurat de temes informàtics que un dia van ser difícils i trencadors, però sense cap lligam intern entre ells, ni cap enfocament propi o unificador. En d'altres paraules, la IA vindria a ser respecte a la Informàtica el que seria un hipotètic repertori de troballes de la creativitat humana (que inclouria la roda, la màquina de vapor o l'aviació) estudiades i examinades sense cap relació amb el context tècnic de necessitats en què van néixer (el transport, la mineria de carbó i l'espionatge –i possible bombardeig– des de l'aire, respectivament, fa 5000, 300 i 120 anys). O ¿es pot imaginar algú quina possible línia metodològica pot unir coses tan dissemblants? Quina metodologia podria haver generat alhora la roda, el motor d'aviació i les ales? Una «teoria de la locomoció»? Un camp d'estudi anomenat «locomoció artificial»?

Segons que penso, la IA és i ha estat sempre un inventari *ex post* de totes les troballes informàtiques relacionades amb problemes que habitualment –quan parlem d'humans– considerem enquadrades en allò que diem «intelligència» o que per la seva complexitat ens obliguen a un tractament innovador i creatiu. Fer possible que un ordinador jugui un joc, per exemple, ens ha obligat a fer consideracions profundes i generals sobre l'essència i estructura del joc, sobre les transformacions funcionals en un espai d'estats, sobre la tipologia i eficiència de diverses estratègies, sobre les funcions d'avaluació heurística, etc. I el resultat s'ha pogut aplicar més tard, no cal dir-ho, a problemes diferents (de lògica, de grafs, etc.). Però això no vol pas dir que hagi estat la IA, com a metodologia prèvia, la que hagi resolt tots aquests problemes. El problema de jugar (intelligentment) va ser resolt per informàtics que, aplicant els mètodes propis i coneguts i improvisant-ne de nous, hi van descobrir idees que més tard, com sol passar, han estat aplicables a altres problemes diferents i no relacionables amb el primer. I això es fa amb la garantia dels mètodes de la bona enginyeria (de la qual la Informàtica és un cas especialment significatiu, com ens recorda Wirth). Al seu costat, la IA ha estat, com a matèria o metodologia, un tema de dimensió variable segons el moment o l'èmfasi (n'hi ha prou amb repassar els diferents manuals d'IA de moda al llarg del temps per comprovar que n'és de làbil i voluble aquest «camp»), amb pocs mètodes prou característics o invariants

que puguin justificar la trama d'una ciència pròpiament dita.

D'acord amb aquest esquema de coses –que, clarament i confessadament, no deixa de ser un prejudici com un altre– voldria repassar breument l'estat de coses en la Informàtica avançada actual, els problemes que s'hi estan plantejant i resolent avui i que, per tant, aviat passaran automàticament a engrossir i enriquir les nostres reflexions en allò que diem IA. Vist així, la segona part, també breu, es concentra en l'intent d'albirar quin possible conjunt d'idees hauran pogut entrar aviat en la IA, considerada per definició com la nostra reflexió general –sempre actualitzada– sobre la «intelligència».

En primer lloc, sembla que hi ha consens avui a dir que el 2001 real no ens ha dut cap Hal ni res que s'hi assembli, i que –seguint per la línia Kubrick– tampoc no és previsible que hi hagi mai cap David, el noi androide del film de Spielberg (*Artificial Intelligence*). Com ha dit algú fa poc, la recerca en IA s'ha fet famosa, després de mig segle d'atzucacs, marrades i frustracions, no per l'èxit sinó pels fracassos... la qual cosa diu molt, precisament, en el seu favor: com saben tots els bons pedagogs, no hi res que ensenyi més que la frustració o un bon fracàs.

En canvi, el que ha anat d'èxit en èxit, i de sorpresa en sorpresa, ha estat la Informàtica pròpiament dita. L'èxit fulgurant d'un Internet i d'un es comunicacions permeables i flexibles, l'extensió i la intensitat dels quals no havia previst ningú (ni tan sols Bill Gates), ha obert un cicle econòmic –el cinquè dels comptats per Schumpeter des de l'inici de la revolució industrial– que, si l'esquema històric es repeteix –com ha fet les quatre primeres vegades, com més va més de pressa–, toca que arribi al màxim d'influència sobre l'economia real la dècada que ve. No tota aquesta tecnologia transformadora és informàtica, és cert (també s'hi ha d'afegir la bioinformàtica i la genòmica), però la veritat és que, vist en què es basen i com es plantegen aquestes (totes dues han nascut de la potència dels nostres processos informàtics i s'autoformulen com a tipus especialitzats de bescanvi d'informació), sembla que les tecnologies amb més potencial s'estiguin convertint en branques d'una futura Informàtica en què coexistiran ordinadors i gens d'una manera còmoda i natural. I és possible també que la revolució econòmica i conceptual que ens arribarà sigui el mateix fenomen

que havia previst Alan Kay (posat al dia, fa poc, pel malaguanyat Michael Dertouzos) quan deia que la veritable revolució informàtica no arribarà –com no va arribar amb la impremta, sinó uns dos-cents anys més tard– fins que la gent haurà «interioritzat» no els ordinadors (o la impremta) sinó el tipus de pensament que hi ha portat: la idea és que, així com el racionalisme nascut de les obres impreses no qualla socialment fins que tothom se n'impregna en la seva actuació social, la Informàtica no canviarà les nostres vides fins que l'estil de vida que imposa i les conseqüències que se'n segueixen no hagin estat assumides per tothom com l'aire que respira (*projecte oxigen* es diu l'esforç col·lectiu que tirava endavant Dertouzos al MIT quan va morir).

Doncs ara va quedant clara una cosa: la «potència» (en el sentit d'assignar significats a les dades) bruta de l'humil PC arribarà a superar la del cervell cap al 2020, i la d'alguns superordinadors ja hi arribarà pel 2010. Això vol dir que aviat tindrem una «hiperintel·ligència» disponible i que probablement no sabrem què fer-ne (i que no podrem «desinventar», com tampoc no podem desinventar les armes atòmiques). Afegim-hi, per fer un bon còctel, que la Genòmica (i la Medicina) tractaran probablement més d'«informació» que d'objectes concrets i que possiblement els futurs ordinadors (o no tan futurs, que potser són a tocar) es basaran en processos moleculars o atòmics (quàntics? de spin?) i/o biològics (des de fa poques setmanes ja tenim dos enzims capaços de «calcular» totes les transposicions genètiques que els indica una tira d'ADN), amb conseqüències i trasbalsos teòrics de tota mena (us imagineu que un procés actualment incalculable o intractable esdevingut tractable –i potser lineal i tot? Renoi, com canviaria el món dels problemes informàtics!). Doncs molta gent intueix –fonamentadament– que tot això és a punt d'arribar.

Qualsevol terrabastall mental i social com el pròxim que s'anuncia ha de tenir, per força, enormes conseqüències pel que fa a la nostra concepció de tot i, en particular, del que és la intel·ligència. És en aquesta mesura que la IA canviarà, en la mesura en què reflectirà els processos intel·ligents que s'executaran socialment, a fora. Alguns trets ja es deixen veure ara, i se'n pot fer cinc cèntims sense gaire risc d'equivocar-se. La intel·ligència es veurà cada cop més com un procés semblant a la vida, només que infinitament

accelerat: els sistemes ja no es dissenyaran ni «programaran» sinó que evolucionaran i s'autoadaptaran; i presentaran, com nosaltres, fenòmens de parasitisme, virus (d'això ja en tenim), malalties i arquitectures immunodefensives. (Per tenir-ne un tast, com recomanava *The Economist* aquest Nadal passat, es poden seguir les sorprenents troballes de programes ja veterans com Tierra o dels *a-lifers* que ja fa anys que es reuneixen en congressos i es dediquen a estudiar les infinites varietats de la «vida artificial».) I també arribarem, tard o d'hora, en aquell punt en què ens adonarem –per fi!– que l'aspecte *racional* dels humans no és *tota* –ni tan sols la *base* de– la intel·ligència sinó més aviat, simplement, «la capa més feble del pensament, un procés deliberat molt fluix que, si és gens efectiu, és perquè recolza i se suporta en uns processos cognitiu sensoriomotors molt més antics, potents i consolidats, si bé normalment inconscients [...] el pensament abstracte és un nou truc que sabem des de fa segurament no més de 100.000 anys i que encara no dominem». Aquesta reflexió, feta pel divulgador robòtic Hans Moravec l'any 1988, està en la línia del que anem descobrint de la mà dels neuròlegs (Gazzaniga, Damásio) o de gent com Searle. La conclusió sembla doble i, més o menys, òbvia: (1) la intel·ligència que ens interessa extreure de les màquines ja els l'anem traient, i amb molt d'èxit fins ara, gràcies; i (2) la «nostra» intel·ligència és el que és i no hi ha cap motiu per voler imitar-la ni suplantar-la, excepte per a aquelles funcions que ens convinguin, com hem fet i fem amb la nostra potència muscular o locomotora.

Pel que fa a «la nostra», d'intel·ligència, potser el millor que podem anar fent, mentrestant, és conèixer-la bé i aprofitar-la, equilibradament, al màxim. Que d'això se n'ha dit, tradicionalment, «felicitat».





## Les «arroves» i la Informàtica

**A**ixò que segueix, de caient històric i pretensió erudita, és la versió no abreujada i lleugerament actualitzada d'un article que va sortir al diari Avui ja fa un parell d'anys i encara deu estar penjat –com el seu autor, anava a dir amb malícia– en algun racó d'Internet i/o del Termcat. Com que sembla que la desorientació sobre l'origen i el significat de les arroves, les que pesen –les clàssiques– i les que no pesen –les d'Internet, abusivament ubiqües–, sembla continuar sense treva fins i tot entre els informàtics, que teòricament en són responsables, m'ha semblat que no fóra pas sobrer de reimprimir aquestes anotacions en un context tan propi com és ara el nostre Butlletí, encara que només sigui per practicar el/la «fitness» mental o simplement per estimular la perspectiva crítica del personal i les seves neurones.

BUTLLETÍ DE  
L'ACIA  
NÚMERO 26  
HIVERN-PRIMAVERA DE  
2001-2002

Qualsevol usuari de la Informàtica s'ha degut adonar que, a les «adreces electròniques» de l'*e-mail*, hi apareix un estrany signe: @, fet d'una *a* i un cercle semitancat al voltant. I també s'ha degut adonar que «els entesos», d'aquesta cosa, en diuen arrova (si bé avui els entesos ja és tothom, per definició). La pregunta és: per què «arrova»? De fet, cal llegir-ho així? A qui i com se li va ocórrer semblant cosa?

Per a la resposta ens hem de remuntar a l'Edat Mitjana. Els copistes, carregats amb una feina lentíssima i laboriosa, feien sovint enllaços econòmics entre les lletres. Així, quan havien d'escriure una *n* darrere una *o* o una altra *n* hi posaven un cargolet o «titlla» al damunt, així: *õ* i *ñ* (signes que han quedat incorporats al portuguès i a l'espanyol, respectivament). Igualment, quan havien d'escriure la conjunció llatina *et* o la preposició *ad* feien un signe que enllaçava les dues lletres en un traç calligràfic simple. Així, *ετ* i *αδ* amb el temps van acabar & i @ (que es consideren, per cert,

l'origen, al segle XV, dels nostres signes aritmètics + i ×). Aquest últim signe, el @, volia dir simplement 'a', 'vers', 'cap a', 'en', igual com la preposició catalana 'a' actual. Aquí s'hauria acabat tot si no fos perquè en llibres d'aritmètica comptable impresos durant el Renaixement el signe @ apareixia, convenientment, com a enllaç entre el camp de les unitats (lliures, florins, cols o el que fos) i el preu que se'n pagava. Per exemple, 5 @ 2 £ volia dir '5 unitats, a(l preu de) 2 lliures (cada una)'. A Anglaterra, on @ es llegia *at*, al segle XIX encara es publicaven textos en què s'especificava que els rengles de mercaderies s'havien d'escriure així. O, avui mateix, els bancs americans fan préstecs als seus clients aclarint que són \$20,000 @ 6.5% *interest*.

A l'Espanya de mitjan segle XIX (¿o potser va ser a Catalunya, on la indústria obligava a copiar els costums comercials –i els manuals comptables anglesos?) aquesta convenció de comptables es va interpretar fantasiosament com si el número que precedia aquell signe de copistes es referís a una unitat concreta (aparentment, en molts contextos, de pes). Quina? Tenint en compte que les càrregues comercials estibades als ports eren, sovint, sacs (que solien ser, en aquella època premètrica, d'una rova cadascun), l'elecció, i la confusió, semblava òbvia. Cada rova (o arrova), unitat de nom àrab (on vol dir 'un quart [de quintar]' i deriva del mot que vol dir 'quatre'), era 25 lliures, això és, uns deu o dotze quilos. Així, rengles com ara 50 @ 10 *duros*, de significat transparent (els 50 podien ser sacs de patates o bales de cotó, i darrere el @ hi havia, clarament, el preu), van donar lloc a coses com ara 50@ (llegit «50 arroves») sense cap referència al preu. Probablement el comptable confonia aquest operador binari (dit en termes matemàtics) amb la a minúscula escrita a mà a final de ratlla (i, per tant, fantasiosament allargada o cargolada, segons el costum i l'estètica de l'època) que, aquesta sí, es referia a les arroves com a quantitat descrita (tècnicament, fent la funció d'operador unari sufixat) i que ja havia estat usada pels venecians –seguits pels florentins, pisans, genovesos i catalans, per aquest ordre– -del segle XV, si bé el fet que fessin servir casualment una 'a' com a unitat de descàrrega de naus es deu a la paraula *amphora* (i més tard àsino –volent dir 'la càrrega d'un ase o d'una mula'– o qualsevol altre identificador d'embalum, com ara aquí –per què no?– el típic sac

«d'arrova», de 10 o 12 quilos). La gràcia de la cosa (si és que en té cap, de gràcia) és que aquest malentès es va exportar d'Espanya (de Catalunya?) cap a França, on el signe @ es coneix des de llavors amb el nom d'*arobase*, nom de ressonàncies hispàniques. Per cert, a la península ibèrica la paraula sempre s'ha escrit arrova, amb ve baixa, tant en català i portuguès com en l'espanyol de Nebrija, tot i que a la seva normalització del segle XVIII els castellans van optar per posar-hi una *b* alta, vés a saber per què (...i això sempre que no vagi ser, com no fóra pas d'estranyar, un error perpetrat i fomentat per mercaders catalans hispànicament il·lustrats que, al moment d'escriure en la «llengua del cadastre» –com en deien de la imposada per la Nova Planta– i tot just autoritzats pel rei de Madrid a comerciar amb el món mundial, es castellanitzessin llibèrrimament mutant la *v* en la *b*, com ja s'havien vist obligats amb els pretèrits acabats en *-ava* o amb l'haver en textos que, si no fossin tràgicament supervivencials, resultarien còmicament absurds).

Tot això no hauria passat d'anècdota oblidada si no hagués estat per la màquina d'escriure, que va aparèixer al darrer terç del segle XIX a Amèrica de la mà de Sholes i de la Remington (la dels fusells) i que va incloure el signe @ si més no en les versions adreçades a usuaris comptables. I de la mecanografia a la informàtica, on, quan es van haver de codificar (amb set bits) tots els signes imaginables, naturalment es va pensar en el mig oblidat @ (*l'at sign* o *a mercantil*) dels comptables, així com el # (*el hash sign* o *sostingut*) dels tipògrafs (i dels carnissers!). I aquí hauria tornat a caure fàcilment en l'oblit si no hagués estat, aquesta vegada, per l'ARPAnet (l'antecessor immediat de l'Internet), en el qual cada usuari tenia un nom (de fet, el nom del seu servidor) i una localització (dividida en camps separats per punts). Entre l'una i l'altra, el signe natural era precisament el @, o així li ho va semblar el Nadal de 1971-72 a Ray Tomlinson, informàtic de BBN a Cambridge (el de Massachusetts), que, a l'hora de fer un programa per gestionar el nou correu (electrònic) i d'haver de triar un signe per remarcar clarament la separació entre el nom de l'usuari i el de la màquina des de què l'usuari operava, va usar la tecla «@» que tenia el seu terminal (un consola de teletip amb teclat comercial estàndard) i que, fora de frases «comercials» més o menys fòssils, amb prou feines es feia servir, i encara menys per informàtics,

la majoria dels quals n'ignoraven la funció i el significat. Així, una adreça com ara sales@lsi.upc.es (que és, per cert, la meva) resulta transparent: «[en] Sales [, que s'està] a [l dept. de] LSI [de la] UPC [, ubicat al ram o zona] es» (o *Sales ad LSI*, o *Sales at LSI*, que és com es llegiria). Poc s'imaginava Tomlinson, programador jove i pencaire, que estava inventant, tot modestament, una potent icona del segle XXI, o que ens estava posant a tots, catalans inclosos, en un compromís a l'hora de trobar-li un nom a la cosa, com veurem immediatament que ens ha passat a tots, suecs i swahilis, tant se val que siguem gent de ciències com de lletres (que aquests no en tenen cap culpa, els pobres).

I així ara, doncs, torno a repetir la pregunta: del desconegut i pràcticament innominat signe @, una vegada reconseguit internacionalment (una mica per casualitat), com n'hem de dir? (Aquesta pregunta, per cert, se l'ha fet tothom en tots els països.) Les respostes que s'hi ha acabat donant són varies, però es poden agrupar en set categories –a les quals goso afegir, pel meu compte, un vuitè epígraf a tall de suggeriment– dividit en dos subepígrafs :

1. Els noms espontanis: *l'at sign* (en anglès) o la *a*, la *a* de lloc o la *a* d'adreça (en català).
2. Els noms oficials (en general força encarcerats). Hi trobem la *commercial a* o *mercantile a* (en anglès) o *l'a commerciale* (en francès i en italià). També en català se n'havia dit *a comercial* o *a de Comerç* (és a dir, dels que estudiaven *Comerç*, una FP administrativa *avant la lettre*).
3. Noms que descriuen el signe per la seva similitud a una flor: *rose*, *cabbage* (=col), o a un remolí: *whirlpool*, *vortex*, *cyclone* (tots ells exclusius d'Anglaterra).
4. Noms que hi veuen un pastís en rosca (un tipus bastant universal de pasta). Aquí hi entra l'ensaimada (que va ser com es va batejar espontàniament el signe @ a Catalunya, i des d'on es va propagar a universitats de tot el món, com explico més avall), l'*strudel* (que és com en diuen a Àustria i Israel), el *pretzel* (en diversos llocs), el *kanelbulle* o el *kringla* (a Suècia) i d'altres noms de pastes rodones equivalents a la balear.

5. Noms d'animal. Aquí hi caben els micos (que és el que és el @ segons alemanys, polonesos, serbis i russos), els cargols (que és el que hi veuen francesos, italians, israelians i coreans), els cucs (hongaresos), les arengades (txecs), els gossets (russos) i els aneguets (grecs)
6. Noms de parts d'animal. Aquí hi escauen l'*snabel* a danès i suec («a amb trompa» [d'elefant, és clar]) o la «cua de mico» (l'*apestaart* holandès i el *Klammeraffe* alemany), la «cua de porc» danesa o la «pota de gat» sueca. Com a corollari un pèl més abstracte d'aquestes denominacions hi ha la «a amb cua», on l'animal ha desaparegut, o la suggerent «a amb rínxol» (*alfakroll*) dels escandinaus.
7. L'equívoca *arrova*, de què ja hem parlat més amunt, que compartim amb portuguesos, espanyols i francesos i de l'error, l'origen del qual som potser nosaltres.

A part aquestes denominacions que han sorgit espontàniament pertot arreu, n'hi ha algunes que suggereixo des d'aquí per al català, i que crec que poden ser bons substituïts –i relativament naturals– de la bàrbara *arrova* (la qual, ben mirada, no deixa de ser un disbarat lingüístic). Una és directa i respectuosa amb els orígens:

- 8.1. dir-ne, de @, simplement *ad*, en llatí, que és exactament el que vol dir i ha volgut dir sempre i pertot (i que de passada és homòfon amb l'*at* anglès, cosa que ens estalvia explicacions). Una alternativa és la que inicialment va proposar el Termcat: «a encerclada», solució que és coherent amb la © (la «ce encerclada» dels *copyrights*) i amb la ® (la «erra encerclada» de les marques) i que coincideix exactament amb la solució del japonès. Tot i que, poc més tard, el Termcat va descartar aquesta solució (el cercle no és ben bé tancat), jo no la trobo pas prou desencertada, i més quan es pot reduir fàcilment a versions més col·loquials com ara «a amb cercle» o «a amb rodona» (o «amb rínxol», com els holandesos i els escandinaus) o simplement «a-rodona» (que, per la seva relativa similitud sonora amb «a-rova», potser hauria permès d'una vegada desterrar la lamentable pífia «arrova» més fàcilment).

Per això, vista l'actitud del Termcat, proposo llegir el signe (quan apareix en una adreça) com ho fa tothom: *ad* (així, en llatí, com no ho ha deixat de ser mai), que, pronunciat amb la preceptiva neutralització de l'arxifonema *-d* en català (com es veu a salut i actitud, o a Madrid, totes tres correctament pronunciades t en català), es converteix en *at*, que és just com ho pronuncien (i ho escriuen) els anglosaxons. O, si no hi ha problema de confusió, simplement *a* (com fem quan diem «la Mercè a Barcelona, i la Tecla a Tarragona»), potser afegint, després d'una petita pausa respiratòria –i mig murmurejant– «...rodona!» o «-ep, amb rinxol!» (o «amb cua!»). Per cert, no faria pas cap mal estendre la utilitat i vivacitat de la poc útil tecla (i signe) a l'abreujament d'expressions corrents (com encara avui fan els anglosaxons i fèiem nosaltres no fa pas gaire sense cap vergonya) tals com ara:

«Ton @ Girona» volent dir en Ton de Girona, o que hi viu («resident a») o que se l'hi pot trobar

«Teatre Principal @ Lleida» (per distingir-lo del de Barcelona)

«signat @ Terrassa» (aquí l'operador binari apareix com si fos unari, prefixat)

«Sabadell @ 21 de gener de 2002»

«el 21/1/02 @ 17 h» (= «1/21/02 @ 5 pm», absolutament corrent als EUA, en citar-se)

«20.000 EUR @ 6,5 % interès» (igualmente, allà– i també, tradicionalment, aquí)

«4 barres de pa @ 0,50 EUR» (l'ús català tradicional dels pesos i els preus unitaris)

etc., etc.

totes elles prou funcionals i pràctiques, així com fidels a l'esperit i la lletra dels copistes i comptables de casa nostra que un dia van cargolar la preposició llatina *ad* (=la nostra *a/en*) convertint-la en *ad* i, tot seguit, en el rínxol que tenim ara.

Com a cosa suplementària, goso proposar també:

8.2. dir-ne (del nom del signe @, no vull pas dir *llegir-lo* així, intercalat, dins d'una adreça) *ensaïmada*, mot que no sols resulta natural i gràfic sinó que, contra el seu aparent localisme, ens situa –amb suecs, alemanys, austríacs i israelians– en el terreny cultivat universalment, com hem vist més amunt, de la descripció del @ via pastissos en espiral. Pot ser que fent-ho contribuïm a cultivar i propagar, a més del català, l'alimentació mediterrània. (Per cert, el primitiu ús d'ensaïmada com a nom universitari semiformal del @ es va propagar bastant amb l'*e-mail* dels primers temps –vull dir, en el cas de les tres universitats barcelonines i el seu enllaç amb el CERN, el 1984-86–, de manera que jo m'he trobat més d'una vegada amb científics estrangers –des de suecs a taiwanesos– que «sabien» que a Catalunya el @ es llegia *ensaïmada*.) Ara, si es vol una paraula prou neutra i que no sigui l'ensaïmada, sempre ens queda el *rínxol*, que és bastant universal com a concepte (el capillar abraça usuaris des de Burundi a la Carmen de Bizet i a Greta Garbo) i que com a paraula descriptiva d'un signe d'ús corrent ens agermana, com l'euro, amb els nòrdics. O, si voleu, *a amb cua* (concepte encara més universal). O el més universal i lliure de tots: la *a*, simplement, acompanyada de pauses respiratòries o no (o amb *t* final, o amb gemecs, o el que es vulgui). Que no fos cas que tot i tenint prou llibertat de triar féssim curt.

(ítem arqueològic recuperat el 13/1/02 @ 22h)





# Intel·ligència és més que un mot

*Perquè ens afecta tant la paraula intel·ligència?*

BUTLLETÍ DE  
L'ACIA  
NÚMERO 27  
ESTIU DE 2002

L'ÚLTIMA setmana de juny ens ha arribat una notícia que trobo reveladora. L'exèrcit americà ha prohibit, en principi, que volin sobre l'Afganistan, i pertot arreu en general, avions d'observació o de bombardeig sense pilot (els *drones*, o borinots, que en diuen ells) completament automatitzats i provadament segurs: Per què? Doncs perquè els avantatges enormes que tenen aquests aparells (entre altres, i no és pas poc, el d'estalviar la vida d'un pilot en cas d'estimbament) no compensen la *prevenció* que causa tant en civils com en militars tenir una *màquina* completament autònoma volant de cap a peus pel seu compte, prenent les seves pròpies decisions i garantint la seguretat pròpia, fins allà on és possible, i també (i això amb una prioritat absoluta) la de terceres parts. Aquesta prohibició s'ha fet pensant en la probable cohabitació en el futur, vista la gran fiabilitat d'aquestes màquines, d'avions sense pilot i civils (i el món en general). El sol pensament que hi pogués haver mai un accident causat per un *borinot* d'aquests i la mala cara que els tocaria fer als militars davant la mort d'un sol civil, especialment si aquest fos americà, fa empal·lidir l'exèrcit de cop. En termes civils, això és fàcil d'entendre: quina companyia d'assegurances fóra capaç d'assegurar un borinot? Ja se sap que les asseguradores compten fins al mil·lieuro les probabilitats d'un desastre multiplicades pel cost (allò que els estadístics innocentment anomenen *esperança*), ho comparen amb el total de primes ingressades fins al dia del desastre i, si la diferència entre la segona i la primera quantitat no és pròxima a infinit, ho deixen córrer i diuen que la cosa no els

interessa gens.

Fins aquí tot és comprensible, pràctic, neutre i corrent. Ara proposo un simple experiment mental. Imagineu que apareix als diaris la notícia que ara ja disposem d'un sistema intel·ligent (subratllo la paraula intel·ligent) que és capaç de dirigir tots els aspectes d'un hospital (i prendre, si cal, totes les greus decisions pertinents) o d'un supervisor intel·ligent que atén la cura intensiva d'un pacient (incloent-hi la decisió, possiblement tan ponderada o més que la d'un metge humà, sobre qualsevol aspecte delicat) o la gestió d'un aeroport sencer o, per fixar idees i per analogia amb el borinot, del vol complet –des de l'embarcament fins a la descàrrega– d'un avió ple de passatgers humans (amb tanta prudència i precaucions com es vulguin a l'hora de prendre decisions que afectin la seguretat dels viatgers o dels sistemes de terra). Segur que la notícia es comentaria al carrer i ens faria plantejar la qüestió: «Vostè, volaria en un vol així?» o «Confiria la seva vida o la dels seus a una màquina?» i altres titulars fàcils d'imaginar. I observeu que la mera inclusió de la paraula *intelligent* produeix el miracle. Ara imagineu-vos que la notícia es dóna sense esmentar enlloc la paraula i dient, per exemple, «Tots els aspectes del vol estan constantment controlats, i la seguretat està absolutament garantida en tot moment» (em podeu objectar que la seguretat mai no és *garantida*, tampoc en el cas de pilots humans –i tothom sap que, de pilots, n'hi d'estúpids a matar–, però això, tot i saber-se, no s'esmenta mai). Feu l'experiment amb un grup reduït de gent (jo l'he fet) i veureu la diferència. Si no hi surt la paraula intel·ligència ni per associació, la qüestió no passa de carn d'asseguradora, de suma algebraica dels beneficis (que és la cosa normalment esperable) ingressats menys els costos *esperables* en cas de desastre i si fa no fa aritmèticament estimables i afitables, un assumpte que no mobilitza pas ni filòsofs ni periodistes ni gent de barra de cafè, i sí en canvi –només– estranys éssers matemàtics anomenats probabilistes o actuaris; tot es redueix a la pregunta, força més simple i poc propensa a esdevenir titular de cap diari: «Vostè, subscriuria una pòlissa d'assegurança per un vol en aquestes condicions?» o «Deixaria la vida d'un familiar en estat crític en mans d'un aparell molt provat i sofisticat més que no pas en mans d'un metge segurament fallible i possiblement antipàtic (i fins i tot, cas no pas insòlit, incompetent)?» Preguntem-ho, o

preguntem-nos-ho, i em sembla que pot quedar clar que la paraula intel·ligència sempre introdueix soroll en qualsevol conversa, i a segons quin interlocutor el posa immediatament a la defensiva.

L'exemple de la notícia que he esmentat em sembla que il·lustra també un altre fenomen que crec consubstancial a la Intel·ligència Artificial com a enginyeria que és. Un cop fetes totes les consideracions que es vulguin sobre què és i què no és la intel·ligència, el sistema que acaba essent projectat i que es materialitza ha de ser, necessàriament, fiable, és a dir, no ha de fallar i ha de ser controlable, en tot i tothora: no valen, doncs, màquines que aprenguin dels seus errors; ras i curt: no hi d'haver errors. Dit altrament: a una persona se li poden perdonar –potser– certs errors; a una màquina cap, mai (i si és que se'n preveu algun, la màquina no es construeix, i llestos). I ara, després d'aquest petit *excursum*, vaig a la part que més m'interessa a mi personalment, i qui sap si a algú altre:

### *Què coi deu ser, doncs, això de la intel·ligència?*

Sobre això, no se m'ocorre cap resposta millor que la que n'han donat sistemàticament els militars americans durant els darrers vuitanta anys. No ho han escrit enlloc, fora dels seus propis informes interns, ni cap civil no ho ha descobert i divulgat, ni cap filòsof se n'ha assabentat i ho ha inclòs en cap llibre de Filosofia. Però la història és molt coherent, i ve de lluny. Vegem-ho.

A tombant del segle vint hi havia en plena florida una branca de la psicologia experimental amb laboratoris i noms il·lustres com Fechner, Weber o Wundt i també el francès Alfred Binet. Aquest va mirar de parametritzar els diferents components de la intel·ligència i va acabar confegint una escala de *punts*, que s'ha fet cèlebre. Aquí hauria acabat la cosa (o hauria continuat colgada enmig de les pàgines d'un manual de psicologia, com la llei de Weber o de Fechner-Wundt sobre la percepció), si no fos perquè l'exèrcit francès, a propòsit de la primera guerra mundial, va mirar d'aprofitar la feina de Binet per *predir* el grau de fiabilitat que es podia raonablement atribuir a cadascun dels cridats a files un cop convertit en soldat i, per tant, en un context dramàtic com és la guerra, causa possible d'actes desencertats, pèrdues de vides o

qualsevol altra greu responsabilitat. La mateixa idea els va sobrevenir als militars americans. Decidits a intervenir a Europa al 1917, els soldats i voluntaris reclutats apressadament, venint com venien d'un camp laboral totalment diferent al de l'estressant guerra de moviments a camp obert i de trinxeres, no oferien gaire garantia de *funcionar* adequadament en aquest context tan nou per ells i en el qual, sense gaire experiència, tanmateix havien de prendre, sobretot els comandaments, decisions sovint de vida o mort per a centenars o milers de soldats (o de civils). L'exèrcit americà, ple de presses, va fer el que va poder per preseleccionar el personal, però, un cop acabada la guerra i desmobilitzada la tropa (i preveient que allò pogués repetir-se), va fer l'exercici d'avaluar com havia estat classificat cada u a l'hora de reclutar-lo i com s'havia comportat a l'hora de la veritat. Va demanar l'ajuda de la Universitat de Stanford, la de Palo Alto, fundada per un especulador ferroviari aproximadament penedit. Allà es va configurar allò que ha acabat sent el *test d'intel·ligència* per antonomàsia, el d'Stanford-Binet. No es tractava pas de definir ni de saber què és la intel·ligència; només fer un predictor empíricament adequat, basat en les correlacions vistes a posteriori, que permetés estar raonablement segur de les característiques del portador del test: capacitat d'iniciativa i de resposta, flexibilitat i adaptació davant del canvi sobtat (o no) de la situació, capacitat de veure l'entrellat d'una percepció o d'un complex d'estímul, saber destriar l'important dintre d'un cúmul o veure la línia d'acció a seguir entre diverses possibles, tenir prou sensibilitat per saber fins on es pot arribar sense tibar massa, tenir capacitat d'arrossegar i d'engrescar als altres, saber crear una idea general de direcció que se sobreposi al neguit de l'immediat, i així podríem anar continuant, perquè el que l'exèrcit americà demanava d'un soldat és, fet i fet, el mateix que un cap de personal demana a un candidat a una feina (comprensió ràpida i prou completa de les coses, personalitat, iniciativa, flexibilitat, adaptabilitat, lideratge, etc.) i el que un qualsevol de nosaltres espera que tingui una persona a qui algú ha qualificat prèviament d'*intel·ligent*.

Aquesta definició, pragmàtica i sense pretensions, de la intel·ligència coincideix més o menys amb la definició intuïtiva habitual i a l'exèrcit americà (i a les empreses americanes que l'hi van demanar com a criteri de preselecció de personal) li ha funcionat sempre

prou bé. I, sense que no gaire gent se n'adonés, s'ha convertit en el tema subjacent de la Intel·ligència Artificial (IA) d'avui. Perquè cal no oblidar que, al 1955-56, el primer contacte exterior que va tenir McCarthy al moment d'inventar i proposar els seus programa i concepte –i l'expressió mateixa d'*Intel·ligència Artificial*, consistent en essència, com se sap, a fer programes d'ordinador que fossin, diguem-ne, *intelligents*, o que ho semblessin– va ser un psicòleg d'alta graduació militar que li va exposar el problema que tenia l'exèrcit amb la intel·ligència (la dels soldats, no pas la dels espies) i que li va prometre (i obtenir) un bon finançament per al seu programa, cosa que per cert mai no li va faltar a partir del mateix 1956.

Així que, si per un atzar mai algú em preguntés què vol dir *intel·ligència*, ja tinc preparada la resposta: «Pregunta-ho als militars. D'això, en saben». Si no fos que sé la història, no hauria dit pas mai.

Barcelona, 30 de juny de 2002



# El context de la Intel·ligència Artificial

**L**A Intel·ligència Artificial (IA) va començar quan l'intel·ligentíssim matemàtic anglès Alan Turing va formular amb precisió, al 1936, en què consistia *pensar* o, més concretament, fer de *computer* (*professional calculista*). Llavors, com ara, se suposava que l'activitat mental d'un matemàtic era el paradigma de *la intel·ligència*, almenys dels seus aspectes més racionals i rigorosos. Turing ho va explicar tot d'una manera tan clara que el procés (de *pensar* o *calcular*) acabava resultant executable *mecànicament*. Només faltava una *màquina* real que ho fes, i aquesta va ser, al 1946, l'ENIAC, el primer ordinador. Amb aquests models (els de Turing) i aquestes màquines (els ordinadors) ja podíem accedir a l'estudi i la imitació de la nostra racionalitat. Així va néixer la primera IA, la dels anys 1946-56, que, encara sense dir-se'n (es parlava de «Cibernètica»), ja permetia a uns il·lusionats científics-interdisciplinaris (de matemàtics a metges, passant per lògics i biòlegs) atrevir-se a estudiar *la intel·ligència*, descriure-la, estudiar-la i parlar-ne (i, potser, algun dia, imitar-la o reproduir-la).

Al 1956 això va canviar. Van passar dues coses. Primera: IBM va regalar ordinadors a les universitats americanes. Van anar a parar ràpidament a mans dels estudiants més brillants, que s'hi van posar a programar i a resoldre problemes (els de tipus més complex, creatiu i *intelligent*) mai abans no abordats per una màquina, com ara jugar a dames o escacs, o planificar prèviament una seqüència complexa d'operacions necessàries per assolir algun objectiu (per exemple, reconèixer figures semblants, classificar objectes, guanyar en un joc, reconèixer cares, entendre frases) i, fins i

BUTLLETÍ DE  
L'ACIA  
NÚMERO 27  
ESTIU DE 2002



tot, el *súmmum*: ser capaç de demostrar un teorema rigorosament, com ho faria un matemàtic. La segona: uns quants universitaris americans joves (McCarthy, Minsky, Simon) van proposar una nova estratègia (que ell mateixos van batejar *Intel·ligència Artificial*, nom avui omnipresent) que és, encara, la nostra. En reacció als il·lustres precedents interdisciplinaris de la dècada anterior, aquests matemàtics no volien pas *estudiar* la intel·ligència (i menys el cervell) ni reproduir-la, sinó, merament, *fer programes* (d'ordinador) que fessin coses que, fetes per humans i vistes als ulls dels humans passessin per *intelligents* o que, si més no, *ho semblessin* (i, encara més, que *sorprenguessin* precisament pel fet d'haver sortit de màquines, no d'humans). Aquest era un enfocament molt simplificat de la feina proposada: *no fer coses* –aparells o màquines– que *fossin intelligents*, sinó agafar una màquina coneguda i relativament maniable, *l'ordinador*, i fer-li executar programes que *fessin coses intelligents*. No era *ser* sinó *fer*. Tot i l'aparent elementalitat de la seva concepció, aquest esquema ha donat lloc a èxits espaterrants i han impressionat tant els tècnics com, a través de les columnes dels diaris, al públic en general. Als anys 1970 hi havia *sistemes* (bé, de fet simples *programes*, siguem clars) que feien coses com *parlar* (o, si més no, mantenir una aparença de conversa *intelligent*), *deduir* conclusions no evidents, *decidir* i actuar enmig d'un entorn complex, *interactuar* amb altres agents amb un mínim de coherència, *diagnosticar* i ajudar a resoldre problemes (*raonant-los*), jugar (i guanyar) als escacs, etc.

La línia encetada al 1956, que tants èxits ha donat i que encara continua viva (malgrat els dubtes, debats i canvis de tecnologia) és bàsicament aquesta: com aconseguir que els ordinadors de què disposem (que potser aviat seran quàntics o qui-sap-què) *facin* coses que els observadors presents, no necessàriament ignorants ni crèduls, puguin considerar *intelligents*. Ara, si el que algú reclama no és pas un programa que *faci* (o sembli) sinó una màquina que *sigui* (*intelligent*), en aquest cas val més prendre's una bona dosi de paciència i uns quants anys de coll, i esperar. Llavors –potser– podrem parlar, per primera vegada, de *què* és la intel·ligència i de si l'hem arribat a *materialitzar* (o reproduir) en màquines. Però això, em sembla, encara és ciència-ficció.

*[Aquest petit article va sortir publicat a La Vanguardia del diumenge 23 de juny del 2002 (a la pàgina 28, juntament amb treballs d'en Ramon L. De Màntaras i de l'Enric Plaza). Com que em vaig autoconcedir el dret de publicació de l'original català (que és el que segueix), em permeto, amb l'amable i pacient autorització de l'editor del Butlletí, fer-lo públic aquí com a apèndix de la columna que encapçala aquest número. Excuseu-me l'atreviment i no m'ho tingueu en compte.]*



## Qüestions irresoltes: Potser que hi tornem?

**E**M sap greu pel lector, però a risc de fer-me (més) pesat voldria tornar al tema doble de la meva darrera intervenció. M'hi preguntava, primer, per què som tan rematadament sensibles a la paraula «intelligència» i, segon, quina cosa -ja que no la paraula- pot resumir més bé el concepte d' «intelligència». Per il·lustrar la primera qüestió em plantejava fins a quin punt estàvem psicològicament preparats per volar en un avió sense pilot, com aquests «borinots» americans que volen sobre l'Afganistan i l'Iraq però en versió civil i amb nosaltres com a passatgers, i si cap companyia d'assegurances consideraria un vol com aquest de «tarifa normal». Precisament al mateix moment que m'ho plantejava i ho escrivia va passar a Suïssa una cosa extraordinària que il·lustra molt bé, trobo, l'assumpte que jo discutia; per això m'agradaria tornar-hi, perquè veig que m'hi vaig deixar coses que m'havien quedat al teclat. I sobre la segona qüestió em demanava si no deu ser que els militars, d'això de «la intelligència», en saben un cove (gràcies a la seva veterania en les proves d'intelligència, vull dir, de la mesura del Quocient Intel·lectual o «QI» dels *caloios* que els arriben, presumptes guerrers quasiprofessionals). Doncs bé, amb noves dades a la mà, tretes d'un congrés de psicòlegs que es va fer a Nashville al desembre, m'agradaria abundar-hi. Hi torno doncs, amb perdó, i ho reprenc al punt on ho vaig deixar.<sup>4</sup>

BUTLLETÍ DE  
L'ACIA  
NÚMERO 28  
HIVERN-PRIMAVERA DE  
2003-2004

<sup>4</sup> L'autor es refereix a l'article de la plana 209. (n. de l'ed.)

*Per què ens afecta tant la paraula intel·ligència?*

Mentre jo feia consideracions, aquí mateix, sobre si gosaríem volar en un avió que ens haguessin dit que anava sense pilot (és a dir, «intelligent») o en un altre de tan intel·ligent com el primer del qual no ens haguessin dit res de pilots ni de presumptes intel·ligències, al cel de Suïssa va ocórrer un desgraciat xoc d'avions en el qual van morir entre d'altres molts infants ucraïnesos que venien a Catalunya. Un cop reconstruïts els fets, sembla que el que va passar va ser que els sistemes automàtics, intel·ligents, dels dos avions van predir correctament que hi hauria un xoc i van determinar que un d'ells havia de canviar de cop la trajectòria. I així ho van notificar als pilots que, tot i creure-s'ho, van optar per confirmar-ho a un ésser humà de debò, a terra (Justament Suïssa, on els éssers humans tenen fama de cultes, nets, desvetllats i precisos). Com sol passar en els nostres càlculs mentals, els dos pilots es van deixar guiar per l'estereotip de «persona» (a la qual es pot parlar, i que «entén» les situacions) en contraposició a «màquina» (que és el contrari) i es van fiar d'un suís perquè el digués sàviament què havien de fer. I va passar que el suís que els va contestar era un suplent supernumerari faltat d'hores de descans i amb sobredosi de son, de tensió i de responsabilitat que no va saber o no va poder consultar uns instruments que sembla que tampoc no funcionaven perquè els havien desconnectat. Un seguit de nyaps, vaja, tan poc «suïssos» (segons l'estereotip) com reals. I aquesta persona va ser la que, creient en miracles o directament al·lucinat, els va dir que tranquils, no canviu de trajectòria que no passa res (per la regla de tres segons la qual un xoc tan terrible no pot passar de cap manera, Déu no ho permetria pas). I se'n van fiar, i va passar, i es veu que Déu ho va permetre.

No sé pas si els IAaires, els tècnics d'aviació i la gent normal hi han pensat prou i ho han metabolitzat bé, això, i no sé pas si han arribat a la mateixa conclusió que jo. El que jo he conclòs d'aquest malaurat accident és que, si els pilots s'haguessin fiat menys del company que els contesta i els entén i més de la màquina que decideix sense entendre'ls, la cosa segurament hauria anat d'una altra manera. I ara tindríem uns nens recordant la platja. No dic que s'hagi de confiar cegament en una màquina, només cal

EL QUE JO HE CON-  
CLÒS D'AQUEST  
MALAURAT ACCI-  
DENT ÉS QUE, SI ELS  
PILOTS S'HAGUES-  
SIN FIAT MENYS  
DEL COMPANY QUE  
ELS CONTESTA I ELS  
ENTÉN I MÉS DE  
LA MÀQUINA QUE  
DECIDEIX SENSE  
ENTENDRE'LS, LA  
COSA SEGURAMENT  
HAURIA ANAT D'U-  
NA ALTRA MANERA.

relativitzar les nostres expectatives sobre el que podem esperar d'uns i d'uns altres. I que quan ens diguin que un avió vola manat per un sistema que segur que no parla amb nosaltres i és més fred que el metall no ens espantem més enllà de cap mesura, i que considerem que potser si que, al capdavall, és més fiable aquesta màquina que ha estat fabricada per a això i no es cansa, que no pas aquell pilot amb cara de playboy fatxenda sindicat i sobreassalariat que no sembla pas que estigui més per nosaltres i pel vol que per la broma i el ha! ha! ha! amb les hostesses (i no ho dic per estereotip, és que n'he trobat a munts, d'aquests, i no sempre m'han inspirat la confiança i la tranquil·litat que se suposa que m'havien d'inspirar). Els pilots automàtics, pures màquines, avui són capaços de volar i volen en condicions adverses com ara la de falla d'un de dos motors i, així, no són poques les vides que han salvat sense rebre'n cap medalla. I sense recórrer al dramatisme, ¿ja sap, la gent, que cada vegada que aterra amb boira és sistemàticament una màquina i no el pilot el que ho fa tot i li assegura una tranquil·la i insuspecta caiguda (generalment més suau que els sotrats del playboy)? Si els ho diguessin un cop aterrats, què farien? Creurien més en la IA (és a dir, en la intel·ligència de les màquines, perquè no és sinó d'això de què es tracta)? I les companyies asseguradores, se'n refiarien potser més dels pilots artificials que dels que van de blau (per molt «naturals» que siguin, aquests últims)?

Jo no ho sé, però el que veig clar és que perquè la gent accepti amb tranquil·litat ser pilotada per una màquina –i «passar» d'uniformes blaus– calen si més no tres coses: (1) que l'avió i la màquina siguin – i es demostrï que siguin– completament segurs, (2) que les companyies d'aviació tinguin una bona raó per usar-los i deixar de confiar-se a rars i sobrevalorats pilots de carn, ossos i uniforme, i (3) el més important: que els passatgers estiguin disposats a volar-hi, o que algú els hi hagi venut la idea amb tanta habilitat que no hi tinguin cap inconvenient ni aprensió, a volar-hi (ni les companyies asseguradores a trobar-hi una prima raonable). Els dos primers factors són objectivament aconseguibles i mesurables. El tercer, ai!, és purament **subjectiu**, com tantes altres coses que envolten aquest camp nostre de la IA... perquè, en relació amb les màquines intel·ligents, cal recordar com hi compten –moltíssim!– elements subjectius com ara la *sorpresa* (la intel·ligència és bàsicament sor-

LA INTEL·LIGÈNCIA  
ÉS BÀSICAMENT  
SORPRESA, NO EM  
CANSO DE DIR-HO.

presa, no em canso de dir-ho), les *expectatives* més o menys (sovint menys) racionals, l'*aprensió* (aquesta sí, totalment irracional), etc., factors que, encara que teòricament no tinguin lloc en la ciència, sovint són els que acaben per determinar la qualitat d'un producte IA o la nostra apreciació de la IA com a activitat científica.

Tot això em porta a una avaluació prudent del **futur** de la IA. La IA no tindrà sols el futur que li donin els seus enginyers sinó, sobretot, el que li doni l'**acceptació** psicològica suau i atraumàtica dels seus consumidors o destinataris finals. Cal que ho vegin tot plegat com un avantatge i no com una amenaça, i que ho interioritzin sense complicacions ni trasbalsos (i que estiguin disposats a confiar-hi, sense gaires manies ni esgarips, fins i tot la pròpia vida, en una activitat tan personalment compromesa com és un vol d'avió).

*Què coi deu ser, doncs, això de la intel·ligència?*

I ara torno al meu segon tema. Si la vegada passada vaig arribar a la paradoxal conclusió que no hi ha res com ser militar per saber bé en què consisteix la intel·ligència ara, un cop llegits els treballs d'una colla de psicòlegs reunits a Nashville el desembre del 2002 (i no precisament per sentir-hi música *country*), reprene el fil. Aquests psicòlegs, vinguts de tot arreu, parlaven de les mesures del QI (el quocient d'intel·ligència). Com sempre, s'hi repetia aquell argument tan suat que l'únic que mesura el QI és la capacitat que té el subjecte per superar les proves del QI, o que el QI és molt sensible a l'entorn cultural del subjecte, o aquest altre, un mica més sofisticat, segons el qual el que realment mesura el QI és el factor *g* (el «factor general» d'intel·ligència) i que, a més, el mesura força bé... el problema és que ningú no sap què significa o representa aquest «factor». També s'hi van sentir crítiques de totes bandes sobre la falta d'idoneïtat de la mesura del QI però, i això és graciós, ningú sabia proposar, després de gairebé un segle de vida del QI, res millor, i tothom tendia a constatar que, tot i els seus aclaparadors defectes, el QI era la menys dolenta de les mesures, totes insatisfactòries. I fins i tot es va arribar a dir que, encara que no se sabés què significava *g*, quan s'introduïa com a factor als càlculs i se l'intentava correlacionar amb d'altres, els

resultats no eren absurds del tot. I que, ves, potser sí que el QI volia dir alguna cosa... cosa que d'altra banda tampoc no fóra tan desenraonat: al cap i a la fi se suposa que *g* correspon *grosso modo* a la capacitat d'aprendre, raonar i saber resoldre problemes del subjecte, i que la prediu amb força exactitud (precisament aquest grau de confiança és el que els militars volien mesurar davant d'una situació imprevisible i complexa com la d'una guerra mundial, deia jo al meu escrit anterior).

Una mostra: una sociòloga de Delaware va fer exhaustivament una sèrie d'estudis que no deixen gens de lloc al dubte: el QI es correlaciona magníficament amb l'«èxit» a la vida (en el sentit de bona *qualitat de vida* i satisfacció o sentiment de «realització», segons totes les mesures possibles: dit senzillament, la gent amb QI alt és més *brillant*, i viceversa). Un altre estudi, fet a Austràlia encreuant dades de QI i d'hospitalitzacions, porta a la següent relació sorprenent (... o potser no): la gent amb un QI de 100 o més, dóna una taxa d'accidents de trànsit de 51 per 10.000; els qui estan entre 85 i 99 en donen un 92, mentre que els qui van de 80 a 84 passen de 146 (sembla, doncs, com si la intel·ligència impliqués més control, domini o sort). I un altre troba que els pacients amb poc QI provoquen –deixant de banda els accidents– més despeses al sistema sanitari... ja sigui perquè no entenen les indicacions d'un prospecte (i això es veu que afecta el 40% de la població americana) i per això s'acaben prenent amb l'estómac ple –i potser deixant-la sense efecte– una medecina que s'havien d'haver pres en dejuni, ja sigui perquè, per exemple, no acaben d'entendre bé què comporta signar un document d'acceptació (i això es veu que encara afecta més americans, el 60%). Un altre estudi, aquest fet a Texas sobre una població d'infants en entorns desestructurats i «d'alt risc», hi descobreix un invariant, un percentatge estable d'infants «resistents» (és a dir, que tot i les circumstàncies difícils se'n surten sense caure en la depressió, la delinqüència o el fracàs escolar); doncs bé, els més «resistents» (*resilient*) coincideixen amb els de més alt QI (quan tenien 8-9 anys) mentre que com més baixa el QI més baixa la «resistència». I encara un altre, aquest d'Edimburg: resulta que encreuant els QI d'infància (que s'han conservat) de persones que ara tenen 80 anys, es constata clarament que els QI més alts es correlacionen molt precisament amb els que

SEMBLA DONCS  
COM SI LA INTEL·  
LIGÈNCIA IMPLI-  
QUÉS MÉS CONTROL,  
DOMINI O SORT.



han tingut més qualitat de vida, més longevitat sense problemes i que, vaja, han funcionat més bé; el recercador corresponent, un tal Gottfredson, aventura que si hi ha res que resumeixi el factor *g* potser no és altra cosa que, «senzillament, la capacitat de tenir una existència relativament exempta d'errors».

Tot això, extret del gremi dels psicòlegs, sembla que s'avé prou amb les consideracions dels militars americans d'entreguerres que feien servir el QI per veure quina «intel·ligència» general es podia esperar dels aspirants a guerrers. Però és que la conxorxa sembla estendre's al col·legi de metges. En efecte, ja fa 12 anys que, analitzant escànners per emissió de positrons, un metge de la Universitat de Califòrnia d'Irvine ja es va adonar que com més QI menys energia consumia el cervell per fer una mateixa feina (en aquest cas es tractava de jugar un joc d'ordinador que prèviament s'havia après). Doncs ara resulta que el mateix passa quan es fan feines de baix nivell i consum intel·lectual (com ara la de veure passivament un vídeo arrapat al coixí del sofà), segons explica un metge de Stanford a *Intelligence*, que hi afegeix que «sembla que el cervell dels qui tenen un QI alt funciona de manera diferent -fins i tot quan no es tracta de «pensar» (raonar)- amb un llinar d'activitat més alt i amb més economia energètica». L'article el presenta un col·lega seu d'Stanford, que postilla amb la següent hipòtesi: «sembla que els (homes i primats en general) que «pensen» amb més rapidesa i més flexibilitat són capaços d'activar més selectivament i més ràpidament les zones cerebrals que més convenen per a cada funció... en resum, que el cervell els funciona més bé».

No sé si això us ha deixat pensarosos. A mi sí, m'ha activat el còrtex. I és que potser un dels efectes col·laterals de la intel·ligència és que es pugui plantejar fins i tot «què coi (amb perdó) és la intel·ligència?». És clar que no sé si això porta enlloc. Però almenys serveix per fer segones parts d'un article.

## *Podrà mai un ordinador emular el cervell humà?*

**L**A resposta és simplement i rotundament sí! Només cal esperar. Simples consideracions ens indiquen que, a les taxes actuals –o menors– de progrés del *hardware*, la «potència» bruta (en el sentit d'assignar significats a les dades i en el de capacitat de càlcul i de memòria disponible, en gigabits) de l'humil PC arribarà a superar la del cervell cap al 2020, i la d'alguns superordinadors ja s'hi podrà comparar cap al 2010. Això vol dir que aviat tindrem una «hiperintel·ligència» al nostre abast, i que probablement no sabrem què fer-ne. (Hiperintel·ligència, per cert, que no podrem «desinventar», com tampoc no podem desinventar les armes atòmiques.) Afegim-hi encara això: potser els ordinadors del futur (o de demà mateix, qui sap) seran força diferents de concepció, aspecte i funcionament dels actuals. És prou possible que estiguin basats en processos moleculars o atòmics (quàntics? de spin?) o també biològics (des de fa uns quants mesos ja disposem de dos enzims capaços de «calcular» totes les transposicions genètiques que els indica una tira d'ADN).

La pregunta que segueix també és molt simple: i llavors, què farem nosaltres? Com assimilarem el fet de veure'ns «superats» per les nostres màquines? I aquí la contesta no és tan elemental. Depèn. Depèn de com ens preparem des d'ara fins a l'esdeveniment fatal. Si l'experiència històrica serveix de res, no és gaire difícil aventurar dos desenvolupaments paral·lels i aparentment contraris. En primer lloc, segur que n'hi haurà uns quants que s'espantaran –fins i tot molt– amb la perspectiva, i que escriuran llibres per encomanar-nos la seva por. Això se sol fer molt; de

BUTLLETÍ DE  
L'ACIA  
NÚMERO 29  
TARDOR DE 2003

ENS ACOSTUMEM A  
TOT, ENCARA QUE  
PUGUI SEMBLAR  
O HAGI SEMBLAT  
MONSTRUÓS; N'HI  
HA PROU QUE ENS  
CONCEDEIXIN UN  
PERÍODE DE DIGES-  
TIÓ

fet, les successives apocalipsis que hem patit han anat trobant variades excuses i motius: la misèria industrial del segle dinou, la nuclearització del planeta del vint, el desastre ecològic irremeiable que es postula per a l'actual, etc. En realitat, part de la feina (d'espantar-nos amb supermàquines pensants) ja la tenim feta: autors clàssics de ciència-ficció ja es van plantejar que tinguéssim robots amb problemes ètics (Asimov ho va fer el 1939, després de sortir impressionat d'una visita als «robots» que s'exposaven a la Fira Mundial de Nueva York d'aquell any) o superordinadors amb problemes de consciència (i omnipotència neuròtica) com el Hal del 2001 –de fet, del 1968– de Clarke, entre moltíssims altres. El segon desenvolupament és més benigne i, tot i soterrat (poca gent se n'adona) i poc espectacular o literari, ha ocorregut en quasi totes les revolucions tecnològiques anteriors, i és que ens acostumem a tot, encara que pugui semblar o hagi semblat monstruós; n'hi ha prou que ens concedeixin un període de digestió. Ens hem acostumat a viure en ciutats (això va ser fa gairebé deu mil anys), a llegir, a poder continuar llegint (a partir dels quaranta) gràcies a les ulleres que van estrenar els intel·lectuals i copistes medievals (a qui l'«intolerable» invent va salvar el sou), a veure un llibre «escrit» per una premsa de prémer vi, a tancar-nos en una nau per treballar-hi concentrats i sense parlar durant deu hores seguides o més sis o set dies a la setmana), etc. No cal pas aclarir que totes aquestes coses es van percebre al seu dia sense excepció com a totalment antinatural, monstruoses i difícilíssimes de comprendre i acceptar, i que van anar acompanyades de les corresponents literatures apocalíptiques. Les més recents brutalitats tecnològiques són també les més compartibles i fàcils de recordar. Avui ja no se sorprèn ningú del fet que una màquina tingui la força de innumbrables peons (penseu en una excavadora) o faci la feina de centenars de persones en molt poc temps i a menor cost (qualsevol procés industrial mecanitzat) o ens permeti anar a llocs prèviament inaccessibles en hores. Al primer món, hi ha algú avui dia que consideri que aquestes màquines siguin una amenaça per algú o que hagin «desplaçat» l'home de la seva posició preeminent al món? Al contrari, ara (quan ja les hem metabolitzat) ens semblen grans ajudes i totalment imprescindibles. I què fariem sense el nostre estimat automòbil (que apreciem com un amic i que no

sentim pas que ens hagi privat de la nostra humanitat o desplaçat de res sinó tot al contrari)? I com escriuríem si no fos pels nostres editors de text?

Doncs, previsiblement, allò que ha passat sempre tornarà a passar. Tot i que ara pugui semblar mentida, ens acabarem aclimatant als ordinadors que vénen i que tindran la nostra capacitat. I, després d'haver-nos espantat molt, ja no ens semblaran tan amenaçadors. Al contrari. Com el nostre cotxe, els farem anar per on vulguem perquè així augmenti, encara més, la nostra sensació de sentir-nos *dominants* (però per quins set sous devem tenir aquesta necessitat de dominar-ho tot?). I, en el cas més que previsible que els ordinadors d'aleshores vinguin proveïts, de fàbrica, de *consciència* i d'*emocions* i de consciència ètica i tot), hi actuarem com fem amb els nostres animals domèstics. Per anar fent boca, ens podríem proposar l'exercici col·lectiu de començar a veure els futurs ordinadors com a mascotes. Ens faran grans serveis i els trobarem útils però rarament ens perjudicaran gaire més que l'ocasional dòberman incontrolat; més aviat seran com de la família, com ho són el gos i el cotxe. Per aquí no hi ha gran cosa a témer.

El que és segur és que el nostre paper i el dels ordinadors haurà canviat, s'haurà redistribuït. I que canviarà molt també la concepció que tenim del món i, en particular, del que és la informàtica i la intel·ligència. Aquesta serà vista cada vegada més com un procés semblant a la vida, només que infinitament accelerat: els sistemes informàtics ja no seran dissenyats ni «programats» sinó que evolucionaran i s'«autoadaptaran»; i presentaran, com nosaltres mateixos, fenòmens de parasitisme, de virus (d'això ja en tenim), malalties i arquitectures immunodefensives. Així arribarem, tard o d'hora, al punt en què ens adonarem –per fi!– que l'aspecte racional dels humans no és pas tota –ni tan sols la base de– la intel·ligència sinó més aviat, simplement, «la capa més dèbil i superficial del pensament, un procés deliberat molt feble que si és efectiu és perquè es basa en uns processos cognitius sensoriomotors molt més antics, potents i consolidats, si bé normalment inconscients [...] El pensament abstracte és un nou truc de què disposem des de fa segurament no més de 100.000 anys i que encara no dominem». Aquesta reflexió, feta pel divulgador de robòtica Hans Moravec el 1988, està en la línia del que anem descobrint de la mà dels

ELS SISTEMES IN-  
FORMÀTICS JA NO  
SERAN DISSENYATS  
NI «PROGRAMATS»  
SINÓ QUE EVO-  
LUCIONARAN I  
S'«AUTOADAPTARAN»

neuròlegs (Gazzaniga, Damásio) o de lingüistes com ara Searle. La conclusió sembla triple i potser òbvia: (1) la intel·ligència que ens interessa extreure de les màquines ja la n'estem extraient ara, i amb força èxit per cert, i més que ho farem en el futur; (2) delegarem en els ordinadors els aspectes més reproduïbles i rendibles de la nostra intel·ligència (com hem fet amb les màquines i els automòbils, a qui hem delegat plenament la nostra força física i la nostra mobilitat, respectivament) sense que per això ens hàgim de preocupar del fet que, fent-hi aquesta concessió, la màquina agraciada ens suplanti de cap manera; i (3) la «nostra» intel·ligència –o el que en quedi després de la seva externalització en forma d'ajudant mecànic– és encara moltíssim, és una cosa essencialment personal i creativa, i allò que «sembla» – però no és pas necessàriament– «intel·ligència» haurà estat separat conceptualment d'allò que hi és més essencial, així com la part racional haurà estat separada de la *intuitiva* i de la personal. La intel·ligència humana, cada vegada més ben coneguda, és el que és, i no hi ha cap raó per voler imitar-la ni suplantar-la (tret de, com s'ha dit, en aquelles funcions en què fer-ho ens convingui, com hem anat fent i fem amb la nostra potència muscular o locomotora). I com en els casos de la megaproducció industrial, de la vida en grans ciutats o del vehicle privat en propietat, segur que ja tindrem ocasió d'adaptar-nos a les sorpreses (que Déu n'hi do si n'hi haurà!). I qui ha passat per un pont ja té el secret per passar pels següents. Al capdavall, tampoc no ens demanen que patim més de quatre o cinc brutalitats tecnològiques per generació, i això, com demostra la història, pot arribar a ser relativament digerible.

*(Una versió traduïda al castellà i resumida d'aquest text ha estat publicada a La Vanguardia, per encàrrec, aquest estiu.)*

## *A propòsit de conills i meravelles que ningú no s'espera*

**C**harles Lutwidge Dodgson (1832-98) era fill d'un bisbe provincial anglicà del fred nord d'Anglaterra que de petit, a mitjan segle dinou, passava les llargues hores de missa assegut en una enorme cadira de fusta del cor de la catedral plena de baixos relleus mig gastats, on la seva imaginació li feia veure, allà on hi havia un capellà amb una gran sotana, un bonet al cap i un missal a la mà aixecat enlaire fent fila en una processó, una senyora amb grans faldilles i una corona (una reina, potser?) amb una carta enorme –de les de jugar a cartes– i un seguici de cortesans al darrere. Semblava molt amoinada, la reina, i feia com si cridés «que els tallin el cap a tots!» Però tot seguit el noi es despertava i tornava a veure el capellà i el sèquit, tot borrós, en processó. Més endavant aquest noi va fer cap a Oxford, on es va fer capellà i matemàtic i va somiar que escriuria llibres saberuts i definitius sobre la revolucionària nova *Lògica simbòlica* de «mister George Boole» que acabava de descobrir al seu *college*, i fer-ho tot amb el nom elegant, culte i llatí de «Carolus Aloysius Dodgson» o simplement «Aioysius Carolus» o encara, potser, el menys pedant de «**Lewis Carroll**». Escriuria un llibre on explicaria a tothom, en un to planer com si fos en un sermó, que la Lògica, aquesta gran auxiliar de la Ciència, en realitat no té cap mena de contingut. Al contrari, el primer que fa el lògic és desentendre's del que li diu l'interlocutor, traduir-ho a un esquelet formal («simbòlic», com li agradava de dir) i tot seguit comprovar si el raonament que li ha sotmès l'interlocutor perquè l'examini i el diagnòstic és correcte –i és, doncs, utilitzable– o no (i, per tant, és una *fallàcia*,

BUTLLETÍ DE  
L'ACIA  
NÚMERO 29  
TARDOR DE 2003

a rebutjar immediatament). Carroll es va adonar, seguint la idea del matemàtic De Morgan, que el contingut del raonament no tenia res a veure amb la seva validesa, que les paraules podien significar coses diferents de les habituals i, tot i amb això, la validesa del raonament no en quedava afectada gens ni mica. Va pensar fins i tot d'explicar-ho en forma de petit conte per a pàrvuls, on un nen li demanava al personatge rabassut i ovoide de Humpty Dumpty, el ninot que tenien a molts jardins d'infància anglesos (les anomenades *nurseries*) i sobre el qual es feien cançons i amb qui es parlava com si fos de la família – i que originalment havia estat una caricatura afectuosa de l'almirall Nelson, tan llegendari i tan rodó de cos com familiarment popular entre els anglesos. Al conte Carroll s'imaginava que un infant li demanava a Humpty Dumpty (HD) on era el jardí, i el personatge li deia que a l'esquerra tot senyalant-li la dreta; el nen es quedava perplex, i replicava que per què HD deia «esquerra» quan volia dir «dreta», a la qual cosa HD contestava que «dreta» o «esquerra» no volien dir res o, més ben dit, volien dir exactament el que ell (HD) volia que diguessin; que això no tenia cap importància, mentre es fes de manera sistemàtica, vull dir, sense caure en contradiccions. Carroll volia fer entendre amb això el que era una idea bàsica de la Lògica moderna, i és que les paraules en Lògica no han de significar res en particular (al contrari, ho han de poder significar tot) i que l'important no és el que signifiquin sinó que, quan parlem en aquest llenguatge (i hi expliquem, per exemple, tota una història), l'únic que importa no és el que hi passi (a la història) sinó que sigui creïble (que no entri en cap contradicció, que ho arruïnaria tot). Vaja, com a les novel·les (o les pel·lícules), en què una història aparentment llunyana o fins i tot absurda sempre pot voler dir alguna cosa per a algú (només cal que ens la facin creïble o possible). Carroll també havia pensat a il·lustrar el seu llibre de Lògica –que era el projecte de tota la seva vida– amb exemples de raonaments trets de la vida de cada dia i amb figures i jocs que fessin la cosa més agradable i passadora.

Dodgson (o *Carroll*, com vulgueu) era molt modern. No sols era un matemàtic encantat amb la Lògica de la seva època (la de «Mr. Boole»), que segons ell faria que qualsevol raonament fos tan fàcil de resoldre com un càlcul matemàtic (i això que no es podia ni imaginar que després de la revolució booleana vindríem les màqui-

nes i els informàtics), sinó que va ser dels primers a interessar-se per una novetat tècnica tan enlluernadora i «ultramoderna» per l'època com era la «photo-graphy» (=escriptura de llum), la de «mister Fox-Talbot», que permetia revolucionàriament reproduir coses de la realitat fidelment sense haver de saber-les pintar. Carroll, que tot i capellà no deixava d'admirar la bellesa d'una dona jove, s'havia entusiasmat fent «photographies» de nenes maques, com ara les tres filles preadolescents del seu director en cap, un home de gran cultura anomenat Henry George Liddell (el que va fer el primer gran diccionari de grec clàssic en anglès) que dirigia el Christchurch College d'Oxford on vivia i ensenyava Dodgson.

Avui podem apreciar ben bé què va significar per a nosaltres, passat l'any 2000, que un cert cap de setmana del 1865 en què casualment feia bon temps el seu cap li demanés (o manés) que portés les seves filles d'excursió al camp. Dodgson/Carroll, que estava obsedit a aprofitar les esclertes més petites de temps per enllestir el seu fabulós projecte de llibre de Lògica (que no va sortir fins vint-i-un anys més tard) va cedir i acceptar, sempre obsequiós, sense posar-hi objeccions. La més eixerida de les filles aquell dia era la mitjana, l'Àlicia Liddell, d'uns deu anys. Quan, després d'haver dinat sota un arbre, l'Àlicia va veure un forat a terra i va demanar al Rev. Charles què devia ser allò, aquest, mig absent i semicolgat enmig dels seus projectes d'explicació del quid de la Lògica, li va respondre que allò era l'entrada d'un cau de conill. Va ser la primera baula d'una cadena (que ens arriba fins aquí, on som nosaltres): que un cau és «la casa» del conill, que per què el conill –«que acabo de veure entrant al cau» corrents– sempre corre, que doncs perquè deu tenir pressa, i què hi fa un conill en un cau, doncs hi viu, i així també hi invita gent, doncs és clar, deu anar corrents perquè avui deu celebrar alguna cosa, i què hi deu celebrar, avui, si no és cap dia, doncs potser és el seu aniversari, però que estrany que avui precisament sigui el seu aniversari, doncs deu celebrar que avui no és el seu aniversari, cosa que és més probable, i què hauria de fer jo per entrar a la casa i veure-ho, doncs t'hauries de fer petita, i així fins al vespre, tot teixint una realitat que, tot i els disbarats (el *nonsense*), era lògicament coherent i perfectament plausible (sempre que l'oient ho amanis amb una bona dosi d'ironia intel·ligent). Quan el reverend es va



retirar a la seva cambra va pensar «quina manera de perdre el temps, tota una tarda –i el meu llibre aquí, sense progressar!», mentre les nenes, entusiasmades, li deien a son pare que el rev. Charles els havia explicat una història molt divertida i al·lucinada. El cap Liddell l'endemà mateix li va demanar (o manar) que ho escrivís de seguida, que ell ja havia contactat amb un il·lustrador (el cèlebre Tenniel) que hi afegiria els dibuixos. I vet aquí el pobre Carroll entretingut durant dies i dies refent la parida aquella que havia hagut d'improvisar per compromís. I vet aquí que el llibre, que s'havia de dir *Les aventures de l'Àlicia sota terra* (*Alice Under Land*), va acabar amb títol canviat (*in Wonder Land*) i essent potser el primer *best-seller* de la història (centenars de milers exemplars venuts en poc temps). Va entusiasmar tothom, des dels petits fins als lògics (que hi veien «Lògica» pertot arreu), passant pels humoristes i pels lingüistes (que hi veien un món fabulós d'absurd, de *nonsense* –que ell va inventar– i d'inventius jocs de llenguatge) i els poetes i els pintors (els surrealistes el tenien per sant patró) i fins i tot a la reina Victòria que, entusiasmadíssima, li va demanar, un dia que el va voler saludar personalment a Oxford, que li dediqués explícitament i afectuosament el seu pròxim llibre.

«AIXÒ M'HO HA ENVIAT AQUELL CAPELLÀ D'OXFORD TAN ATENT, PERÒ: ON SÓN, ELS DIBUIXOS? I ELS NENS? I L'ÀLÍCIA?...SI NO S'ENTÉN RES!»

Tot això, i l'èxit, no va fer gaire el pes al nostre reverend que, quan a la fi d'any reflexionava en el seu diari sobre aquell 1865 que acabava de passar, hi anotava: «Quin any més eixorc! Enguany, realment, no he fet ben bé res d'útil... I, mentrestant, el meu llibre (de Lògica) encara per escriure!». Pobre, si hagués sabut que avui el coneixem per aquell conte que no volia explicar ni publicar i no pas pel seu famós llibre (que avui ningú no recorda)... Ironies de la vida. Com quan el reverend va acabar el seu llibre (que va titular, com el d'en De Morgan, *Symbolic Logic* i que tenia tantes pàgines i quilos com per matar-te si et queia a sobre), el 1886, i el va enviar, afectuosament dedicat, a la seva reina. Aquesta, estranyadíssima de rebre un llibre així, va comentar al seu vailet de torn: «això m'ho ha enviat aquell capellà d'Oxford tan atent, però: on són, els dibuixos? i els nens? i l'Àlicia?...si no s'entén res!». Curiosa i impensable fi d'una trajectòria vital (com la nostra, la de tots) que, si és alguna cosa, és que és totalment imprevisible i al·lucinatòria. No puc deixar d'imaginar-m'hi John Lennon, un altre anglès dels Midlands practicant del *nonsense* i de l'al·lucinació, segons el qual,

com deia en una de les seves últimes cançons, «la vida és allò que passa mentre tu et penses que te la planifiques». Segur que en Lewis Carroll no hi hauria pas estat menys d'acord que jo (que vaig començar a IBM amb grans plans i he acabat de funcionari).

*Ton Sales. Professor de Lògica (ai!) i de Raonament, a la FIB — la Delegació d'Alumnes la qual és simpàticament la responsable d'haver-me fet escriure aquest article, cosa que li agraeixo.*



## *Quin dia tindrem intel·ligència? I quan ho sabrem, que en tenim?*

**L**A primera d'aquestes dues preguntes volia ser: «Quan tindrem intel·ligència artificial?» Perquè cal suposar que de la natural ja en tenim i que ara toca l'altra. Ara bé, vist com estem fent anar el món darrerament tots plegats, em sembla que de la primera no n'anem gaire sobrats. Ho pensava veient a la TV com persones segurament educades i encantadores es converteixen en soldats i soldades que combaten l'avorriment torturant civils i fent-s'hi fotos. I això que la guerra de l'Iraq havia de ser una operació quirúrgica feta amb bona voluntat. No vull ni pensar què deu passar doncs en una guerra feta sense tan bones intencions. Ho dic perquè recordava la famosa frase del 2002, fa dos estius, de l'actual cap màxim de DARPA (l'antiga ARPA, ja sabeu, la que va crear el protointernet). El jerarca, que per cert és un civil, va explicar que la prioritat actual màxima de la seva agència era aconseguir la «intel·ligència artificial» (IA). I què calia entendre per IA? El senyor ho va explicar de seguida, exactament com ho transcriu: «la IA consisteix en» –i ho va dir i es va quedar tan ample!– «la creació d'un soldat que combinarà l'enorme potència de les màquines amb l'enorme [sic] potència de la intel·ligència humana». Pobres de nosaltres (i dels iraquians), vaig dir jo, quan tot seguit m'hi vaig haver de repensar. Ben mirat, en el fons potser és millor que l'ARPA tingui raó i el seu objectiu –també el nostre?– tan clarament expressat s'acompleixi aviat i bé. En efecte, no m'imagino cap màquina, i encara menys si és intel·ligent, que pensi a fer-se una foto amb un iraquiana vexat i enviar-la als parents o que, en lloc d'aplicar la convenció de Ginebra amb el

BUTLLETÍ DE  
L'ACIA  
NÚMERO 30  
PRIMAVERA DE 2004

rigor cec de què només són capaços les màquines i els buròcrates, hi pugui trobar cap mena de sentit –i encara menys trobar-hi gràcia– al fet de distreure’s molestant iraquians. O sigui que no tinc gaire clar si tenir màquines intel·ligents serà pitjor o millor que no pas ara. Segons com, potser hauríem de confiar en la intel·ligència (i el seny) d’una màquina més que no pas en la d’un humà eventualment trastocat. Per als qui això de construir soldats metàl·lics no els acabi d’abellir sempre tenim un consol, però. En totes aquestes històries de relacions extramatrimonials entre exèrcits i IA, que dura des de sempre, es pot observar la següent llei: sol esdevenir-se que després d’un militar, que parla –com sempre– de fer soldats multifuncionals i indestructibles, ve un metge o un enginyer que transforma o adapta la proposta primitiva en una altra de més èticament digerible, que permeti que tots nosaltres continuem treballant amb les mateixes coses, però amb la consciència tranquil·la. Deixant a banda els benèvolos propòsits de convertir els soldats en guerrers humanitaris (vull dir, participants en les «guerres humanitàries» que fem els països civilitzats avui dia), hi ha dos tipus de reconversions tranquil·litzants que sovintegen molt. Una és canviar el soldat per un bomber que en comptes de matar-nos ens salvi, ens curi i ens aconhorti psicològicament enmig del post-trauma. Una altra és convertir-lo en un infermer-conseller-company-pròtesi personalitzat que ens ajudi a passar amb afecte un mal tràngol com ara el de ser vell o malalt. Això es fa molt, i tots quedem en pau amb nosaltres mateixos i amb tothom; però és que, a més d’humanitari i de sedant de consciències, em temo que sortosament és l’únic recurs que tenim els *IAaires* catalans com a membres d’un país subaltern incapaç de fer ni plantejar cap guerra. (Almenys per aquí ens salvem; com sempre, fem de la impotència virtut.)

«LA IA CONSISTEIX EN» –I HO VA DIR I ES VA QUEDAR TAN AMPLE!– «LA CREACIÓ D’UN SOLDAT QUE COMBINARÀ L’ENORME POTÈNCIA DE LES MÀQUINES AMB L’ENORME [SIC] POTÈNCIA DE LA INTEL·LIGÈNCIA HUMANA»

I pel que fa a la segona pregunta del títol («I quan sabrem que ja la tenim, la IA?»), cal recordar que el 1950 l’Alan Turing, el nostre sant patró i pare putatiu de tots nosaltres, va publicar aquell famós article a *Mind* en què es plantejava aquesta mateixa pregunta i se la responia imaginant-se una situació en què un humà es trobava davant d’una màquina suposada «intel·ligent» i d’un interlocutor humà (també suposat intel·ligent, és clar). Aquests dos últims estaven ocults, de manera que l’humà-jutge es podia

comunicar amb tots dos per mitjà d'una màquina d'escriure i a base d'interaccions havia de deduir qui era l'altre humà i qui era la màquina. Si arribava un moment en què l'observador no podia discriminar entre tots dos, llavors ja podríem dir que havíem aconseguit fer una màquina *intelligent*. Òbviament, per passar la prova (el famós «Test de Turing») calen dos ingredients: (1) la capacitat (i la voluntat) d'imitar la intel·ligència humana, per despistar el jutge, i (2) la capacitat (i la voluntat) d'*enganyar-lo*. El primer implica confondre l'imitació amb la intel·ligència, i per això ha estat molt i ben criticat. El segon (l'engany), que perquè no soni tan malament podem pietosament rebatejar com a *superació d'expectatives* (les de l'observador-jutge), lliga perfectament amb el taló d'Aquil·les que repetidament s'ha retret a la IA, el que deriva de la definició si fa no fa improvisada que va fer Minsky un dia (al 1964) i que jo no m'estic de repetir tot sovint, aquí mateix: «intelligent» és «tot allò que fet per un humà resulta natural i que en canvi fet per una màquina resultaria impropï i, per tant, sorprenent». I això porta, vulguis no vulguis, al *relativisme* del concepte i del criteri classificatori d'intel·ligència: per detectar-la ens hem de *sorprendre* i, és clar, n'hi ha que se sorprenen per qualsevol cosa i n'hi ha que no es deixen sorprendre per res (tan altes són les seves expectatives). Això ho converteix tot en relatiu i patològicament depenent de l'estat d'ànim i expectatives de l'observador-jutge (i, és clar, de la seva competència, que cal suposar-li). Com es pot veure fàcilment, com a criteri científic el test de Turing n'és poc, de científic. (La qual cosa no detreu gens del mèrit d'en Turing, que prou que va fer: l'any 1950, que per a l'*arqueologia de sistemes* representa l'edat de la pedra dels ordinadors, tan sols imaginar-s'ho parlar de màquines intel·ligents i mirar de trobar un criteri de demarcació de la intel·ligència a partir d'això tenia tot el mèrit del món, com tot el que va fer Turing en condicions prou adverses.) I així s'ha dit que, segons el criteri de Turing, les màquines ja són intel·ligents des del 1966, de quan milers de badocs es van convèncer que l'Eliza, el *programet* que amb recursos limitadíssims simulava ser un psiquiatre, era un psiquiatre de veritat. O bé, com diuen els experts en general, Turing ja ha passat la prova amb èxit diverses vegades, sempre que prèviament s'ha delimitat el terreny i s'ha tractat de raonar i/o actuar sobre dominis relativament *tancats*, en què no cal que l'ob-

PEL QUE FA A LA  
IA JA SOM AL CAP  
DEL CARRER, I  
QUE –SEGUINT  
EL CRITERI D’EN  
TURING– JA PODEM  
DIR QUE «PER FI  
TENIM MÀQUINES  
INTEL·LIGENTS»

servador sigui un ignorant ni tampoc cal enganyar-lo: es convenç ell sol i no li queden dubtes de la intel·ligència (=permutabilitat amb un humà) del que té al davant.

És prou clar que, a la velocitat amb què augmenta la potència dels nostres ordinadors (gràcies als Moore i les seves lleis), aviat hi haurà un punt en què la corba monòtonament creixent i probablement exponencial de la «intel·ligència» de les màquines (o d’allò que s’hi confon) arribarà a creuar –o ja l’ha creuat?– la nostra cota d’intel·ligència. Preveient aquest moment i la necessitat de detectar-lo per quedar-nos tranquils (o per espantar-nos), i tot i tenint en compte les clares mancances del test de Turing, aquest es podria reconvertir en un diagnòstic útil de l’instant fatal: es tractaria de triar un observador-jutge molt competent i incrèdul, i a més particularment interessat a evitar i desemmascarar possibles fraus. I, posats a triar, qui millor que un client atret malaltissament per un producte que desitja i que coneix bé i sobre el qual no es deixarà enganyar? Doncs a Wolverhampton, Jim Wightman, informàtic anglès benemèrit, potser ens ha resolt el problema. Ha estat construint el que ell en diu *nanniebots*, uns programes-agents que actuen amb la pietosa i santa idea de detectar pedòfils a la xarxa. Es tracta de programes que «parlen» amb els presumptes pedòfils imitant molt versemblantment la parla del jovent, que aprenen –via pràctiques prèvies en una xarxa neuronal– a partir de la parla d’adolescents reals. Tant, que els pedòfils creuen que ho són (adolescents, vull dir), hi cauen de cap sense sospitar gens ni mica el frau, i concerten les cites-parany. La idea original és caçar-los i que els visiti la policia, però és que, de retruc, hem aconseguit sense voler, potser el primer cas d’èxit rotund de la prova de Turing. Si admetem –amb perdó– que els pedòfils puguin fer aquí el paper de jutge o d’observador científic, tenim aquí potser una constatació objectiva: que pel que fa a la IA ja som al cap del carrer, i que –seguint el criteri d’en Turing– ja podem dir que «per fi tenim màquines intel·ligents» –malgrat que hàgim d’agrair la constatació, ai !, a pedòfils de col·laboració involuntària. I és que el món és ple de paradoxes: perquè cap soldat no torturi cap civil en un país estranger haurem d’haver arribat a la IA gràcies a l’Internet i la pedofília... ves quines coses!

## *És perillós profetitzar, sobretot quan es tracta del futur*

**S**EMPRE he pensat que **imaginar el futur** és un esport d'alt risc. Quanta gent normal (entre la qual compto, si m'ho permeteu, els escriptors de **ciència-ficció**) hauria mai previst que un dia veuríem persones pel carrer parlant-se a si mateixos com insensats i gesticulant privadament amb una cosa de plàstic metal·litzat enganxada a l'orella? M'hi jugo el que vulgui en Miquel Barceló que les obres en què hi surt una cosa remotament semblant es poden comptar amb els pèls d'una granota (com irònicament deia James Cagney en un clàssic americà a propòsit de l'honradesa d'una senyora). I ves si hem acabat caient-hi tots, en aquesta càotica pantomima col·lectiva en què hem convertit els nostres carrers. Jo tinc una filla tan avesada a «textejar» sobre la maquineta que l'altre dia li vaig sentir dir que una certa pel·lícula dels anys noranta era increïblement poc realista, perquè «no hi sortia cap **mòbil**» (i els actors havien de quedar prèviament en un lloc i una hora per telefonar-se, cosa que es veu clarament que és inútil i totalment inversemblant). En fi, que una vegada ens acostumem a una cosa, és com si l'haguéssim tingut amb nosaltres tota la vida. I si no pregunteu-ho als baleners de Nova Anglaterra, capità Ahab inclòs, que es van oposar amb dents a l'electricitat i les bombetes d'Edison perquè els suposaria, i els va suposar, la ruïna més sense palliatius possibles; o als copistes que van veure escombrada d'un buf la seva refinada feina amb la «premsa de vi» del senyor Gutenberg que, deien esverats, «escrivia els llibres ella sola», sense intervenció humana, miraculosament (*escriptura artificial*, en deien, us sona?). Per no parlar dels pintors que als anys

BUTLLETÍ DE  
L'ACIA  
NÚMERO 31  
TARDOR DE 2004



1840 es van quedar sense feina amb la *repulsioa* i *inhumana* –deien– *nova escriptura de llum* o *photo-graphy* del senyor Fox-Talbot, que –on s’era vist!– reproduïa escenes sense que ningú les hagués de pintar. Ni de nosaltres mateixos. Us imagineu què faria jo sense un editor per escriure aquest article? Doncs em veuria desesperat, incapacitat irreversiblement per tornar a la màquina d’escriure aquella que s’encallava cada quatre lletres, i al món llunyà i prehistòric de l’original-i-la-còpia (de carbó), un món sense fotocopiadores i ni tan sols «típlex». O què fariem tots nosaltres sense el cotxe?

Bé, alguns potser –m’hi compto– podríem sobreviure, però una gran majoria es pansiria de tristesa amb un tal greu atemptat a la seva llibertat personal tan heroicament aconseguida. Doncs ara imagineu-vos què fariem tots nosaltres, tots sense excepció, sense electricitat a casa...? Descobriríem tot de cop que no sabem fer res, que ens havíem convertit en uns inútils integrals, malaltissament addictes al corrent elèctric i a Fecsa-Endesa. Avui, que ens treguin l’electricitat de casa és deixar-nos sense sang i despullats. I qui havia previst cap d’aquestes coses? Va haver-hi algú que, cinquanta anys abans del fet, hagués pogut pensar que es podria escriure sense escriure, reproduir sense pintar, fer llum sense oli d’«espelma» (és a dir, d’«esperma») de balena, desacostumar-se a la sala de concerts o de teatre i sentir en canvi música a casa nostra mateix o anar a emocionar-se veient ombres que es mouen en un llençol penjat a la paret d’una sala sense un sol actor? I, sobretot, qui hauria pogut imaginar les (enormes) conseqüències que tot això ha tingut i té sobre la nostra vida i sobre nosaltres, que vivim en un món en què tot error és reversible, tot és corregible, ningú tolera equivocacions i ningú no té la paciència d’aguantar una cua o una estona sense parlar amb algú, encara que sigui al teu propi telèfon i per a no dir res? Ho hauríem dit, uns anys abans de cada cosa? Ho hauria pogut dir un professional de la imaginació com ho són els escriptors predictors/reveladors/creadors de futur? Doncs no, rotundament.

Només cal veure l’imaginatiu Jules Verne. Va pensar, sí, en un fax *avant la lettre* que copiava, és clar, facsímils, però –ah renoi!– fets de lletres encadenades, escrites a mà, amb tinta. Lletres separades, generades per un mecanisme? No, home, a qui se li ocorre haver d’escriure amb lletres de motlle, com si fóssim tan pobres d’esperit

VA HAVER-HI ALGÚ  
QUE, CINQUAN-  
TA ANYS ABANS  
DEL FET, HAGUÉS  
POGUT PENSAR  
QUE ES PODRIA  
ESCRIURE SENSE  
ESCRIURE, REPRO-  
DUIR SENSE PINTAR,  
FER LLUM SENSE  
OLI D’«ESPELMA»  
DE BALENA,  
DESACOSTUMAR-SE  
A LA SALA DE  
CONCERTS O DE  
TEATRE I SENTIR  
EN CANVI MÚSICA  
A CASA NOSTRA  
MATEIX O ANAR A  
EMOCIONAR-SE VE-  
IENT OMBRES QUE  
ES MOUEN EN UN  
LLENÇOL PENJAT  
A LA PARET D’UNA  
SALA SENSE UN SOL  
ACTOR?

com una impremta? I és que fins i tot Verne, tan imaginatiu, tenia límits posats a la imaginació. I potser no els hauria tingut tant si hagués vist el *maquinam* complicadíssim, impràctic i antiestètic (una mena d'*invent del TBO*) que aquells mateixos anys construïa amb penes i treballs un americà que es deia Sholes i que pretenia inventar una pretensiosa «màquina d'escriure» (una *type writer*, literalment *màquina d'escriure amb lletres d'impremta*) per no haver d'escriure, una mena de teranyina de palanques que martellejaven immerescudament un sofert paper (i les oïdes del pobre operador) en el qual deixava impresa una taca de lletres de motlle més o menys allargada en el sentit horitzontal (això suposant que la tinta taqués el paper sense foradar-lo). La màquina funcionava quan volia, però la companyia Remington, que després de la guerra civil ja no venia armes i s'esllanguia sense saber què fabricar, va atrevir-se a construir la cosa i fins i tot vendre-la. És clar que si Verne l'hagués vist potser hauria imaginat el seu fax per reproduir missatges treballant sobre un input de tipus d'impremta discrets i no de lletres enganxades cal·ligràficament. El que no és clar és que Verne hagués pogut imaginar que aquella «màquina d'escriure» d'en Sholes, tan *cutre*, es faria general, primer a les oficines (de la mà pionera, com sempre, de les dones, ara «mecanògrafes») i més tard a tot arreu. I tampoc no és gaire probable que ho pensés el mateix Sholes inventor, a qui no se li va poder ocórrer altra cosa que posar-hi, com a dispositiu d'entrada a la seva *type-writer*, un teclat. Com sona. Un teclat de piano, en què cada lletra tenia una tecla «blanca» i els signes estranys una tecla negra, com de bemoll. I és que Sholes no havia vist mai un «teclat» de les nostres «tecles», rodones i còncaues, i sí les blanques-i-negres i allargades del piano o l'acordió de l'època. Doncs sí, tot i la imaginació i futurisme que se li suposi, tothom té límits a l'hora d'imaginar futurs: com pot algú imaginar una cosa que no ha vist mai ni s'assembla al que ell veu cada dia ni n'és remotament extrapolable? Com pot donar crèdit a una cosa insignificant que està passant davant dels seus ulls (com als 1890) i creure que d'allà sortirà, inopinadament i sorpresivament, tot un futur, **el** futur? Verne no es va saber imaginar un reproductor d'escrits que no fos amb lletres humanes, com Sholes no va poder pensar en un tauler de lletres a tocar que no fos com el d'un piano amb notes, o com ningú va pensar el 1901

que al segle que començava es farien guerres per ràdio o TV, que aniríem a tot arreu tancats dins d'una closca amb quatre rodes o que veuríem gent parlant sola pel carrer imitant un boig.

Totes les consideracions anteriors em servien d'excusa l'altre dia en què, com és de témer en aquest ofici, un periodista em va demanar, a mi, **com a informàtic** (bufa!), què pensava sobre el futur que ens esperava a propòsit d'una notícia que acabava de sortir als diaris (la del Pod, que us explico més avall). Vaja! Anar a trobar un informàtic (precisament un informàtic?) perquè t'expliqui el futur? (el futur?). Què devem tenir els **informàtics** que tothom es veu amb cor de demanar-nos sempre això a nosaltres preferentment com si fos evident que veiem el futur millor que, posem per cas, els venedors de síndries (o de cogombres o d'hipoteques)?, vaig pensar. Es tractava, pel que vaig veure, d'anar seguint el joc del periodista, que ja tenia la idea feta del que passaria (en el futur). Jo havia merament de justificar o il·lustrar –des de la meva suposada professionalitat i sapiència– els seus preconcebuts (prejudicis). (En fi, el *déjà vu*: el científic convocat per abonar i adornar «objectivament» amb llenguatge tècnic i/o opac la decisió que ja han pres d'altres.) La notícia-excusa era que la **Toyota** presentava a Europa el seu model futurista de cotxe batejat **Pod** (no em feu dir per què; és un acrònim, com sempre) que estava pensat perquè, com a **agent** (en el sentit nostre, de la **IA**) autònom i amb iniciativa, interactués amb l'amo del cotxe com si fos un company; que, per exemple, s'adonés (abans que la consellera Tura) que el conductor anava pujat d'alcoholèmia i que intentés compensar aquesta delicada situació primer exposant-li raonadament el cas i després, si es feia l'orni, tirant-li una boira de microgotetes fredes a la cara per aclarir-li les idees i finalment, si no se'n sortia, recorrent a mesures dràstiques com eixordar-lo amb els seus propis bafles, abaixar-li l'aire condicionat perquè se li gelés una mica el temperament o, com a cas ja extrem, aturar-li el cotxe (és a dir, aturar-se ell) o potser desviar el trajecte per dur-lo a un hospital o directament al llit o a cal psiquiatre. Tot això feia riure qui-sap-com el meu periodista (ho trobava propi «de japonesos» o de minusvàlids mentals), mentre jo pensava, ja no com a **informàtic** sinó merament com a bona persona, que, vista la d'accidents que hi ha, potser tota aquesta cosa del Toyota que aconsella i remulla no era el disbarat tan gros que a

ell li semblava. I vaig provar de dir-li-ho, però, ca!, l'home anava sobrerevolucionat. El que el feia trencar de riure era que la notícia afegia que la Toyota havia aprofitat l'antena del cotxe per convertir-la en una eina d'expressió: vull dir, que quan el cotxe veia l'amo movia l'antena, com un gos. L'home no se'n sabia avenir. «Com són aquests japonesos, renoi», anava dient. I jo, que sóc escèptic i descregut de mena, m'imaginava aquell mateix periodista d'aquí a uns quants anys tenint un Toyota Pod o equivalent i fent l'idiota, en les seves mateixes paraules, dient per exemple «Pod, quina illusió em fa saber que estàs content de veure'm». Perquè ara pot semblar ridícul, o potser propi de suposats americans o japonesos infantilitzats, però: qui és capaç d'assegurar que ell mateix mai no reaccionarà així? Jo, que no em tinc ni per japonès ni per un ésser especialment ridícul o sentimental, comprenc perfectament que, un matí d'aquests en què tot sembla que surt malament, vegis el teu cotxe, aquest et saludi i tu pensis: «coi, això és ridícul, però me n'alegro que algú estigui content de veure'm». Qui pot jurar que no li passarà? I llavors, és clar, ja no ens farà riure, sobretot a nosaltres, com ja no ens fa riure algú gesticulant pel carrer amb el mòbil enganxat, especialment quan aquest algú som nosaltres mateixos.

Jo no sé –ni ningú, en podeu estar segurs– què ens espera d'aquí a uns anys, com viurem i quines emocions sentirem. Només puc extrapol·lar a partir del que tenim. No tenim una exigència de gratificació instantània, el plat cada dia a taula i comoditats a l'abast? I no tenim cotxes i màquines i ordinadors personals amb el quals ens comuniquem de manera prou satisfactòria com si ells fossin realment companys nostres? Doncs el més previsible és que això s'intensificarà i s'estendrà –tant de bo!– a tota la humanitat en general, juntament amb la irresistible tendència a la comoditat, a la nostra *fluixera* de caràcter, a la nostra estupidesa general i democratitzada, a la nostra tirada a menjar-nos literalment el paisatge, o a la nostra demostrada ineptitud a tenir el planeta net i ben conservat. Això, segur. Ara bé, serà això el futur? Doncs no en tenim ni idea. Qui ens pot assegurar que no hi haurà un **tall** tan gros com, per exemple, el que hi va haver ara fa 115 anys, a la dècada dels 1890? N'hi ha prou a recordar que tres petites coses d'aquesta època, tres, aparentment menors i continuistes, van suposar que l'abans i el

SI JO ARA DIGUÉS:  
«RES, EL FUTUR  
SERÀ...» I AQUÍ HI  
POSÉS QUALSEVOL  
DE LES FUTURO-  
LOGIES EN ÚS QUE  
HI HA ALS DIARIS  
I A LES REVIS-  
TES, SEGURAMENT  
M'EQUIVOCARIA,  
I MOLT. NO CAL  
SER HUMIL PER  
ADONAR-SE QUE  
LES SORPRESES HI  
SÓN I HI SERAN.

després acabessin tallats per un cisma **irreversible**: gairebé als mateixos anys va aparèixer el cotxe sense cavalls, el cinema i la ràdio, com a coses separades i aparentment incrementals i diferencials, i tots tres pràcticament alhora. Des d'aquest moment ja no hi va haver retorn possible. Vivim en ciutats que no tenen res a veure amb les anteriors, que estan concebudes per i per a l'**automòbil**, i no cal recordar on anem a comprar o com anem als llocs, o què tenim tots al cap generalment. I gràcies al tàndem ràdio/cinema, finalment integrats i introduïts a les cases amb el nom de «TV», ens hem acostumat a rebre la informació, quantiosíssima, per via electromagnètica (amb unes «ones» invisibles que fins al 1887 i gràcies a Heinrich Hertz no se sabia ni que existissin) i audiovisual (en un format que ens ha canviat fins i tot la nostra manera d'imaginar i de pensar). Vivim en un món que hauria semblat veritablement **màgic** (recordeu Clarke: «Qualsevol tecnologia prou desenvolupada és indistingible de la màgia [per a qui pertany a una tecnologia anterior]») a qualsevol persona que s'hagués mort precisament el 1890: anem encapsulats com fetus o centaures dins d'un cotxe –i també pel carrer– totalment aïllats, escoltant música o sentint allò que es diu o està passant en aquell mateix moment a l'altre cap de món. Si jo ara digués: «res, el futur serà...» i aquí hi posés qualsevol de les futurologies en ús que hi ha als diaris i a les revistes, segurament m'equivocaria, i molt. No cal ser humil per adonar-se que les sorpreses hi són i hi seran. Sorpreses sorprenents (valgui la redundància). Sorpreses com les que tindria l'hipotètic iaio del 1890 que ens visités ara, o les que vam tenir tots quan al cap d'un segle just –i passades dues guerres, i les atrocitats impensables que en pengen (que segur que no havien de passar i que ningú no havia previst)– tots ens vam quedar de pasta de moniato veient el 1990 que la gran novetat del segle XX, la cosa soviètica, es fonia en un tres-i-no-res, com es dissol un terròs de sucre al cafè amb llet... per no parlar del segle XXI, un segle que havia de ser de pau, modernitat i democràcia i que va començar ben bé a l'inrevés: uns avions ultramoderns van aterrar edificis també ultramoderns gràcies a uns homes aparentment racionals i educats, i inesperadament suïcides, infosos d'una concepció absoluta i ultramontana de la religió que ens retorna a quatre-cents o cinc-cents anys abans. Ah, i això no és tot. L'endemà, en comptes de veure els nostres

dirigents destil·lant racionalitat i pau, vam haver de veure com dos antics amics, l'Osama i en George, partien peres davant nostre i es declaraven la guerra, sense consultar-nos, en nom d'un mateix Déu. Així, començava el segle nou amb un vell retorn a les guerres i els déus tribals. Qui és el savi que hauria pogut predir un u per cent de tot això quan un periodista se li hagués acostat? Doncs ara, en ple 2004, qui és el maco que gosarà explicar què ens passarà? Ah, i ser informàtic i haver tingut intimitat amb ordinadors no ens dóna cap excusa ni patent especial, no ho oblidéssim.

Per això, quan m'ho demanen, contesto amb la frase que encapçalava aquest article i que va dir Samuel Goldwyn el 1938 (també va dir una cosa semblant Bohr, però no volia fer gràcia) quan preparava *Allò que el vent s'endugué* i algun periodista li va demanar les seves previsions sobre l'èxit de la pel·lícula. Goldwyn, que era un immigrant autodidacte, va enunciar la cèlebre frase, amb la mateixa gran propensió a la relliscada semàntica que exhibia aquells mateixos anys en Joan Pich i Pon, el nostre primer batlle de Barcelona amb gràcia (un meritori precedent d'en Joan Clos, l'alcalde que mata mosques «a cops d'elefant», segons diu)(o eren «elefants a cop de mosquits»?). Segur del que deia, Goldwyn ens va regalar la impagable frase de la capçalera, que jo m'he apropiat i dic tan sovint com puc. I goso fer servir *profetitzar i futur* alhora, com abans he dit *sorpreses sorprenents*, i solc afegir, com feia en Pich, *valgui la repugnància!*. El que no va fer aquest home com a alcalde ho ha acabat fent com a humorista.

AH, I SER INFORMÀ-  
TIC I HAVER TINGUT  
INTIMITAT AMB OR-  
DINADORS NO ENS  
DÓNA CAP EXCUSA  
NI PATENT ESPECI-  
AL, NO HO OBLIDÉS-  
SIM.



## *Hi ha una sola intel·ligència?*

**H**AIG d'avisar que la pregunta que faig no va de filosofia i menys de teologia. Només vol parlar, modestament, de la nostra feina (la IA) i del seu objectiu: fer màquines [que siguin/semblin?] intel·ligents. I, si les hem de fer intel·ligents vol dir que les hem de dotar concretament d'intel·ligència. Ah, conxo! És una cosa, això? És un una propietat que tenen certes coses? És un element d'un conjunt? (és a dir, n'hi ha més d'una, diverses intel·ligències?) ¿Com reconeixem que l'hem trobat o que, si més no, hi anem de camí? I com ens aprofitarà la digestió? Ja es veu que inevitablement això va de futuribles, i la vegada passada ja em vaig mirar de curar en salut explicant que això de fer profecies és relliscós, i hi podem prendre mal.

Per això, continuant amb els meus rotllos, avui volia mirar de delimitar, tant com jo pugui (que és poquíssim), com podem imaginar que serà la intel·ligència, la nostra i la de les màquines intel·ligents que segur que anirem creant. Ja sé que això m'obliga a desbarrar. Al darrer butlletí ja vaig explicar per què no em volia embarcar en el procel·lós oceà de les perillósíssimes prediccions de futur. No sé si va quedar clar que, si no es vol fer riure, això de predir val més no tocar-ho. Sempre podem dir, és clar, una cosa tan òbvia com aquesta: «als pròxims anys, un cop dominat –gràcies a les comunicacions rapidíssimes– l'espai i el temps, els nostres ciutadans passaran la vida plens de comoditats però angoixats per l'excés d'informació». Aquesta és una predicció tan i tan segura que la podria subscriure tothom ara mateix sense problema. Tan segura i universal, de fet, que no deu tenir gaire valor perquè ja la va fer, així exactament, l'any 1900 Sir William Hunter, president de

BUTLLETÍ DE  
L'ACIA  
NÚMERO 32  
HIVERN DE 2005

«ALS PRÒXIMS ANYS,  
UN COP DOMINAT  
–GRÀCIES A LES  
COMUNICACIONS  
RAPIDÍSSIMES– L'ES-  
PAI I EL TEMPS, ELS  
NOSTRES CIUTA-  
DANS PASSARAN  
LA VIDA PLENS DE  
COMODITATS PERÒ  
ANGOIXATS PER  
L'EXCÉS D'INFOR-  
MACIÓ».



la BAAS, un assenyat i docte victorià que veia claríssim que el segle que venia, el vint, seria de bojós. I és que, una vegada conquerit «l'espai» amb els vertiginosos ferrocarril i vaixell de vapor i «el temps» amb la transmissió instantània del telègraf, el bon home –com el seu contemporani el metge Max Nordau (especialitzat en «degeneracions»)– s'imaginava per tot arreu ciutadans aviciats i empatsats d'informació, després d'haver llegit, és clar, tants diaris, de viatjar incansablement pel territori i de rebre «cables» a tothora (i, fins i tot, parlar a distància, pel *telèfanu*, com se'n deia aquí). La recepta és tan bona que passa fronteres i avui encara és aplicable. Com que avui, paradoxalment, llegir diaris, viatjar amb tren o vaixell i parlar per telèfon són precisament activitats més aviat relaxants i no gaire enneuradores, per mantenir la profecia n'hi hauria prou canviant diaris per canals (de TV), el tren i el vapor per l'autopista i l'avió, i el telègraf per l'Internet (o, si hom és onanista, per la consola de videojocs); així l'angoixa predita continua de la mateixa magnitud, i la cosa es manté aproximadament invariànt malgrat el pas dels segles. Segons com ho mirem, hem potser perdut cent anys per anar –això sí, molt més ràpid– al mateix lloc, i el resultat, avui com ahir, és l'enneurament perpetu («córrer per no anar enlloc», com ho definia lúcidament la reina vermella del clarivident Lewis Carroll que al cel sigui). Sort que durant aquest segle vint hem inventat els psiquiatres (i els argentins) i hem passat de Nosferatu a Groucho Marx i a Woody Allen. Ja ens va advertir el científic Peter Medawar que cal tenir expectatives, però que només els rucs fan prediccions. Si no, recordeu que, com va dir el seu compatriota i col·lega Haldane a propòsit de la imprevisible història de la ciència (que és com el Barça, el Madrid i el futbol en general: una muntanya russa): «qui havia d'imaginar que d'una mateixa cosa, l'*argent viu* (o *mercur*' com se'n va acabar dient, aquell «metall líquid» dissolvent conegut dels miners), se'n seguiria el concepte de pressió atmosfèrica (i tota la Meteorologia sencera) i alhora, i estranyament, també el control de la febre dels malalts i el futur de mitja Medicina?» (cito lliurement), i, posat a afegir, hi afegeixo que «[del peculiar comportament d'aquest metall] també se'n seguiria l'observació feta a Anglaterra, a partir de la lectura de la columna de mercuri, que el bombament d'aigua dels pous de les mines de carbó no es pot fer i, (com diria el

nostre alcalde Pich i Pon, «és que, a més de no poder-se fer, és impossible») quan el desnivell passa de 30 metres», observació que, com se sap, és a l'origen del concepte de la «màquina de vapor» de Newcomen i de Watt). ¿Qui hauria pogut sospitar una cosa així, ni en el més imaginatiu dels estats etílics? O bé, com explicava (qui, si no?) el mai no prou lloat Von Neumann: «¿qui hauria pogut mai endevinar que la solució del greu problema dels naufragis de vaixells a l'Atlàntic es resoldria no a base de millorar els enormes giroscopis d'últimíssima tecnologia que portaven incorporats sinó a partir dels estudis que feia Hertz el 1888 al seu laboratori intentant casolanament –amb miralls d'armari!– que s'hi reflectissin com si fossin la llum les «ones elèctriques en l'aire» que ell mateix s'havia inventat del no-res» inspirat per la lectura de les equacions de Maxwell (ones de les quals miraculosament han sortit el SOS., la ràdio –i la tele, i el mòbil– i tots els nostres *gadgets* per fer-ho tot sense fer res)? (torno a citar libèrrimament). Als mateixos científics que l'any 1895 declaraven, com va fer Lord Kelvin, que el vol amb «avions» era físicament impossible, o que la Física estava en un estat terminal de crisi del qual ja no es podria recuperar mai («no es dediqui a la Física, jove, no té cap futur» li deien a Planck) els va costar penes i treballs de reconèixer que els (molt imaginativament) anomenats «raigs X» de Röntgen no eren pas una estafa de la magnitud de l'aigua miraculosa, com assegurava un cèlebre físic, ni els «raigs Becquerel» –és a dir, la radioactivitat (llavors cadascú tenia els seus propis raigs)– no eren un truc de circ. I ho dic perquè aquestes dues descobertes són la mare de tota la nostra Física actual (i de tots els hiroshimes) i perquè van ser fetes l'any 1895, just el mateix en què naixien tot de cop *no-res-menys* que l'automòbil, el cinema i la ràdio (i els «còmics!»), uns invents que m'estalvio de dir com van canviar com un mitjó el segle que venia.

Bé, una vegada posat en guàrdia novament contra els perills de fer veure que «predius» sobre el futur, em dispo a fer el ridícul provant de fer «prediccions raonables» (notable i descarada *contradictio in terminis*). Si creiem, com a bons creients i practicants de «la fe» (la IA, vull dir), que el futur anirà ple de màquines intel·ligents, ¿podem imaginar com seran, aquestes, o com haurem de redefinir el concepte d'«intelligència» després d'això? No és

«SEGONS COM HO MIREM, HEM POTSER PERDUT CENT ANYS PER ANAR –AIXÒ SÍ, MOLT MÉS RÀPID– AL MATEIX LLOC, I EL RESULTAT, AVUI COM AHIR, ÉS L'ENNEURAMENT PERPETU («CÓRRER PER NO ANAR ENLLOC», COM HO DEFINIA LÚCIDAMENT LA REINA VERMELLA DEL CLARIVIDENT LEWIS CARROLL, QUE AL CEL SIGUI)»

pas, tampoc, que el concepte el tinguem gaire clar ara mateix. Si mirem el diccionari, ens diu que intel·ligència és la «capacitat» –alguns afegeixen «mental»– «d’aprendre i d’entendre». Molt bé. No és gaire científic ni precís, però ja ens entenem. També podríem afegir-hi tot seguit «i, a partir d’això (d’aprendre i d’entendre), d’adaptar-se amb èxit a l’entorn i als canvis», si no fos que ara, amb això de l’adaptació, ja trepitgem un terreny massa vague i tou. En fi, de moment hem de continuar acontentant-nos a reconèixer una màquina com a «intel·ligent» simplement perquè ens ho sembla; no tenim cap definició ni cap regle de mesurar, desgraciadament. Ara bé, tot i que, previsiblement, reconeguem a satisfacció la intel·ligència de certes màquines, vol pas dir això que totes hagin de tenir la mateixa mena d’intel·ligència? Vull dir: les persones bé en tenim una bona pila, de varietats d’intel·ligència: unes són intuïtives, les altres emocionals, unes terceres analítiques o comunicatives o sintètiques o visuals o *holístiques* o geomètriques o autistes o mecàniques o pràctiques o manuals, etc. Això és un fenomen normal; se’n diu *divisió de treball*, i de la conjuminació de totes aquestes peces del *puzzle* se’n diu societat. Potser no tindria cap interès per a ningú que tothom fos consumidor d’un mateix tipus o patró d’intel·ligència: a més d’avorridíssim, fóra improductiu i estèril. Per tant, es pot assegurar amb un mínim de credibilitat –és a dir, poca– que les màquines que engendrarem seran, com nosaltres mateixos (o com els ximpanzés de Tanzània, o el nostres animals domèstics), un zoològic de personalitats i sensibilitats diverses, amb les típiques sortides de mare i extrems de la campana (la de Gauss, vull dir): l’escadusser individu impresentable, el savi, el ric, el miserable, l’obsés, el que se’n va d’osques (o que se’n va de l’olla), etc. Com sempre i tothora, com ara i aquí, vaja.

Així doncs, res de nou. Només cal obrir el corral o l’armari perquè n’hi capiguem més, i la vida encara serà més apassionant (o irritant) que ara. El que passa és que aquí hi ha dues coses que s’hi afegeixen (o s’hi adhereixen), imperceptiblement i sense voler. La primera és aquesta: com que aquestes màquines intel·ligents seran diferents de nosaltres (com tots nosaltres som diferents entre nosaltres) però, això sí, semblants a nosaltres (perquè tots serem intel·ligents, i cooperarem –perquè per això les hem fet– i ens hi comunicarem com fem entre nosaltres), la primera sensació que

es desperta és: així doncs, ja no estarem sols? Seran ben bé com nosaltres les màquines que farem? I, la pregunta més inquietant: ens substituiran, doncs? Aquesta última, no se sap si pregunta o amenaça, sí que va molt més enllà del que és raonable i del que hem fet sempre fins ara. Com jo mateix he dit sovint i no em canso de repetir: oi que tot i potser haver començat espantant-nos hem acabat paint bé i sense traumes els antics perills de «substitució» i, així doncs, avui ja ningú veu amenaçant que els cotxes corrin més que nosaltres, que els avions volin o que les màquines i motors tinguin molta més força de la que puguem tenir mai? (I, per cert, tots ells ens superen mitjançant mecanismes molt diferents dels «naturals»: la roda no té res a veure amb una cama, un avió no té res a veure amb un ocell i un mecanisme de potència no s'assembla de gaire res als músculs.) Doncs probablement prendre-s'ho com una amenaça està bé per espantar els nens o fer articles de diari, però totalment fora de lloc per tenir una vida social descripada i sana i encarar responsablement la IA. La segona cosa que s'adhereix insensiblement al fet de tenir màquines intel·ligents és aquesta altra: quina imatge ens en fem, d'aquestes màquines? Han de ser (o les hem de veure com a) col·laboradores? (sembla que sí, oi? per això les hem fet) I companyes? (això ja depèn de cadascú: no hi ha gent que parla al seu gat, i fins i tot al seu cotxe, com si fos de la família?) O potser enemigues? (això hi serà segur: sempre hi ha gent propinqua a l'odi sideral i disposada a competir i/o emparar-se amb qui sigui, incloent-hi màquines com ara la (pobra) pantalla del PC, o el cotxe, o la rentadora del veí).

El més curiós del fet és que no sols «veiem» les màquines d'una certa manera, sempre –inevitablement– antropomòrfica, sinó que nosaltres mateixos provem de comprendre i «veiem» la nostra pròpia intel·ligència –que no ens és directament observable i analitzable– com si hi veiéssim una d'aquestes màquines intel·ligents. (Per fer-ho normalment agafem com a paradigma la més complexa o sofisticada –o misteriosa, o incomprendible– de les màquines que tenim en cada època.) Així no és estrany que les metàfores de la ment hagin anant canviant i que primer (al segle XIV) la intel·ligència humana es veiés «com un rellotge» (!) (per l'efecte enorme que feia que un mecanisme simple «penetrés», concretés i manifestés una magnitud com el temps, tan abstracta

i inaprehensible com la mateixa «intelligència»), més tard (segle XVI) «com una impremta» (!) (perquè la planxa d'impressió «contenia» resumida –també una mica com la nostra intelligència– tota la pacient i creativa feina d'un il·lustrat i experimentat copista i ho escrivia tot ben acabat automàticament, «artificialment»[sic], i a gran escala), encara més tard (segle XVIII) com un mecanisme complexíssim i coordinat com els «autòmats» que feia en Vaucanson i que van veure a Suïssa els lletraferits turistes anglesos que s'hi van inspirar entusiasmats el 1816 per inventar-se en Dràcula i el Dr. Frankenstein tot en un mateix dia. I la cosa no acaba aquí: al segle XIX el cervell es veia com una extraordinària màquina de vapor (què, si no?) amb els seus reguladors automàtics que ho fan anar tot ben suau (clar que, per a això, també s'hauria pogut trobar un símil més senzill com és ara el WC, que també té el seu regulador, però aquesta atrevida analogia potser hauria ofès algun pensador políticament correcte). Al segle XX les metàfores han estat dues, segons la meitat del segle de què es tractés: una d'innocent i inofensiva que consisteix a comparar el cervell amb una de les complexíssimes centrals telefòniques megaurbanes, amb el seu imprevisible i molt múltiple i paral·lel flux d'informació. Però també hi ha l'altra, una imatge datada concretament el febrer del 1946 (i que devem als periodistes i a una broma imprudent de Von Neumann), que sembla clarivident i òbvia però que ens ha fet molt de mal sobretot als informàtics de la IA (per no parlar del mal que els ha fet i encara fa als psicòlegs, filòsofs i gent de professió indeterminada), que l'han incorporat sense entendre-la gaire com si fos el seu propi principi inspirador i metodològic (amb gran detriment i desastre de les corresponents disciplines). Consisteix aquesta metàfora, només disponible enllà dels anys cinquanta i d'èxit fulgurant i rabiós, a admirar-se de la potència dels grans ordinadors, de la seva manera de treballar, de la diferenciació conceptual entre la màquina pròpiament i el «software» (una broma de llenguatge que va fer l'Ed Berkeley i que va causar furor en totes les llengües) i acabar declarant solemnement –com no podia ser d'una altra manera– que «el cervell és [com] un ordinador».

«NO NOMÉS «VEIEM» LES MÀQUINES D'UNA CERTA MANERA, SINÓ QUE NOSALTRES MATEIXOS PROVEM DE COMPRENDRE I «VEIEM» LA NOSTRA PRÒPIA INTEL·LIGÈNCIA –QUE NO ENS ÉS DIRECTAMENT OBSERVABLE I ANALITZABLE– COM SI HI VEIÉSSIM UNA D'AQUESTES MÀQUINES INTEL·LIGENTS»

Aquesta metàfora, que ha passat de símil de treball a dogma de fe (i que és la base efectiva no sols de qualsevol carrera d'Informàtica –de 1960 endavant– en general sinó en particular de l'*Artificial*

*Intelligence Project* original, fundat amb ajuda militar l'estiu del 1956 a Dartmouth per McCarthy i el seu exèrcit de sequaços) ha estat especialment nociva perquè tot de psicòlegs han deixat de fer la seva feina pròpia per analitzar on és el *hardware* i on és el *software* del cervell de cadascú, i com els comportaments individuals o socials són «procediments efectius» i altres despropòsits del mateix estil. És clar que encara que creguem que realment el cervell (o la natura, o el que sigui) és un ordinador i que l'evolució és un algorisme (o el que sigui) més o menys executable per algú (qui?) i que qualsevol cosa (incloent la paret de casa meva, segons el cèlebre símil d'un famós informàtic/filòsof) és un ordinador –o, més laxament, és «com un ordinador»– la nefasta creença no només ha aturat uns quants anys el progrés d'aquestes matèries sinó que també ha donat lloc a dues menes d'excessos fàcilment previsibles: l'un, el de capgirar i reforçar la fletxa i dir que el cervell no és com sinó que és un ordinador, així, ras i curt, amb totes les conseqüències (d'això se n'ha dit «IA forta» i encara hi ha qui es passa la vida discutint-ho). L'altre excés consisteix a fer una translació mecànica de la dualitat *hardware-software* a tot arreu com si aquesta dualitat fos realment clara fins i tot en els ordinadors. I no ho és, gens. (I, a part, a fora de la informàtica és demostradament emmaranyadora i inútil.) A un observador marcià li és absolutament indiferent i incognoscible si un mòbil o un comandament o un receptor de TV fa tot de coses complicadíssimes perquè algú l'ha construït així a partir de plànols o perquè és el resultat d'un programa que s'ha compilat i que s'executa eficientment, obra d'un autor potser ja mort i tot: ell només hi veu electrons i funcions que s'executen; la hipotètica existència d'un programador ve a ser com endevinar, veient la natura i l'evolució, si tot allò és un producte directe de la necessitat funcional o ha estat «pensat» per algú. Pregunteu a qualsevol informàtic si els microprogrames (emuladors de *hardware*, etc.) són una cosa (*hardware*) o en són l'altra (*software*), o considereu vosaltres mateixos si realment es pot tractar igual un programa acabat d'escriure, amb el *font* encara fresc i recordat, i un *procediment dedica'* (probablement en forma *cablejada* i tot) que fa la seva feina eficaçment i que ha estat compilat fa temps i, precisament per això, impossible d'analitzar en termes del seu algorisme generador precompilatori, és a dir d'un *software-font* que ja ningú no recorda i

probablement no entendria. Això que diem, l'existència de *software compilat* (en expressió feliç i cèlebre) i la seva difícil distingibilitat del *hardware* «dedicat» és fàcil d'entendre en el cas de la natura: qui ens diu que un animal, planta o òrgan que observem no ha estat «dissenyat» com si fos *hardware* per fer la seva funció, o bé potser és el resultat de la «compilació» (consolidació, integració, adaptació i potser fins i tot canvi de funció) d'un esquema funcional teòric, concebut per la ment d'algú («Déu»? «la natura»? «l'atzar»? ) i compilat aviat en forma d'animal, planta o òrgan específic que, un cop posat a fer la seva feina, ja no es recorda per a res de què deia el seu programa.

Que hi ha diferents nivells d'integració, d'imbricació o de *hardwa-rització* del *software* del nostre cervell és evident quan es consideren els diferents «cervells» que tenim: l'olfactiu, que és el més primitiu (compilat ja fa una pila d'èons), el visual (tan compilat que ja ningú tampoc no pot treure l'entrellat de quin devia ser l'hipotètic «programa», probablement del tipus «reconeixement ràpid i identificació de formes»), el lingüístic (que evolutivament sembla ser una adaptació, també «supercompilada», d'un programa que originalment havia de regular els gestos, incloent-hi els moviments de mà i braços, dels músculs de la cara i de la llengua), i així successivament fins a arribar al cim aparentment més impressionant i alhora el menys madur, consolidat i segur com és el del raonament «lògic», una mena de programa –aquest sí, *software* gairebé pur– que ressegueix superficialment (i «des-emotivitzadament») el que l'hemisferi esquerre del nostre cervell fa a nivells més profunds i «compilats» (i que justament per la seva espontània formalització simple és més fàcil de comprendre, teoritzar i imitar, com saben els professors de Lògica i els constructors de «demostradors de teoremes»).

Així doncs, i resumint: «Com seran, les nostres màquines intel·ligents?» Doncs així, com ja són, i com som nosaltres i les parts del nostre cervell: dedicades, eficients, cada una amb el seu nivell de profunditat («profunditat» aquí ve a ser la mesura de com és de difícil entendre què fa, i per què i com) adaptada al compliment de la seva funció assignada. No és així com dissenyem les nostres màquines o com funciona el nostre cervell o com es conjuminen les diferents persones/cervells que actuen coordinadament en una so-

cietat, cadascú amb la seva funció i des del seu lloc? Doncs així és com serà la coexistència entre nosaltres i les nostres futures màquines intel·ligents: exactament, com entre nosaltres (és a dir, difícil). Això no treu que, als conflictes socials, culturals, generacionals, pre- i postpuberals, etc. hi haurem d'afegir els nostres conflictes (o amors, o indiferències) amb certes màquines (com ara els tenim amb la sogra, amb en Bill Gates i les seves coses, o amb el gat del veí), tot, segurament, en el marc d'una, diguem-ne, harmonia general. I tot (i especialment això de l'«harmonia») dit sense ni un bri de sarcasme: fet i fet, no es pot negar que –tot i que n'hi ha que ho viuen malament i ho pateixen amb sentiment– la societat és una cosa complexa d'allò més i, això no obstant, assegura si fa no fa un permanent –precari, si es vol– estat d'equilibri, tendencialment organitzada i aproximadament funcional, una olla creativa que bull en estable i difícil equilibri dinàmic «a la vora del caos» (dit això en sentit tècnic, no pas irònic). Si jo fos amic de les metàfores, que no ho sóc, diria allò tan original (i suat) que «la ment no és un ordinador, la ment és com l'Internet» i m'hauria quedat tan ample. I potser l'analogia no hauria estat tan estúpida com sembla: al capdavant, totes dues coses són xarxes descentralitzades i caòtiques (com molts cervells, el meu, sense anar més lluny) i alhora capaces d'autoorganitzar-se i donar lloc a fenòmens emergents (anava a dir pornogràfics o *freakies*, pensant en certs continguts de la Web) com potser la mateixa intel·ligència (per no parlar, ai!, de les «emocions» i, ui, de la «consciència»). En fi, deixo l'exploració d'aquest símil barat en mans dels metafòrics recalcitrants.

En una època com la nostra d'un riquesa i una abundància insultants, d'una llibertat d'opcions exagerada i d'una diversitat més o menys tolerada, potser ens hauríem de preparar per acollir entre «els nostres» els nous productes de la IA, com ja hem acollit plenament entre nosaltres cotxes propis, banderes sagrades i animals exòtics. I és que, com deia l'inefable Pich, «som tan rics que nedem en l'ambulància».





## Després de l'hivern ve la primavera

**C**OM saben tots els lectors d'aquest Butlletí, la història de la Intel·ligència Artificial (IA) ha sofert moltes pujades i baixades, i alguns daltabaixos. Va començar, als anys 1940, amb l'alegria i empena encomanadisses de matemàtics sense manies (i en particular sense cap mena d'animadversió –al contrari– als biòlegs) com eren Turing, Von Neumann, Pitts, Wiener o Kleene.

BUTLLETÍ DE  
L'ACIA  
NÚMERO 33  
PRIMAVERA DE 2005

Però aquesta edat primerenca i potser ingènua es va acabar de cop l'estiu del 1956 amb la insurrecció d'uns matemàtics joves, anti-biòlegs, purs i imparables, encapçalats pels rebels-en-cap i futurs gurús McCarthy i Minsky, que van consagrar i acabar imposant el *model simbòlic* de la IA (ja sabeu: ser intel·ligent és semblar-ho i, a més, tenir un bon *software* i anar-lo executant en la novetat del moment (1956), els ordinadors electrònics, tan més nets i presentables que el *hardware* llefiscós i informe que es coneix com a «cervell»; ah! i, sobretot, sorprendre el personal fent-li coses que, «si haguessin estat fetes per un humà, hauríem titllat d'«intelligents» –Minsky dixit). Aquest va ser el model dominant durant trenta anys; fins als vuitanta, en què un nombre creixent d'informàtics, entre els quals em compto, així com biòlegs (amb les seves xarxes neuronals) i altres altermundistes de divers plomatge van –vam– començar a dubtar que tot fos tan simple o que l'eslògan aquest de «tot són símbols» permetés anar gaire més lluny. I així, després de la crisi dels vuitanta va venir l'anomenat «hivern de la IA» dels 1990. Ni els productes que sortien complien el que prometien, ni els filòsofs de la IA (una espècie nova, apareguda els anys seixanta) tampoc no hi ajudaven gaire: creien que hi veuríem clar quan «definíssim» (bufa! com?) els *qualia*, les *emocions* i la *consciència* –plantejament que, com es pot sospitar, representa un problema mal definit (els

bromistes dirien que, venint de filòsofs, això era inevitable) i probablement irresoluble. La cosa va acabar en un atzucac i en una confusió de ca l'ample, amb la sensació d'uns anys perduts i amb nosaltres gens ni mica il·luminats. Semblava que ningú no tenia ni la més remota idea de cap on podia estar anant la IA i, nosaltres, de per on enfilars la recerca que ens és pròpia.

I ara, on som? Doncs com diria un antic ministre espanyol cognomenat Acebes, hi ha dues línies d'investigació: l'una —especialment familiar a Catalunya— és el pessimisme congènit: «això no hi ha qui ho adobi», «sempre estem igual», «no ens en sortirem pas», etc., tot esperant a veure com se'n surten els militars americans i adlàters (els polítics i l'agència científico-militar estatal DARPA), carregats de dòlars, que ens van prometent resultats brillants —no sé si «brillants», però, és l'adjectiu que més s'hi escau— pel que fa a armes immobilitzants a base de descàrregues a distància (les *stun guns*), o pel que fa al *new model army* fet de «soldats» (artificials) efectius i resoladors, invencibles, insensibles i altres «in-» apropiats (per exemple, inequívocs, inequívocables, no culpables —ni jutjables—, no enterrables —ni plorables, per cap mena de família—, etc.) o d'altres *gadgets* també promocionats per la DARPA (*Defense Advanced Projects Agency*, la mare de l'Internet quan encara estava en estat fetal) com ara els *pilots-on-command* (avions tipus espia o bombarder, autònoms i supervisats de lluny), *autobots* o robots de comandes indirectes genèriques (vull dir, no de dir-li «apunta l'arma cap allà» sinó «troba l'enemic» o «conquereix el turó» o «[fes-t'ho com vulguis, però] ocupa la casa, i identifica i deté tothom qui hi trobis»), d'aspecte ja francament «intelligent», així com els *chatbots* de batalla (assessors o consoladors de soldats (humans)) o els «robots rucs» (*dumbots*) que naveguen i fan coses complexes només perquè prèviament se'ls ha atipat (literalment) de dades i coordenades sobre tots els moviments i escenaris remotament possibles (no se'ls demana que sàpiguen què fan ni que ho sàpiguen explicar). I més productes que segur que sortiran de la factoria DARPA-Rumsfeld-Bush i que, segons ens prometen, acabaran configurant la IA de l'endemà, ja neta de conflictes de consciència i de dubtes filosòfics i problemes de futur.

L'altra línia acebesiana d'exploració em sembla més probable, més controlable, més sensata, més intel·ligent, i més «IA» (vull dir,

més pròpia del nostre ofici i mentalitat), i és la que us proposo tot seguit, perquè hi anem pensant tots plegats. No sé pas si els fets ens donaran la raó, però almenys hi veurem una sortida raonablement pensable, filosòficament imaginable, professionalment satisfactòria i tecnològicament factible, que ja és molt d'agrair. Es tracta d'ajuntar les reflexions següents: (1) la constatació simple consistent a reconèixer que, amb la xarxa i els navegadors-miracle, els PC d'ara ja no són el que eren, ni s'hi assemblen gens, els ordinadors d'abans; (2) el conegut concepte de singularitat; i (3) tres propostes actuals, entre altres de possibles, que jo personalment he triat (m'*empatxoquen* força) i que cito més avall com a possible futura sortida de l'«hivern» informàtic: (a) la d'en Jeff Hawkins de PalmComputer, (b) la de l'ínclit i conegut Douglas Lenat (ex de Stanford, avui amb empresa pròpia i suport de DARPA), i (c) la hipòtesi electromèdica visualitzada per gent diversa però ben sintetitzada per Tom Mitchell de Carnegie Mellon. Es tracta de visions no imaginables fa un parell o tres d'anys però que ara, vista la vertiginosa successió d'improbables productes informàtics (i mèdics) que tenim, potser ens permetran «predir» (ai!) una mica el futur que, en el fons, és el que tothom vol (per sentir-se tranquil) i que tothom —sobretot els periodistes— ens demana als informàtics (!! ) com a dipositaris de l'Oracle. L'exercici potser servirà per saber on és i què serà aquesta primavera que tothom espera i sobre la qual ens obliguen a pronunciar-nos en públic i pitoníssicament. Anem per parts.

*«El futur de la IA ja no és el que era», i de la temible singularitat...*

Ray Kurzweil, l'informàtic-futuròleg americà que teoritza sobre els futurs soldats-màquina d'eficiència humana (com a mínim), ens fa notar que un PC que avui costa 1.000 euros té, si fa no fa, la potència de càlcul del cervell d'un insecte i que, si continuen les tendències com fins ara, d'aquí 15 anys és més que raonable, per extrapolació simple, que un ordinador equivalent en mida/preu tingui la d'un humà. Què farem, doncs, l'any 2020? No és per espantar ningú, però ja podem anar preparant el terreny per quan arribi la cosa. Per cert, la «cosa» fa tot l'aspecte de ser una singula-

ritat, que la Wikipèdia defineix així: «punt més o menys previsible en el desenvolupament d'una civilització en què el progrés tècnic s'accelera de tal manera que supera qualsevol capacitat humana actual de comprendre-la o de predir-la [més enllà del punt singular]». (Penseu en Súmer, hi afegeixo jo, o en la impremta i el seu corollari: la Il·lustració.) Parafraçant Sartre, és clar que «el futur [de la IA] ja no és el que era» i que el d'ara no és el que serà.

El cas de Jeff Hawkins és, potser, professionalment encoratjador. Més de deu anys després d'inventar el *palm computer* (i fundar una companyia per fer-ne i fer-se d'or venent-los a raig) un dia va tenir una revelació: estant al seu despatx, de cop va «entendre» què era la intel·ligència (com explica al seu llibre *On Intelligence*, un èxit editorial recent (2005) a Amèrica). Diu que va entendre que la intel·ligència (humana) és, en primer lloc, molta, moltíssima memòria que ens permet recordar tota mena d'escenes, contextos, moviments, maniobres, etc., que hem vist o après o que podem arribar a veure algun dia. En segon lloc i complementàriament, un mecanisme de reconeixement i comparació de formes molt precís, que «nota» variacions ínfimes entre dues situacions i, un cop detectades, avisa perquè immediatament se'n treguin, deductivament, les conseqüències. I, en tercer lloc, un mòdul heurísticodeductiu i calculatori molt potent que genera sense parar «hipòtesis» explicatives del canvi, estima molt expertament la probabilitat de cada hipòtesi, en dedueix conseqüències esperables, les espera o les provoca, i reavalua la línia deductiva seguida i la probabilitat de les hipòtesis segons el cas (això últim, que els humans fan molt eficientment, ens és molt fàcil d'implementar gràcies a l'experiència acumulada per la IA en sistemes bayesians d'inferència).

«EL CAS DE DOUG LENAT ÉS DIFERENT. VA COMENÇAR A STANFORD DINTRE DEL PARADIGMA CLÀSSIC DEL «TOT SÍMBOL» I VA PRETENDRE FER UNA MENA DE DICCIONARI UNIVERSAL EN QUÈ TOTA ACCIÓ QUEDÉS EXPLICADA I DETERMINADA PER LA SEVA DEFINICIÓ»

El que fa gràcia de Hawkins és que trenca amb l'esquema tradicional i reiteratiu dels autors de manuals d'IA que consisteix a explicar que la intel·ligència és això o això altre, que els altres no tenen raó i jo sí, i que la meua filosofia de la IA és tan bona que algú ja s'afanyarà per construir ell mateix una màquina seguint-la i veurem què passa. No. Hawkins ha engegat de seguida els seus projectes, basats en la seva visió de la intel·ligència (que no pretén que sigui millor que la de qualsevol altre), i per això ha fundat una institució (el *Redwood Neuroscience Institute*, RNI). Per construir-la ell mateix, per avaluar-ho tot plegat, per rectificar quan cal, i per

aprendre'n.

El cas de Doug Lenat és diferent. Va començar a Stanford dintre del paradigma clàssic del «tot símbol» i va pretendre fer una mena de diccionari universal en què tota acció quedés explicada i determinada per la seva definició. Tot era un joc de símbols i semblava que hauríem d'acabar posant Lenat i el Cyc (el seu projecte/producte, pronunciat *psic*) a la llista de les bones però equivocades intencions de què, segons diuen, està empedrat l'infern (i també el cementiri de la ciència). I vet aquí que no: una dotzena d'anys més tard Lenat ens reapareix, com Hawkins, amb companyia/institució pròpia (Cycorp) i grans projectes. Ens diu que, tal com té ell la base de dades després d'anys d'anar-la omplint amb definicions (en té vora mig milió!), ara ja pot començar a dir, sense que li caigui la cara de vergonya, que el seu Cyc comença a «entendre» algunes coses: efectivament, les sap caracteritzar des de molts punts de vista (que li han anat donant els contribuïdors), pot comparar-les, establir rangs i diferències, deduir-ne com més conseqüències millor, acarar-les amb la realitat, reavaluar o corregir, etc., etc., tot una mica com fem els humans, que des de petits no hem fet altra cosa que recollir (moltíssima) informació sobre les coses, treballar-hi, deduir, comparar, etc.

A més, Lenat, a qui molts ja havien donat per derrotat, està eufòric. Per una banda creu que amb la barator de les memòries i dels PC del futur i gràcies a les sorpreses inesperades del Google (o d'un postGoogle) potser aviat ens podrem «aclarir els conceptes» capbussant-nos *on-line* en un patracol de coneixement acumulat i compartit, nascut de l'experiència de la gent i comunicat per aquesta amablement al Cyc via web. En aquest pròxim hipotètic Cyc + Google, cada concepte i situació tindrà (quasi) infinites entrades, relacions i perspectives (no recorda pas això una mica la «intelligència» d'en Hawkins posada a l'abast de qualsevol?). I, per l'altra, Lenat ofereix a tothom la possibilitat de contribuir des de cada PC al seu projecte Cyc, per millorar-lo com més va més, a base de fer-li preguntes i jutjar les respostes que el Cyc ens doni i explicar-li (al Cyc) fins a quin punt és bona o no la seva contesta i, així, enriquir-lo, de manera que cada vegada domini més conceptes, entengui la complexitat de cada situació i no només sigui —i no merament sembli— més intelligent sinó que tingui

més d'allò que diem sentit comú (Déu n'hi do l'èxit si el procés ens portés fins a aquest punt!). Ningú, almenys jo no, es podia haver imaginat durant la crisi hivernal dels 1990 que la solució vindria no de les xarxes neuronals sinó dels desprestigiats «símbols» (apa!) de Lenat, fent marrada pels PC i la seva ubiqüitat i memòria barata, pel Google (qui hi hauria somiat!) i pels projectes coparticipats tipus SETI, tan de moda aquests dies, aplicats a una «base de sentit comú» construïda per acumulació col·lectiva. Cap d'aquestes coses no existien aleshores ni eren remotament imaginades.

Encara voldria afegir un tercer cas, el que explica clarament, per exemple, Tom Mitchell, de CMU. Es tracta d'un costum mèdicoexperimental a què els «neurocientífics» s'han acostumat darrerament, el de fer exàmens de cervell via imatges de ressonància magnètica funcional (allò que en l'argot se'n diu fMRI). La de coses que hi aprenem! (i, atenció!, la de coses que hi podrien arribar a aprendre —horror!—, si no hi fem res, els publicitaris i els polítics!). Aviat sembla que arribarà el dia en què podrem correlacionar íntimament els components sintàctics i semàntics mentals de la parla, de les emocions i records i dels plans i les accions del nostre cervell amb les zones (i àdhuc circuits) en què s'hi generen i «couen». Tant, que potser ja no caldrà que parlem amb les nostres màquines sinó que —mai més ben dit— aquestes ens «endevinaran el pensament» (i ves que ningú aprofiti per canviar-nos-el). El resultat potser serà una estreta simbiosi màquina-cervell que, si passa, segur que ho canviarà ben bé tot.

«AIXÍ DONCS, COM EN TANTES ALTRES COSES POTSER QUE FEM LA FEINA QUE ENS TOCA FER; PERQUÈ -ELS CATALANS HO SABEM PER EXPERIÈNCIA- SI NO HO FEM NOS-ALTRES NO ENS HO FARÀ NINGÚ»

I a propòsit de tot això, heus aquí una postil·la terminal: vista l'(exagerada?) atenció crítica que desperten imaginades —o reals— amenaces socials, com ara la transgènia dels nostres aliments o el cèlebre i hipostasiat canvi climàtic, o la radioactivitat (ben real) del plutoni de Vandellòs, i vista, contràriament, la poca o nul·la sensació d'amenaça que percep la societat i els *media* (per ignorància? per incompetència? perquè els entesos no ho sabem explicar?) pel que fa a la postulada singularitat de la IA que se'ns acosta (qui sap si més amenaçadorament i ràpida que la fi del petroli), la pregunta que em faig és: qui preveurà —i qui s'hi prepararà, i qui controlarà— les conseqüències (que no dic pas que hagin de ser nefastes) que se'n puguin derivar si ningú no ha previst ni en somnis que hagi d'arribar cap singularitat? No estic parlant d'espantar-se'n, que

tampoc no fóra bo, de cap manera, sinó tan sols de tenir-ho en compte, d'explicar-ho bé (vull dir, donant claredat i confiança) i de començar a engegar el sistema immunitari social contra el que pugui passar (i si no passa res, doncs tal dia farà un any). Oi que ja tenim el cos i les vísceres mig preparats –gràcies a la por espontània, i a la premsa (no tan espontània), i a l'exageració que a tots ens és pròpia– per afrontar qualsevol dels suposats desastres que hipotèticament ens vindran dels aliments, del clima o del petroli (o així ens ho sembla)? Doncs potser fóra qüestió d'anar esmolant les eines per a possibles sorpreses que puguem donar els del nostre gremi. Que si no avisem després ens diran traïdors... i serà en el pitjor moment, i amb la nostra credibilitat oracular per terra.

### *...a la primavera*

Així doncs, com en tantes altres coses potser que fem la feina que ens toca fer; perquè –els catalans ho sabem per experiència– si no ho fem nosaltres no ens ho farà ningú. I potser que anem preparant la primavera aquesta de la IA que s'atansa després de l'hivern dels 1990. Perquè, això rai: si fem els deures, passarà que «quan haurem arribat al pont, prou que veurem com el podrem passar» (seguint la metàfora i el consell d'un il·lustre barbut renà mort fa poc menys de cent vint anys a Londres). Ja que ens han encarregat l'oracle, potser que el fem.





# Com construir un cervell

*Primer pas: El mirall, la padrina i l'olla de grills*

BUTLLETÍ DE  
L'ACIA  
NÚMERO 34  
TARDOR DE 2005

**C**OM tothom sap, una de les coses que més caracteritzen –a estones– els humans (segons quins) és l'empatia, la capacitat de posar-se en el lloc dels altres, i/o de «sentir les coses que ells senten». Això va des de compartir un estat d'ànim concret fins a l'**horror** que sentim de veure com una taràntula peluda que hi ha al coixí d'en Sean Connery li amenaça la integritat facial en una famosa escena (de llit) d'un episodi de James Bond. Veiem com la fera és a punt de llançar-se-li sense pietat a la cara i també, en un acte automàtic, la de tots els espectadors. Fins ara, com podia passar això («passar», com si diguéssim, a través de la pantalla) era un misteri. Però ara (vull dir, durant l'any 2005) ja ho està deixant de ser una mica, gràcies a recerques fetes a Groningen, UCLA, Parma, UC San Diego i Harvard, que sembla que s'hi hagin posat d'acord (fa goig que hi hagi tanta gent avançant en aquests temes). Resulta que les neurones –nostres i dels micos– que s'activen amb els circuits que controlen la percepció i el moviment, i que en aquest cas estan lligades a la repugnància i la por (unes emocions per cert fortament «sentides») i al compulsiu «reflex de fugida», també s'activen quan tot això veiem que li passa a qualsevol altre que es troba amb el desagradable aràcnid, sempre que nosaltres ho estiguem veient. Ah, i el sentiment compartit s'estén a la comprensió immediata de la situació, context i finalitats de l'escena vista (de seguida entenem la sorpresa i la por d'en Bond davant la taràntula, o què estava fent i què volia fer el personatge que hem vist amenaçat).

Un 20% de les neurones que s'activen quan tot això ens passa a

UNA DE LES COSES QUE MÉS CARACTERITZEN –A ESTONES– ELS HUMANS (SEGONS QUINS) ÉS L'EMPATIA, LA CAPACITAT DE POSAR-SE EN EL LLOC DELS ALTRES, I/O DE «SENTIR LES COSES QUE ELLS SENTEN»

nosaltres es continuen activant quan veiem la mateixa escena amb algú altre: són les anomenades *neurones especulars* o *miralls* (*mirror neurons*), aparentment redundants, que solen residir a la nostra insula anterior. I, curiosament, en tots **dos** casos unes ones anomenades *mu* (de « $\mu$ »), que circulen pel nostre cervell a 8-13 hertz i que vés a saber quina missió tenen, queden automàticament bloquejades, en tots dos casos. I, més curiosament, aquestes mateixes ones es bloquegen en els infants autistes només en el primer cas (experiència directa), no pas en el segon (experiència indirecta, transferida visualment, però sentida pels no autistes pràcticament com si fos directa). I, encara més curiosament, resulta que el que sol faltar als infants autistes és justament aquesta transferència vivencial o «comprensió» de la situació i emoció aliena; allò que tècnicament s'anomena *teoria de l'altre* –que només tenim els primats (i no tots)– i que nosaltres solem dir **empatia**. Doncs: ah, carai! Potser tot això deu voler dir que les bèsties intel·ligents nostres (vull dir, de la IA) hauran de tenir *cèl·lules-mirall*, també? Doncs hi haurem d'anar pensant (si no volem, és clar, que ens deixin «tirats» per qualsevol cantonada sense que entenguin que potser ens hem fet malbé el genoll i no podem seguir).

UN 20% DE LES NEURONES QUE S'ACTIVEN QUAN TOT AIXÒ ENS PASSA A NOSALTRES ES CONTINUEN ACTIVANT QUAN VEIEM LA MATEIXA ESCENA AMB ALGÚ ALTRE: SÓN LES ANOMENADES NEURONES ESPECULARS

I no és tot. Fa un temps, un neuròleg dissortat va suggerir que teníem al cervell neurones tan i tan especialitzades que en teníem una que estava dedicada a la nostra padrina: cada vegada que la veïem, la neurona en qüestió (la *grandmother neuron*) se'ns posava contenta i engegava a funcionar, a *carregar en memòria*, com si diguéssim, tot el que en recordàvem o s'hi referia. (Naturalment, el proponent volia dir que a certs conceptes »prou freqüents hi teníem consagrada una neurona prou ben dotada i completa.) No cal ni dir com el pobre home va ser ridiculitzat. Doncs, heu-nos aquí que acabem de saber que hi ha gent, gent concreta (i normaleta), que té una neurona dedicada a la Jennifer Aniston (la de Friends) o a en George Clooney; com sona. Tota una neurona dedicada! Ara els neuròlegs ja parlen de la *Jennifer Aniston neuron* [sic], com l'han batejada, i a cap d'ells ja no se'ls ha escapat mai més el riure. ¿Deu voler dir això potser que hem de construir neurones dissenyades una a una, a la mida del seu objecte particular, i que ja no tot va de *xarxes neuronals complexes* i de *càlculs col·lectius* autoorganitzats com creïem, i que els objectes mentals/conceptuals (la «padrina»)

no necessàriament «emergeixen» (concepte tan intangible com poc comprès) de la collectivitat i la connectivitat? I ja no caldrà que fem el miracle que se'ns demana, de fer sortir d'una olla de grills un orgue de gats? Doncs pot ser. I si és, ja tenim una nova sorpresa. (Carme Torras, haurem de rellegir-nos la teva tesi doctoral! al final hauràs tingut raó en l'escala i l'objecte d'estudi.) Es fa interessant, la cosa, oi?

*Segon pas: El joc, la curiositat, la traça,... i l'ajut de la cuca Portentosa.*

Això no vol dir que hàgim de deixar estar els plantejaments col·lectius, d'«eixam» (*swarm*), que sembla que continuen enlluernant sobretot els militars (ara s'ha sabut que a Corea del Sud volen substituir els soldats que vigilen la frontera amb *swarm robots* que facin la feina cooperativament i sense cansar-se ni pensar!! en una mena de «joc col·lectiu»). La paraula joc aquí no l'uso en sentit pejoratiu; al contrari, sembla que els robots que construïm tenen un caràcter com més va més lúdic (o infantil). Preneu-ne tres exemples il·lustratius: (1) a París els col·legues de Sony estan fent Aibos dotats del que en podríem dir una curiositat innata: com els nostres fills petits, no paren d'emprenyar i forçar els límits de cada situació: allò que tècnicament es coneix com treballar *a la vora del caos* (com hi treballen el nostre cor o el nostre cervell, vísceres que, quan es comporten de manera que l'*atractor* en l'espai de fases ja no és *estrany* sinó periòdic, vol dir que estem a punt de dinyar-la). I ves que els de Sony no en facin de nous, d'Aibos, que, també com els nostres fills petits, ens torturin a preguntes (i per què?!, i per què?!») com segur que els ho suggeririen els **Hawkins** o **Lenat** que citàvem l'altre dia, conseqüents amb la seva (nostra?) nova visió de la intel·ligència; (2) molts centres de robòtica continuen provant de fer, però ara amb força èxit, robots «amb traça», vull dir amb destresa (*dexterity*), a més de prou sentit d'equilibri per caminar gràcilment i sense caure: els que hàgiu vist els vídeos que reparteixen els laboratoris us heu degut quedar parats, dels saltirons que fan, i com els costa de desequilibrar-se o fúmer-se d'oros encara que vulguin; i (3) el robot Kismet, qualificat d'«emocionalment expressiu», de l'escuderia **Rodney Brooks** està

NO HI HA NI PRESSA  
NI NERVIS. IBM  
COMPTA QUE, SI LES  
PRESTACIONS MA-  
TERIALS SEGUEIXEN  
LA LLEI DE MOORE  
COM FINS ARA, ES  
POT ESTIMAR QUE  
EN QUINZE ANYS  
JA TINDREM RESUL-  
TATS SIGNIFICATIUS  
I COMPORTAMENTS  
MÍNIMAMENT REA-  
LISTES.

dissenyat perquè actuï interactivament, juganerament, diríem, amb tothom que troba, i sembla que amb prou gràcia (especialment quan es troba dones amb faldilles: les martiritza).

En un llibre que ha editat fa poc, l'amic nostre (vull dir, de Catalunya) **Michael Arbib** es planteja això de les emocions i els robots i, encara que els tècnics invitats no acaben de saber com «programar-ho» tot això, la impressió que fa tot plegat és que estem sortosament ben lluny dels primers plantejaments balbucients de quan es trobaven en McCarthy i la seva colla pels porxos «romànics» de Stanford i encetaven a parlar del tema; llavors sí que no sabien gens com posar-s'hi. Ara, afortunadament, tenim al campus d'Irvine (per cert no gaire lluny de Disneylàndia i del museu de monstres de cine de Los Angeles) un escarabat gegant –fa la mida d'una rata de claveguera ben alimentada!– anomenat *Gromphadorhina portentosa* que, a part de xiular com una serp en zel, fa por. Doncs li han foradat el cervell d'elèctrodes per anar recollectant i correlacionant tots els seus moviments i pulsions i, vatua, els del laboratori en qüestió estan tan contents que, quan els ha passat la por, diuen meravelles del que han après de la bèstia: com camina, com busca, com evita obstacles, com s'equilibra, com grimpa, com es distreu (o no) amb un nou (sub)objectiu. Tot ho tenen enregistrat, i saben quin moviment correspon a cada «decisió» neuronal, on hi ha i com funciona el centre de transacció de decisions (en cas de canvi o de conflicte o distracció), com alterar i amb quin èxit els moviments o intencions de l'animal. Veient la cosa, és fàcil d'entendre que estiguin contents com un gínjol i que creguin que ja han arribat al moll de l'os de la intel·ligència (de la cuca, si més no). I si alteren o activen els moviments de la bestiolassa, qui no ens diu que podrem aviat fer el mateix amb un malalt de Parkinson o de cadira de rodes?

*Tercer pas: Com fer-ho, això de construir un cervell (i de passada millorar el nostre, en el cas que en tinguem)*

Això ja ho ha començat a fer la gran IBM per nosaltres. Ha agafat una supermegamàquina Blue Gene de les quals presumeix, concretament una que té a Suïssa (a la Politècnica de Lausana), i l'ha dedicada a temps complet a «simular el cervell». Apa! No es

pretén pas que la simulació sigui gaire predictiva ni precisa, només se'n vol fer una rèplica *in silico* [sic] a base d'anar construint l'equivalent informàtic d'uns cilindres neocorticals (CNs) de  $3 \times 1/3$  mm cada un contenint  $10^4$  neurones i les seves corresponents  $10^4 \times 10^4$  connexions, que es van estenent en una mena de capa (de 1 CN de gruix) feta de CNs juxtaposats i que ve (vindrà) a ser l'equivalent de la «matèria grisa» o neocòrtex del nostre cervell, justament el que correspon a la zona en què s'aprenen, s'arxiven i es mantenen els records, es negocien les pulsions sentides adaptant-les a la situació percebuda i, en general, s'estableix el «criteri» d'actuació i els passos a seguir. Vaja, el que en diríem **reflexió i pensament complex**. Que el sistema no es comporta com un cervell de debò? Doncs ajustem paràmetres i tornem-hi. No hi ha ni pressa ni nervis. IBM compta que, si les prestacions materials segueixen la llei de Moore com fins ara, es pot estimar que en quinze anys ja tindrem resultats significatius i comportaments mínimament realistes.

A més, entrelaçat amb la simulació també hi ha un conjunt de mòduls selectius per respondre preguntes concretes, com ara «què i com és la diferència i/o concatenació, en un cervell, entre **voler** fer una cosa i fer-la, entre **saber** que una cosa passa i **actuar-hi**?» Una idea interessant és anar fent aquesta experimentació sobre la **màquina** i, en **paralel**, actuant sobre cervells humans normals o que tinguin algun possible handicap mental. Per exemple, si cal compensar dèficits observats, fer-ho en paral·lel: amb ajustaments de programa (a l'ordinador) i amb injeccions d'un factor de creixement del teixit deteriorat (al malalt), i comparant. La idea és anar progressant fins a objectius més ambiciosos: crear i controlar moviments (i recuperar els perduts), crear i/o modificar **sensacions** (bufa!), i fins i tot **estats d'ànim**, poder modificar estructuralment la capacitat o les condicions d'**aprenentatge**, **simular comportaments autistes**, **eliminar inhibicions** (i potser deslliurar creativitats?), augmentar o reduir les capacitats i característiques de la parla, millorar la memòria i l'**aprenentatge**, etc., etc. Com es veu, aquest programa és doble: per la banda del Blue Gene retoquem i potinegem com a vulgars informàtics en plena feina; per la banda del malalt mentalment lesionat, li anem donant progressivament els subproductes tecnològics de la nostra recerca.

En fi, de moment deixem-ho aquí, que ja n'hi ha prou per

COM ES VEU,  
AQUEST PROGRAMA  
ÉS DOBLE: PER  
LA BANDA DEL  
BLUE GENE RETO-  
QUEM I POTINEGEM  
COM A VULGARS  
INFORMÀTICS EN  
PLENA FEINA; PER  
LA BANDA DEL  
MALALT MENTAL-  
MENT LESIONAT,  
LI ANEM DONANT  
PROGRESSIVAMENT  
ELS SUBPRODUC-  
TES TECNOLÒGICS  
DE LA NOSTRA  
RECERCA

avui. Al pròxim episodi potser plantejarem si de camí cap a la IA, en resum, **anem cap amunt o cap avall**? Ho dic perquè aquests darrers dies hi ha una discussió oberta, inconclusa, en què figuren **Jon Huebner** i **Eric Drexler** per la banda diguem-ne negativa («no avançarem gaire perquè estem arribant a tots els límits teorico-pràctics, o ja els hem passat de llarg, i la nostra productivitat científicotecnològica és la que és») i el nostre amic i convidat permanent a taula **Ray Kurzweil** i el científic de nom ben rellevant i oportú **John Smart** que opinen, sempre positius, que els altres dos s'equivoquen clarament per múltiples raons que ells exploren i que els porten a pensar que aviat (un «aviat» estimat en 35-75 anys a partir d'ara) ja tindrem el que volem: genuïna IA, de la bona. Potser caldria afegir a aquest quartet la veu sempre interessant i discordant d'en **Roger Penrose** (no cal pas que us expliqui què hi diu, oi?): doncs que si la Física Quàntica no fa un avanç espectacular (un «salt quàntic», per aprofitar la broma) en la comprensió dels microfenòmens neuronals no hi haurà res a fer (i que si el fa, llavors estem tots salvats). Queda pendent la pregunta que em feia jo l'últim dia: I què passarà amb la famosa **singularitat** (quan arribi, i els nostres fills siguin vellets)? No, no és per res, és per avisar-los ara que hi ha temps.

## Encontres a la tercera fase?

**L**A «Intelligència Artificial» (IA) es pot entendre de dues maneres: en general, com a concepte i, en particular, com a moviment. La segona, la IA com a moviment (de reacció), neix, com és sobradament conegut, l'estiu del 1956 en una estació d'esquí de New Hampshire de la mà del «grup de Dartmouth»: John McCarthy, Marvin Minsky i alguns altres il·lustres prohoms (o pronois, *que hauria dit Pla*).

BUTLLETÍ DE  
L'ACIA  
NÚMERO 35  
HIVERN DE 2006

Com sabem, presenta dues fases: la primera (la «IA dura» o «simbòlica») és la de l'èxit fulgurant i imprevist; la segona —a partir dels 1980— és la de la sensació de no veure-hi més enllà, i de sentir-se superable o substituïble per les *xarxes neuronals* i el *connexionisme* (que finalment van ser cooptats com a «IA subsimbòlica»). Si anem per aquí, allò que toca ara (2006) és la tercera fase. Segons Jeff Hawkins (2005) (de qui ja vaig parlar fa dos números), ara es tracta —per diferenciar-la de l'anterior, la «Intelligència Artificial» (IA)— de la «Intelligència Real» (IR). (Ja es veu que en qüestió d'eslògans els americans no es queden gens curts!)

Però també podem entendre «Intelligència Artificial» (IA) no pas com allò concret que els joves matemàtics repatanis del 1956 volien plantejar (i van decidir batejar amb aquesta expressió provocativa i un pèl insolent) sinó com a concepte, referit a «tot allò relacionat amb l'estudi rigorós no basat en un prejudici programàtic previ (per exemple «tot el que és intel·ligent es pot programar en un ordinador») sinó en la intel·ligència real o coneguda, que pot portar un dia a tenir «màquines intel·ligents» o «intelligències no humanes». Si és així, la cosa va començar —sense que llavors ningú no en tingués la intenció ni ho pogués sospitar gens— trenta anys abans, el 1926. Aquesta història té, com a sant protector, inspirador i guia John von Neumann —Johnny per als seus (moltíssims) amics



i coneguts– i inclou la «IA» pròpiament dita (la dels rebels del 1956) com a autèntica «segona fase».

Ara (2006) entrariem doncs, després de la «intelligència artificial», en la tercera, la «intelligència real» de Hawkins. Pel que es veu, en tots dos casos, doncs, som a la tercera fase. Posem-hi una mica d'història.

### *Sant David Hilbert i les seves «recursions» (1926)*

El 1926, amb ocasió d'una disputa entre matemàtics que ara no ve al cas, l'eximi supermatemàtic alemany David Hilbert va estudiar a fons el raonament humà com si fos un mecanisme, un sistema formal sense continguts. Segons ell, un matemàtic (i, per extensió, qualsevol persona racional), quan pensa ho fa seguint unes regles formals. Tant s'hi val en què pensa, mentre ho faci correctament. Si tot aquest procés és «correcte» (tècnicament, vol dir que ha de ser consistent; i una manera d'aconseguir-ho és, com va mostrar Hilbert, fent-ho tot recursiu), aleshores –per un teorema del jove hilbertià Kurt Gödel d'alguns anys més tard (1930)– podem construir un món (*model* segons la terminologia hilbertiana) en el qual allò pensat tindrà sentit (o, dit altrament, del qual allò pensat serà una descripció exacta). Hilbert ho explicava, provocativament, així: «què hi fa que no entenguem una teoria matemàtica! no cal amoïnar-s'hi: si nosaltres no l'entendem, ja l'entendran els nostres néts» (és a dir, ja apareixerà un món en què sigui d'aplicació la teoria i ens permeti entendre-la, i fins i tot trobar-la claríssima).

Sorprenentment, l'aparent *boutade* de Hilbert, l'hem vista fer-se realitat qui-sap-les vegades: per només donar-ne un exemple, els matemàtics es van adonar que podien construir una teoria correcta (és a dir, consistent) dels nombres reals que no els obligués, però, a ser enumerables; semblava impossible que aquesta teoria tingués cap sentit (és a dir, model), perquè tothom sap que els reals no ho són pas, enumerables. Doncs bé, sí. Com havien promès Hilbert i Gödel («si és consistent té un model») existien uns «reals» enumerables (són, per exemple, els nombres *p-àdics*) que sorprenentment donaven sentit a la teoria.

AQUESTA HISTÒRIA  
TÉ, COM A SANT  
PROTECTOR, INSPI-  
RADOR I GUIA JOHN  
VON NEUMANN  
-JOHNNY PER ALS  
SEUS (MOLTÍSSIMS)  
AMICS I CONEGUTS-  
I INCLOU LA «IA»  
PRÒPIAMENT DITA  
(LA DELS REBELS  
DEL 1956) COM A AU-  
TÈNTICA «SEGONA  
FASE».

### *Sant Alan Turing i la seva «màquina» de paper (1936)*

Aquest episodi que acabo d'explicar, una mica surrealista, forma part del gran miracle hilbertià i de la gran matemàtica del segle XX, en què es va demostrar rigorosament viable construir la ciència (descriure la realitat física i operar-hi) com si aquest procés de descoberta i construcció fos un mecanisme formal. Faltava trobar-lo, i és el que va fer deu anys més tard l'anglès Alan Turing: la seva «màquina» (del 1936) és un mecanisme (formal) universal, capaç de fer tot el que, aparentment, som capaços de fer els humans. Turing mateix va preveure que aquesta màquina imaginària es podria bastir físicament algun dia (és l'ordinador que nosaltres coneixem) i fins i tot li va reservar el nom (*computer*) que més endavant von Neumann va tenir l'amabilitat de donar-li.

### *Sant Johnny von Neumann i els seus cervells i maquinetes (1946)*

Cal constatar en aquesta història la presència ubíqua del matemàtic i geni Johnny, que era a Gottingen quan Hilbert feia les seves coses i més tard, ves quina casualitat, a Princeton (al despatx del costat d'Einstein) quan Turing el va anar a veure per explicar-li la seva «màquina». Que en Johnny ho va acollir amb entusiasme queda demostrat perquè va ser ell que va insistir a (1) instal·lar un «sistema» de tabuladores IBM com si fossin una màquina de Turing, a Los Alamos, perquè calculesin els estralls que faria la bomba atòmica que hi construïen, a (2) recomanar (als militars) dos xicots eixerits de la Universitat de Pennsilvània que estaven construint una complicadíssima cosa electrònica que va acabar (el 1946) sent l'ENIAC, el primer ordinador (és a dir, la primera màquina de Turing físicament construïda, el primer *computer*), i, més tard, a (3) dissenyar i fer ell mateix els plànols de quasi tots els ordinadors que es fan i es desfan (l'anomenada arquitectura von Neumann).

Però no sols això: tan bon punt va sentir el que li explicava Turing el 1939, va encomanar entusiasmat a uns col·legues (el biòleg McCulloch i el lògic matemàtic Pitts) que li fessin una versió més «realista» de la màquina de Turing (aquesta, estranyament,

...TAN BON PUNT VA SENTIR EL QUE LI EXPLICAVA TURING EL 1939, VA ENCOMANAR ENTUSIASMAT A UNS COL·LEGUES (EL BIÒLEG MCCULLOCH I EL LÒGIC MATEMÀTIC PITTS) QUE LI FESSIN UNA VERSIÓ MÉS «REALISTA» DE LA MÀQUINA DE TURING (AQUESTA, ESTRANYAMENT, NO S'ASSEMBLAVA GENS A UN CERVELL HUMÀ, TOT I SER-HI COMPUTACIONALMENT EQUIVALENT)...

no s'assemblava gens a un cervell humà, tot i ser-hi computacionalment equivalent); aquests van tenir aviat la feina feta (el 1943) i, del seu model (que, ara sí, s'assemblava a un cervell), se'n va dir *xarxa neuronal*. Johnny va considerar que era un bon model del cervell humà i, quan va presentar al públic l'ENIAC el febrer del 1946, va tenir la gosadia (o cometre la imprudència) de dir, als ignorants i crèduls periodistes que cobrien la festa, que ja havíem construït dos cervells equivalents a l'humà: un, el model formal de McCulloch i Pitts, i, l'altre, aquella bèstia electrònica de moltes tones i dos pisos d'alt que tenien al davant.

No cal ni dir l'enorme impressió que l'impressionable públic de l'època, que no s'esperava res de semblant, en va treure. Johnny va continuar impertorbable la seva feina: en menys dels deu anys que li quedaven de vida va fer els plànols d'un ordinador-model (l'anomenat EDVAC) que tothom pogués construir (i que tothom va construir), va apadrinar-ne dos (que es van anomenar, en el seu honor, *Johnniac* i *Maniac*), va començar a pensar coses que no va dir a ningú (i algunes que va dir, profundes i superavançades, encara no les entenem ara) i finalment es va morir de càncer, el 1957, no sense queixar-se amargament del fet de morir-se: deia que encara li quedaven un centenar de qüestions interessants per resoldre i que no hi havia dret d'haver-se de morir tan d'hora.

Què va pensar, en Johnny, aquests anys? Doncs, que nosaltres hàgim pogut entendre de les seves anotacions i conferències, aquestes (entre moltíssimes altres que ni esmento): màquines de Turing autoreproduïbles (també en va dir *autòmats replicables*) o disposades en xarxa (ell en va dir *autòmats cel·lulars*), un replantejament de la seva «xarxa neuronal» proposada el 1943 en forma de nou concepte que ell va batejar *autòmat finit*, una visió de la Informàtica com a part de la Física (en què l'error de càlcul hi juga el paper de «soroll» termodinàmic), la recomanació d'estudiar detingudament el cervell humà (per basar-hi, o no, qualsevol teoria sobre la intel·ligència) o, per no allargar més, el concepte de singularitat tecnològica (que ara Ray Kurzweil ha posat de moda i que tindrà lloc segons ell pels volts del 2020 o 2040, i de què vaig prometre que parlaria un dia... si no fos que, francament, com més va més me'n passen les ganes, quan constato que en Kurzweil s'ha declarat fa poc *singularitarian*, és a dir, seguidor de la seva pròpia doctrina).

*Els joves macarthistes i la seva intel·ligència «artificial» (1956)*

En aquesta secció ja hem dit més d'un cop que la «Intel·ligència Artificial» (IA) dels mcCarthys i minskys ha passat diferents fases, concretament dues, des de la seva data fundacional, l'estiu del 1956: la IA original o «simbòlica» i, a partir dels 1980, el connexionisme o «IA subsimbòlica» (que és bàsicament una cooptació de l'enemic). I que la primera, una extrapolació salvatge de la descoberta de Turing del 1936, consistia essencialment –a part la repugnància patològica que els produïa el seu professor i col·lega Norbert Wiener, que aquests joves refinats consideraven bàsicament despistat, desagradable, gras i llefiscós– en la negació de tot el que havia vingut abans (especialment l'estudi del cervell, òrgan que aquests joves i pulcres matemàtics consideraven humit, brut, llefiscós i cosa de biòlegs amb bata), afegint que la «intel·ligència artificial» que s'acabaven d'inventar consistia ni més ni menys a «escriure programes (d'ordinador) que fessin coses que, si haguessin estat fetes directament per humans, hauríem considerat intel·ligents» (en l'ajustada definició de Minsky de la [seva] IA).

Les màquines els resultaven de franc (IBM els les havia regalat), els programadors eren becaris brillants i entusiasmats que, com a mà d'obra, sortia força bé de preu (és a dir, gratis), i els primers observadors astorats van ser els també joves i pulcres militars el suport moral i monetari del quals van tenir la precaució immediata i discreta de captar. També hem explicat aquí manta vegada que després de la IA triomfant (en gran part basada en l'efecte sorpresa) van venir els dubtes, l'«hivern» simbòlic, el connexionisme o la IA subsimbòlica i la possibilitat de substituir els ordinadors per xarxes neuronals, i els programes, per l'aprenentatge i la memòria associativa, etc. I que, tot i el rebombori inicial, tampoc això del connexionisme no sembla haver acabat donant gaire la talla. I també hem dit, als darrers números del Butlletí, que aquesta distinció (IA pura versus IA connexionista) se n'ha anat anant en orris, com tantes altres coses, quan hem vist que els plantejaments que més prometen ara són una altra vegada, a una escala immensament superior, els anomenats «simbòlics», els del 1956, que abans vàiem semifracassats i que ara ressusciten, transformats, a base d'afegir memòria sense parar a les nostres màquines i anar a resoldre-ho

tot sense miraments per la via de la massiva brutalitat de les dades.

*El probable superador de les crisis: el cervell (2006)*

PER HAWKINS, LES  
MÀQUINES DE LA  
IA (COLLITA 1956)  
PODEN, PER EXEM-  
PLE EN FORMA DE  
ROBOTS, SER MOLT  
ÚTILS I BENÈFI-  
QUES PERÒ NO SÓN  
INTEL·LIGENTS NI  
HO SERAN MAI,  
PERQUÈ ELS HEM  
EXPLICAT EL QUE  
HAVIEN DE FER -I  
HO FAN DE CONYA-  
PERÒ NO ENTENEN  
EL QUE FAN

<sup>5</sup> L'autor es refereix a l'article de la plana 257. (n. de l'ed.)

Ara potser ja comencem a tenir una idea de com serà tot l'endemà. O, si més no, això és el que ens diu Jeff Hawkins, qui, després d'haver-se fet d'or inventant i venent el seu PalmComputer, el PalmPilot i el Treo a cor que vols, ara (el 2006, al llibre *On Intelligence*) ens diu que ha vist clar què és la intel·ligència, que les màquines que ha fet la IA no en són, d'intel·ligents, i que si volem la IR («intel·ligència real») hem de tornar tots a estudiar el cervell, com ens recomanava l'inefable Johnny. A més es proposa treballar en l'assumpte ell mateix amb les mànigues arremangades (en una institució no lucrativa que ha creat, el Redwood Neuroscience Institute, a Silicon Valley) per veure si té raó, i en tot cas per saber-ne més i aprendre'n. I, cosa nova en l'ofici, no ens assegura que ell tingui pas tota la raó, i afegeix que, per saber en què s'equivoca, ell mateix proposa uns quants experiments fàcils de fer per invalidar (o no) les seves teories, i així anar endavant rebutjant-les i modificant-les. (Després de tant de fantasma, una mica de racionalitat humil s'agraeix.)

Com ja vam anticipar fa un parell de números<sup>5</sup>, Hawkins explica que un dia al seu despatx va veure clar el que era la intel·ligència humana: (a) molta, moltíssima memòria, (b) una enorme capacitat de reconèixer formes i detectar petites diferències, (c) un mecanisme generador imparabile d'hipòtesis, i (d) un altre d'avaluador (de fins a quin punt una hipòtesi s'ha complert o no). Ho va veure clar perquè un dia es va fixar en un detall del despatx que no era habitual (és a dir, que, de la munió de coses que continuaven igual, va anar a percebre la que el dia abans, ni cap altre, no hi era) i tot seguit instintivament va començar a passar revista a possibles explicacions sobre la intrusa: què era allò, què hi feia allà, qui l'hi devia haver posat, per què, què havia de fer ell -si és que hi havia de fer res- amb allò, què passaria si ho feia, etc., etc. Va veure que, bàsicament, la intel·ligència humana és això que acabo de dir. I en va escriure un llibre (que va ser un gran èxit del 2005): s'afanya a dir que tot el que hi diu ja era sabut i conegut per tothom, però que calia reformular els tòpics sota una nova perspectiva (en forma

de nou *pack*, que diríem) que fes comprendre millor les coses i que guiés l'acció.

Per Hawkins, les màquines de la IA (collita 1956) poden, per exemple en forma de robots, ser molt útils i benèfiques però no són intel·ligents ni ho seran mai, perquè els hem explicat el que havien de fer –i ho fan de conya– però no entenen el que fan (una vella i persistent crítica que feia gent com Searle o Dreyfus a la IA «dura» tradicional). Però tampoc no són intel·ligents les xarxes neuronals (tampoc no entenen el que fan), perquè només organitzen fenòmens «des de baix» (*bottom-up*) sense que ningú fixi o tingui en compte cap estratègia ni jerarquia d'acció (vaja, que hi falta un estructura *top-down*). Per superar-ho tot plegat, proposa tres coses: (1) establir un model del fenomen (la intel·ligència) que consisteix en (a) un procediment algorítmic que detecta ínfimes variacions dins d'un panorama complex, (b) analitza possibles explicacions, (c) en dedueix possibles accions a prendre i (d) les avalua tant en buit (preveient-ne conseqüències) com un cop preses (examinant-ne el efectes); (2) construir mòduls que repliquin les funcions dites (així, ens caldrà un mòdul recognoscitiu, un detector de canvi, un mòdul generador d'explicacions, un mòdul deductiu, un calculador de probabilitats, un mòdul avaluador, etc.), i (3) guiar-nos en tot cas i moment per l'estudi detallat del cervell.

Quan Hawkins parla del cervell no ho fa del conjunt (que és un agregat evolutiu de milions d'anys) sinó – només– del neocòrtex, que és una estructura plana, d'uns cinc o sis mil·límetres de gruix, amb l'extensió d'un tovalló de cuina, la textura d'una escudella espessa (Turing *dixit*) i l'aspecte d'un bròquil, que conté totes les característiques racionals, de coordinació i memòria, i que aparentment treballa seguint un sol algoritme (tot i que serveixi funcions de visió, tacte, llenguatge, etc. tan i tan diferents). És això el que Hawkins i el seu institut estudien. Però, com ja se n'ha degut adonar de seguida el perspicax lector del darrer episodi d'aquest Butlletí, aquest còrtex, fet d'un milió de neurones, correspon precisament a la capa de cilindres corticals, cada un de 3 mm de llarg i 1/3 mm de diàmetre (i no, com deia jo crípticament i confusa al número anterior «de 3x1/3 mm») que estudia a fons IBM a Zuric sense presses en una de les seves megamàquines (una BlueGene que competeix en velocitat amb el nostre Mare Nostrum). I la seva

NO, LES MÀQUINES  
DEL FUTUR NO ENS  
SUBSTITUIRAN NI  
ENS AMENAÇARAN;  
ENS AJUDARAN I  
ENS FARAN COMPA-  
NYIA

brutal massivitat de dades recorda també una altre vell conegut nostre: en Douglas Lenat i el seu CyC (a pronunciar «psic») ple d'infinites referències *googleoides* sobre cada cosa, cada concepte i cada situació possibles. És potser per aquí que anirà el futur de la IA (perdó, de la IR)? En tot cas, podem donar per difunta l'analogia-base del 1956 «Cervell = Ordinador», metàfora fundacional que, segons jo dic i repeteixo sovint, ha resultat prou nefasta (Hawkins també ho creu).

### *Les màquines no faran por (2036?)*

Ah, me'n descuidava: a diferència de molts divulgadors de la IA, que sembla que no es puguin estar de veure esporuguidament els humans substituïts per tota mena de màquines (què farem nosaltres, llavors? i, un cop superflus i redundants, les nostres màquines ens esclavitzaran com a les novel·les de ciència-ficció?), a diferència d'ells, deia, Hawkins considera que les màquines que es construeixin segons la IA tradicional (programa previ i supercàlculs massius) –ja sense la pretensió d'«intelligència»– faran coses –ja les fan– que nosaltres no fem ni farem, ni podem, ni ganes (calcular trajectòries impossibles, lligar xarxes de connectivitat infinita, transmetre o manipular imatges, executar moviments complicats per agombolar un malalt, etc.) i que, de la mateixa manera, les que es construeixin a partir d'ara segons la seva IR (la del cervell) –que potser sí que podríem qualificar d'«intelligents»– tampoc no tenen per què fer res que nosaltres ja fem, sinó al contrari, complementar-nos en tot allò que nosaltres podríem fer però no volem, com ara vigilar permanentment una escena i detectar quan alguna cosa que s'hi observa és remarcable (un curtcircuit o foc incipient, algú que per la seva conducta pugui fer sospitar vandalisme pròxim, un estat de salut que faci presagiar un atac de cor imminent, etc, etc.). Aplicacions totes elles d'ajut o complementació de l'activitat humana, no pas d'imitació ni substitució.

Així doncs, tots contents. Ja ho podem anar a explicar als periodistes. No, les màquines del futur no ens substituiran ni ens amenaçaran; ens ajudaran i ens faran companyia. I sí, ja sé que algunes d'aquestes coses fan por (com evitar que la màquina que ho sap i ho detecta tot per protegir-nos de bretolades no acabi

essent el Gran Germà orwellià?). I sí, també ja sé que tot això serà possible, com sempre fins ara, perquè ens ho pagaran els militars americans de la DARPA, entossudits com estan per fer «soldats artificials» (*agents mecànics de defensa*, en diuen) perquè no se'ls morin més soldats de veritat, amb mare i germans. Però és que tenim el consol que: (a) encara que sembli estrany, o estranyíssim, i increïble, afortunadament cada vegada hi ha més poques guerres i menys morts per guerres (només cal verificar estadístiques amb serenitat, sense deixar-se endur pels telenotícies), i cal esperar que pròximament els soldats mecànics no entrin ni tan sols en acció; i (b) com hem vist sempre, darrere l'aplicació bèl·lica, que és la que justifica la despesa, ve de seguida l'aplicació humanitària (qui sap si per mala consciència): robots apagaincendis, salvadors d'esclafats en terratrèmols, ajudants de malalts, companys de vellesa, etc.

Només queda un problema. Si mai estem voltats de màquines intel·ligents, no resultarà llavors massa evident la nostra estupidesa natural?... Jo, particularment, penso que sí, que molt, i només se m'ocorre afegir-hi la frase «que Déu ens faci bonets!» que deien les nostres àvies quan ja es veia que una cosa no tenia remei.





## Robodependència

**P**ER als qui tinguin aficions històricoetimològiques, el mot **robot** ve del txec. Concretament, d'una obra de teatre que es va representar amb èxit a Londres els anys 1920 traduïda a l'anglès però amb el «robot» intacte. *Robot* vol dir ni més ni menys que «minyona»; de fet, a les llengües eslaves i en segles anteriors volia dir –llisament i brutal– «esclau/-va» (de *robota* [= treball esclau], homòleg exacte de l'alemany *arbeit*; sí, aquell mot de la frase cruelment irònica «Arbeit macht frei» [= «treball és llibertat» o «el treball us farà lliures»] que encara avui hi ha a la porta de l'entrada d'Auschwitz). De manera que, teòricament, el robot està, per definició, *al nostre servei*. (O això hauria de ser: precisament la meua tesi d'avui és que –vist cap on sembla que van les coses últimament– més aviat serem *nosaltres els esclaus dels robots*, contravenint l'etimologia.)

Però continuo amb la història. Després de Londres, el «robot» va passar a Nova York, on a l'*Exposició* («World Fair») que es va muntar el 1939 amb gran parafernàlia a Queens (en un antic femer, a tocar del que ara és la pista de tennis de Flushing Meadows), a la paradeta (*l'stand*) de la General Electric hi va aparèixer un robot que, amb aspecte de Bibendum (el ninot Michelin, però metàl·lic, feixuc i amb forma de brioi), feia de, això, minyona. Servia begudes, feia reverències, fins i tot feia veure que deia coses, i així. Vist ara, ens faria riure compassivament. Però vet aquí que un parell de nois de secundària de Brooklyn, jueus, exrussos i proletaris, hi van anar a veure-ho un diumenge amb son pare i, ostres, en van rebre un impacte que encara ens dura: un d'ells, un tal Isaac (dels Asimov, d'un poblet de Rússia) es va posar a escriure frenèticament un llibre on, després d'identificar-s'hi (*I, robot*), en va

BUTLLETÍ DE  
L'ACIA  
NÚMERO 36  
PRIMAVERA-ESTIU DE 2006

dir) esmicolava els nous manaments que havien de seguir els seus congèneres (els japonesos li ho agraeixen, ara que és mort, batejant Asimo un dels seus antropomòrfics ginys caminadors/pujadors d'escalas). L'altre, l'escassa resistència a la «sangria» del qual jo personalment vaig poder comprovar (en un restaurant de Pedralbes on vam tenir la desgraciada idea de portar-lo), és en Marvin (de cal *minski*, perquè era Minsk, a la Rússia blanca, d'on venien), que tot seguit després de veure el cònic i rabassut robot de la GE va començar a imaginar màquines *intelligents* «artificials» (ell diu que només se li va ocórrer a ell, això). Veient vídeos del pobre Robbie, com em sembla recordar que es deia el robot, tan sapastre, costa d'entendre que una *cutreria* d'aquestes característiques despertés tan inflamades passions. (En fi, potser no costa pas tant si es recorda com xalàvem els pobres nanos dels claustrofòbics anys 1950 franquistes quan anàvem a la fira de Barcelona a allucinar i a fer collita de prospectes en paper couché i fotos en color, i a atiparnos d'exòtics i colesteròlics *frankfurts* que hi feien i venien a la *plça de l'Univers* –ni més ni menys– de Montjuïc: tot el que es veia allí en un dia no s'havia vist mai, ni es veuria, en tot un any –o en la vida–; potser per això és fàcil d'entendre que un nano immigrant arribat de poc a Brooklyn, psicològicament aclaparat i superat, «veies el món» o el futur «per un forat» –expressió catalana, per cert també provinent del món de les «exposicions», en aquest cas del segle XIX, concretament dels *diorames*, llavors de tecnologia (francesa) punta.)

Els qui tinguin un bon QuickTime instal·lat, i prou xafarderia per anar veient els clips animats que subministren els laboratoris, han degut poder comprovar recentment com hi ha, com més va, més robots que fan de tot. Segur que n'heu vist que corren com possessos. Per exemple, l'escocès –i alemany– RunBot, que fa un pam i mig d'alçada però corre com un boig sense caure ni agafar artritis, i tot amb un programa que el controla que, de tan senzill, fa riure. O el RiSE de Carnegie Mellon, que es dedica a pujar arbres com si el perseguís un cobrador d'hipoteques (la raó d'haver dissenyat un robot que grimpa pels arbres o per les parets és una cosa que, de moment, se m'escapa).

O el BigDog, de Boston Dynamics, que és un quadrúpede que carrega, de moment, més de quatre roves de pes (quan saps que

és un encàrrec de DARPA ho entens tot més bé: vol dir que els militars volen inventar la mula.) O els robots japonesos que, a falta de Ronaldinhos, juguen a ser, prou bé, per cert, jugadors de futbol. O els lluitadors de sumo, tan poc antropomòrfics (potser millor!) i tan bons, eficients i equilibrats. O els escarabats mecànics *Insbots* (belgues) que, sense que calgui que tinguin una aparença gaire versemblant, s'introdueixen en un ramat d'escarabats autèntics i, a base d'induir-hi accions i feromones, són capaços de modificar la conducta individual i col·lectiva del grup (una quinta columna *escarabàtica!*). O, encara, la munió de «robots» connectats al PC de casa, multifuncionals i no necessàriament humanoides ni industrials (com al Japó) ni militars (com a Amèrica), que promou el govern sud-coreà perquè els seus 48M ciutadans, mentre s'entretenen, acabin essent, el 2013 (això és el que pretén el govern), els més tecnificats del planeta. Sovint aquests robots (que solen costar de l'ordre del 500 dòlars i viuen connectats a la xarxa) es dediquen a voltar per la casa fent tota mena de coses, des de aspirar o fregar el terra fins a escombrar o fer llits, però un d'ells, que fa tres pams i es diu Júpiter, és capaç de parlar i cantar, caminar i ballar, reconèixer veus i cares i, gràcies a una cara en rotació i a uns ulls de dimensió variable, expressar (=simular) «emocions». Això ja ve a assemblar-se al Toyota Pod que remenava la cua (vaja, l'antena de la ràdio) quan sentia que venia l'amo (per fer-lo content, se suposa) i que jo vaig comentar aquí un dia, potser ho deveu recordar, perquè un periodista que m'entrevistava se'n feia un tip de riure («com són els japonesos!», repetia) mentre jo pensava com, d'aquí uns anys, tots serem japonesos.

### *Els japonesos, són humans?*<sup>6</sup>

El que m'interessa ara, a mi particularment, és anar seguint què fan els japonesos en el camp aquest dels robots perquè, pel que es veu, no sols hi tenen la mà trencada sinó que van dibuixant unes tendències sobre les quals no em sé pronunciar (si bé el que hi veig, francament, em fa una certa basarda). Pel que hi constato, els successors lògics del gos Aibo i del caminador/saltador Asimo no eren pas, només, els «robots-soldats» que ens anunciaven els militars americans ni, més pacíficament, els robots criats/ajudants

<sup>6</sup> Demano perdó pel manlleu que, amb aquest títol, li faig a en Miquel Barceló, que al darrer número es demanava si els enginyers ho érem, d'humans.

(com ho suggereix l'etimologia) que alguns imaginàvem –jo més aviat pensava en assistents de tipus gerontològic– sinó tota una colla d'«objectes» o «gadgets», més que no pas pròpiament robots, que aparentment plasmen necessitats sentimentals que jo no sospitava que fossin tan fortes i que es fabriquen i venen –i molt– en cases de joguines.

Preneu per exemple el cas del Scoty (que vol dir, potser massa forçadament, 'Smart Companion Operating TechnologY'), que fabrica l'empresa de joguines WowWee i que s'anuncia a les botigues com a *minyona digital*. Scoty fa tres pams d'alt, no és (gaire) antropomorf d'aspecte i passa per «company» i es ven com a tal (i ell mateix s'hi presenta, explicant a qui ho vulgui escoltar els «seus» gustos i la seva «manera de ser»), amb les següents ofertes: és capaç d'endegar i endreçar tot el munt –sovint, caos personal– de fitxers, fotos, vídeos o música que hom té a l'ordinador (els segueix la pista, els localitza, els desa, els ordena, etc.), cerca i reproduïx allò que li demanem, per vague que sigui com ho fem (per cert, tot per ordres donades amb la veu, sense haver de teclejar gens enlloc), troba i/o grava programes de TV, llegeix els e-mails (que prèviament ha examinat per veure si ens interessen), escolta el que contestem i tot seguit ho posa per escrit i n'envia el correu de resposta. També fa fotos del que li diguem i fins i tot (per citar una més de les mil coses que pot fer) diu l'hora quan la hi demanem (o ens recorda allò que li diem que ens recordi perquè segur que ens en descuidarem)... i, en fi, infinites coses més, a quina és més imaginativa (i predictable). (La tecnologia que duu a dintre no és japonesa, però; és de la Philips d'Alemanya: tanta fúria endreçadora ja és veu que havia de ser germànica, com a mínim.) I tot sense cap teclat a la vista. No cal dir que, després d'haver conviscut una temporadeta amb l'usuari, la màquina ja en sap tots els gustos i costums, i manies. (Passa una mica com vaig explicar que passava amb el Toyota Pod, que es coneixia el conductor i el cuidava com si fos sa mare.) I com que del «rosse» neix el «carinyo», com diuen totes les sogres del món, només queda per veure quines relacions s'acaben establint entre l'usuari i la seva «minyona digital». Aquí ve el record inesborrable de l'«Eliza», el primer programa (als anys 1960!) que va fer creure, fins i tot als més *ràbids* escèptics, que era un psiquiatre de debò, que «ens entenia»; com se sap, de fet, Eliza

era un mer analitzador lèxic, més poc «intelligent» i més ruc –des de tots els punts de vista– que ni fet d'encàrrec (realitat que, però, NO s'amagava a ningú), i que, això no obstant, produïa efectes fulminants, com aquell dia en què un escèptic rematat que volia deliberadament sabotejar el sistema –vull dir, fer que se li veiés públicament el llautó, la seva falta total d'intel·ligència– i que, quan en un gir inesperat aquest es va trobar encallat sense saber què fer amb la paraula «pare», va construir la frase prefabricada «Parla'm del teu x» [on x = «pare»] va ocasionar insospitadament (per als presents) una insòlita plorera en el fins-llavors-escèptic usuari que, deia: «com ho ha sabut, aquest, que amb el meu no ens parlem?!» Des de llavors sabem que la tendència a lligar-se sentimentalment a qualsevol cosa és irresistible. (De fet, això ja ho sabíem: quan passa amb objectes se'n diu *fetitxisme*, i quan passa amb coses vives se'n diu *animals domèstics*, que ja fa temps que estan inventats). És a aquesta sorprenent i aparentment irrefrenable adhesió a éssers no humans al que em vull referir ara, perquè, si no som nosaltres, els *professionals* de la IA, vivint com vivim del negoci aquest de la «Intel·ligència artificial», els que *observem* el que passa i els que *hi meditem*, amb calma i profunditat, no sé pas *qui* ho farà. Penseu que *cal* (= se'ns demana) que sapiguem què està passant, encara que sigui per mera aproximació. (I cal també que tinguem una resposta preparada, encara que sigui mentida, per treure'ns del damunt els periodistes esbiaixats, pesats i simplistes que sempre ens aniran al darrere perquè els confirmem els seus prejudicis!).

I perquè ens comencem a sorprendre de les coses que passen (i no sabíem que havien de passar), aquí en van algunes. De moment (en el moment que escric això, vull dir) ja tenim el següent repertori, gens exhaustiu, de robots-raretes –tots japonesos, de moment– (amb un sentit record per al primer «virtual pet», el faust Tamagotchi): el FurReal, un ximpanzé força verídica, amb pèl, tacte i caiguda d'ulls a més d'una proclivitat infinita a deixar-se tocar; el Butterscotch, un poni que segueix la llum o gira el cap per veure qui li ha fet pessigolles; l'IDog, un gos robòtic prou diferent de l'Aibo que, com el seu parent l'ICat (un *interactive music companion*), s'illumina per dintre –té el cos translúcid i ple de LEDs– quan se suposa que «expressa» alguna cosa (m'imagino que una de les gràcies és que el seu propietari passi hores per esbrinar què carai

deu estar «expressant», l'animaló). (I no us recorda, per cert, això el somriure o ganyota «dibuixats» amb els LEDs verds o vermells del capó amb què el Toyota Pod expressava l'aprovació/satisfacció o, respectivament, la desaprovació/ira/odi pel que veia que (li) feien els altres conductors?). Segons els venedors de joguines de mitja Amèrica, que ja s'hi freguen les mans i els ve salivera, els USA són, pel que fa els *gadgets* d'aquesta mena, pel cap baix, com el Japó, cosa que afirmen sobretot després de veure com alguns compradors americans de les aspiradores «intelligents» Roomba (ep, no és cap broma: 1,5 milions venuts des del 2002 a 300 dòlars pel fabricant, asimovianament batejat IRobot) són capaços de veure «trets intelligents» –i fins i tot una «personalitat definida i característica»– en les evolucions i tombarelles que fa la seva criatura (que, no ho oblidéssim, és una simple aspiradora en forma d'OVNI d'aquells que abdueixen). D'aquí al següent producte ja només hi ha un pas, que, naturalment, ja s'ha fet: per a 100 dòlars qualsevol que ho desitgi pot tenir l'Amazing Amanda, una nina amb edat i posat (pre)adolescent que reflecteix els seus «sentiments» amb les expressions i moviments de la cara i del cos, que reconeix la veu dels amics i que parla amb un vocabulari i sintaxi prou acceptables per l'ocasió. (I no, no és el que algun mal pensat hi pot imaginar: és una nina perquè hi juguïn les nenes... o això diu el fabricant, que, maldestrament, ha triat de dir-se, ni més ni menys que *Playmates Toys*.) Potser semblant és el Primo Puel, un nino també japonès que passa per un nen d'uns cinc anys i que va proveït amb sensors i cinc «nivells de felicitat» i una expressió de cara que canvia segons que convingui; la cosa curiosa d'aquesta joguina és que estava pensada, com l'anterior, per a nenes però ha tingut un èxit indescriptible entre les dones japoneses, especialment les madures que tenen un fill gran que fa poc que s'ha emancipat i se'ls n'ha anat de casa –es veu que al Japó els joves se'n van de casa, un dia. (I, vist aquest antecedent, qui sap què passarà amb les Amazing Amandas prepúbbers? Cal recordar que al Japó els *salary men*, com en diuen dels empleats d'aquests de 12 hores al dia, de nit van molt cansats, i ja és tradició que comprin joguines infantils –o videojocs o *mangas* o *ànimes* o el que calgui– per relaxar-se.)

I què passarà a la culta Europa? (vull dir, l'Europa tan creguda

de ser infinitament més elegant i madura que els USA o el Japó, que són tan –segons la nostra autoritzada opinió– primaris, infantils i xarons.) I com serà el nostre futur quotidià i dels nostres fills? (I com s’ho prendran els periodistes que ja saben com serà el futur?)

Fóra prou irònic que els nostres professionalitzats, seriosos i científics esforços per estudiar i crear la «intelligència artificial», la IA, acabessin tenint com a llegat principal una sèrie completa de super-tamagotchis (això sí, a l’abast de tothom, i democràticament fets perquè els depressius de totes les races i condicions s’hi relaxin).

Viure, per veure!<sup>7</sup>

Barcelona, 22 de maig de 2006

<sup>7</sup> Els romans feien broma d’aquesta dita i, jugant amb «vi-vere» i «bibere», venien a dir: «Ja que vivim, bevem». Però lluny de mi aconsellar-vos que begueu. En aquests temps que corren, potser és millor que us compreu un robot.

PD1. *Volia continuar l’article amb una secció si fa no fa titulada «En què consisteix «estimar» (i «ser estimat» per) una màquina» aprofitant reflexions que he llegit recentment en treballs universitaris diversos (de Califòrnia i també d’Europa) així com en coses de la interessant Sherry Turkle de l’MIT sobre les relacions que s’estableixen entre persones i màquines intel·ligents, però ho deixo per a una altra ocasió perquè em sortiria massa llarg tot i perquè encara no tinc les idees prou clares jo mateix.*

PD2. *També deixo per a un altre dia les elucubracions que vaig prometre sobre el futur de la informàtica i de nosaltres mateixos, i de la famosa «singularitat» a què suposadament ens atansem, perquè el panorama d’aquesta qüestió canvia cada dia que passa i caldria deixar-ho reposar una mica més. (Afortunadament, veig que gent assenyada com els editors de Nature hi ha començat a ficar cullerada i posar seny; vejam si hi ens acabem aclarint, tots plegats.)*





## Crònica de robots (continuació)

**A**QUESTA columna surt a intervals de **tres mesos**, si fa no fa, en data no precisa. Aquesta circumstància fa que (1) quan escric no sé mai si ha passat una cosa o encara està per passar (exemples: (a) quan el lector llegeixi això, ja s'haurà fet la xerrada del 10 d'octubre d'en Wolfgang Wahlter a CosmoCaixa sobre els «*Cotxes intel·ligents (amb ulls i orelles)*»? (b) i jo ja he fet la meua xerrada de **Perpinyà** del 26 d'octubre<sup>8</sup> per celebrar els 50 anys de la IA?), cosa que em dóna una meravellosa sensació d'intemporalitat, d'anar «penjat», que s'adiu força amb el meu estat de despistament congènit. I (2): com que en tres mesos han passat coses, jo em veig en certa manera obligat a fer-hi de cronista i passar comptes, més que res per posar al dia i en comú les referències que tenim el lector i jo, i així poder continuar sobre una **base compartida** les meves particulars elucubracions més a llarg termini.

Per això vull esmentar –i no hi entro gens, perquè se n'ha parlat força– les encoratjadores notícies del març que venien d'Alemanya i que parlaven d'una màquina d'escriure controlada exclusivament pel cervell, les del maig, del Japó, sobre una mà artificial que fa figures de paper (retallant-lo amb tisores!) també controlada pel *pensament* i, sobretot, les del 13 de juliol, a Nature, molt comentades, sobre la «resurrecció» de Matthew Nagle, un antic esportista ara tetraplègic –per apunyament al coll– que ha recuperat els seus *moviments* de cames i braços: ara els torna a controlar directament des del seu cervell, amb el qual, a més, pot obrir el seu correu electrònic –i treballar-hi– o engegar la tele. És molt gratificant veure com la Intel·ligència Artificial (la IA), quan es dedica (cosa que fa de tant en tant) a feines no bèl·liques o d'interès no estrictament

BUTLLETÍ DE  
L'ACIA  
NÚMERO 37  
TARDOR DE 2006

<sup>8</sup> L'autor es refereix al 9è Congrés Internacional de l'ACIA de 2006. (n. de l'ed.)

SI AMB LA IA (LA INTEL·LIGÈNCIA ARTIFICIAL), COM A DISCIPLINA CIENTÍFICA I TECNOLÒGICA, SOM CAPAÇOS DE FER COSES COM LA QUE S'HA FET AQUEST ESTIU, QUE UNA PERSONA PARALÍTICA COMENÇA A CAMINAR I A FER VIDA NORMAL, POTSER ENS PODREM COMENÇAR A SENTIR CONTENTS DE DEDICAR-NOS-HI, COM EL METGES HO ESTAN DE SER-HO I SALVAR VIDES (TOT I QUE TAMPOC NO S'HA D'AMAGAR QUE BONA PART DE LA IA NO ESTÀ DESENVOLUPANT PRECISAMENT PROJECTES PER SALVAR VIDES NI MILLORAR-LES).

estatal, ens dóna certes esperances –realistes– sobre un futur més aviat benèfic.

I precisament del futur, de l'esdevenidor, volia parlar (dic «volia»). Ho havia promès, però com que ja en parlaré (o ja n'he parlat?) a Perpinyà el 26 d'octubre al **congrés de l'ACIA**, ho deixo per aquella ocasió, que crec més adequada i on «el futur» queda més naturalment subsumit, espero, en el tema de la meva comunicació (que vaig titular, com deveu saber, «**De la intel·ligència artificial a l'estupidesa natural: com passar d'una a l'altra en només 50 anys**»). És una xerrada feta per celebrar el **cinquantenari de la «IA»** (la d'aquell llunyà Dartmouth de l'estiu del 1956) i el contingut de la qual, si s'hagués publicat aquí, hauria estat aproximadament allò que havia promès. (Ara bé, contravenint la meva tradició àgrafa i àcrata potser ho acabaré escrivint, i acabarà publicat en aquesta columna. No prometo res.) De moment, per avui, em conformo –i m'alleujo– deixant de banda la ditxosa «singularitat» de Kurzweil i el grandios tema de com és «esperable» que sigui «el futur» (ui!) i tot això. En canvi, continuaré repassant algunes novetats en el camp dels **robots** més o menys simpàtics (i actualitzant en alguns casos coses ja dites en números anteriors). La veritat és que, inicialment, també volia entrar breument, cap al final, en un tema d'actualitat que pot portar cua: les **neurones mirall** (les «Aniston», com se'n diu en broma), de què ja vaig parlar en aquesta mateixa columna fa dos o tres números. Hi volia tornar, encara que fos ràpidament i amb la meva proverbial frivolitat. Però he decidit que la cosa quedava tota massa llarga, i me n'he estat. Fins a la pròxima columna, potser. El que queda escrit aquí, doncs, va tot de **robots**, i, tenint en compte que ja en van unes quantes columnes, potser m'encavalcaré amb mi mateix, però ja em dispensareu. (I si voleu varietat, us remeto a Perpinyà.)

### *Les «gestes» de la IA aquesta dels cinquanta anys*

Aquí ja n'hem parlat sovint, de les fases que ha passat la IA: l'optimisme inicial (1956-80), l'«**hivern**» (1980-2000) i la «**primavera**» (ara). Avui tothom se sent sortint del túnel (o de l'hivern, vaja, en termes meteorològics), i tothom parla, en comptes de **IA** («Intel·ligència artificial»), de **IR** («Intel·ligència real») –o, com segur que

diria en Joan Saura, «intelligència **de debò**»– o d'«Informàtica cognitiva» o de «Human-level Intelligence» o INH («Intelligència **al nivell de la humana**») (no feu servir les inicials HLI, però; em temo molt que Microsoft les ha patentat). Algunes coses semblen clares: avui les nostres màquines tenen més memòria que mai no hauríem imaginat (i gràcies a això tenim, per exemple, coses com Google) i, com a conseqüència, tenim «*companys intelligents*» que ens endrecen els nostres caos particulars (per exemple, el robot que endreça la casa –o el que en queda– després d'una nit de timba o de trompa casolana amb els amics), o que, tal com vam explicar aquí fa uns números, ens porten el cotxe, ens organitzen la vida, parlen amb nosaltres (i/o parlen per nosaltres), ens recorden les coses que no tenim al cap i que, ja per fer (i pel mateix preu), a més ens escombren la casa. Tots aquests sistemes passen el test de Turing en certes condicions (de fet, els fabricants ja els dissenyen i construeixen pensant-hi). Per tenir, tenim fins i tot un robot (d'Stanford) que, amb un tornavís i el típic prospecte laberíntic, sap muntar mobles d'IKEA (com sona!), cosa que evidentment està fora de l'abast de la majoria de mortals. (Si el test de Turing es fes a IKEA, qualsevol home en sortiria clarament derrotat davant la màquina aquesta.) I tenim un «majordom electrònic» que ho fa tot, des de parlar amb nosaltres i donar-nos conversa (i aguantar-nos, que ja té mèrit) fins a encarregar menjar del gos quan cal. (Aquest majordom, anomenat Chancellor, és de Hecht-Nielsen, vell conegut nostre, que també ha fet detectors de frau a Internet, i del qual tornaré a parlar una mica més avall.) Alguns d'aquests robots (per exemple, els comandaments a distància que ens reconeixen la veu) no són nous, però ara són molt millors, més sofisticats i molt més econòmics (i, per tant, els anem trobant pertot).

*Dels «autònoms» als «cooperants» passant pels «predestinats» i els «confabulats»*

Més robots? Oitant! Heu-vos-en aquí uns quants de selectes. Tenim el *cotxe* (un quatre-per-quatre d'Stanford que sembla un camió blindat) *sense pilot*, totalment autònom i automàtic, que va amb GPS tot sol pel Mohave i que ha guanyat, ara no fa un any, el premi americà de travessar deserts sense perdre-s'hi (però és de DARPA!

ELS GURUS DE LA IA (INCLOENT-HI MICROSOFT) DE VEGADES JUGUEN MASSA AMB EL LLENGUATGE COMÚ: BATEJAR SEGONS QUINS MODELS AMB PARAULES JA RECONSGRADES I SOLEMNES COM ARA CONSCIÈNCIA, PREDESTINACIÓ O CONFABULAR ELS COL·LOCA A LA BANDA DEL NARCISISME FRÍVOL I NAÏF, «GRACIOSET», DE QUÈ SOVINT SE'LS HA ACUSAT DES DE FORA (VULL DIR, DES DE LA CIÈNCIA «SERIOSA»). O ÉS QUE NO FA UNA MICA DE VERGONYA, PER EXEMPLE, PARLAR (I, A MÉS, TEORITZAR) DE PREDESTINACIÓ A PROPÒSIT D'UN ROBOT QUAN EL QUE ES VOL DIR NO HI TÉ RES A VEURE (I ÉS UNA PURA TRIVIALITAT)?

ui, els militars!), o el robot francès que, amb mires benèfiques (aquest sí) es passa la vida badant, **mirant** què hi passa (o, més aviat, què no hi passa) en una piscina, per detectar si el comportament [del cos] d'algú que s'hi banya n'indica una indisposició o ofec (i salvar-lo, és clar). Més? Microsoft ara parla de la «**predestinació**», i no es refereix pas al concepte teològic sinó a la propietat de *preveure* («ensumar», diríem) on deu estar anant un individu (= la seva «pre-»destinació) (ha! quina gràcia), tenint en compte les seves predisposicions observades (els seus «gustos» i tendències) i la història dels seus viatges anteriors. Estirant això una mica, potser aquest nou concepte informàtic (la «predestinació», uff!) ens pot dur a una aproximació d'allò que en diem «intuïció» (per exemple, la dels metges experimentats que, malfiats del que «veuen», de cop «ensumen» cap on va el diagnòstic... i l'encerten [o no!]) (i compte que Microsoft no l'hagi patentat, també).

Tot això per no parlar de l'interessantíssim estudi (de debò, sense broma) d'IBM que continua, a la Politècnica de Lausana, sobre el (primíssim) *còrtex cerebral* (el projecte «Blue Brain» del qual ja hem parlat aquí als dos darrers números), o de les moltíssimes coses que aprenem de les *tomografies cerebrals* de ressonància magnètica, cada vegada més localitzades i funcionals (vull dir, cada cop més precises espacialment i més precisament correlacionades amb factors exteriors, com ara *imatges* percebudes, *estats d'ànim*, etc.), un subproducte de les quals és per cert la «resurrecció» d'en Matthew Nagle què he esmentat al començament.

I què dir de les noves teories de la *intelligència*? Jo ja he parlat aquí abastament de la d'en Jeff Hawkins. Ara potser toca parlar també de la teoria de la «*confabulació*» de Robert Hecht-Nielsen (personatge que he citat més amunt), que ha incorporat en un software propi (recentment presentat en un simposi IBM) i que permet, per exemple, anar seguint informacions de diari (en el seu cas, del Detroit Free Press), «fer-se'n una idea» (per dir-ho en termes familiars) i, i aquí ve la gràcia, comentar de què va un assumpte, especular-hi («confabular») (ha! quina gràcia), i fins i tot gosar aventurar prediccions sobre cap on pot anar la cosa en el futur... o, si més no, fer comentaris «intelligents» que demostrin que sabem de què parlem, com ara: «ui, si no hi intervé Bin Laden, la cosa pot no ser greu», i trivialitats per l'estil. Si fa no fa, com

hem fet tots els homenets i les donetes quan hem volgut quedar bé davant d'un desconegut a qui cal impressionar –per exemple la sogra (l'exemple arquetípic). Si hi hagués cap test de Turing per *semblar llest i interessant*, el programa nielsenian segur que el passaria, o això pretén.

I, mentrestant, de robots n'anem tenint de tota mena: n'esmento només un parell, perquè em semblen especialment curiosos i interessants. Primer, un «chatbot» holandès que, amb molt d'èxit popular, per cert, no para fins que deixes de fumar: et deixa convençut i derrotat. Si l'esmento no és pas perquè no em sembli bé, sinó perquè se m'esborronen els cabells només de pensar que podria fer un publicitari pesat –o, ja no diguem, un fanàtic– matxucant-nos el clatell fins que li compréssim la cosa que ens intenta vendre: només ens calen encara més *persuasius* subtils! El segon robot que volia ressenyar són **dos** (com la Santíssima Trinitat, que deia aquell): es tracta d'un parell de robots amb camp de visió reduït que es veuen obligats a «cooperar» per passar una porta (només la meitat de la qual veuen). Hom s'imagina el diàleg, una mica tipus parella còmica Pryor-Wilder: «què veus? és un tros de porta, això? més a la dreta, no, més a l'esquerra, ara!», etc. L'interessant és que és un dels exemples més precoços de *cooperació* (que, no cal suar gaire per descobrir-ho, és una de les claus de la robòtica futura, igual com ha estat la base de les societats humanes i de la divisió del treball). La qual cosa ens porta a plantejar-nos la següent qüestió.

### *Ètica per a robots*

Pot semblar llunyà, això, però, creieu-me, es planteja ara mateix entre nosaltres. Jo ja n'he parlat alguna vegada aquí mateix. En alguns casos, tothom accepta que tot vagi a càrrec d'una màquina, que suposem infallible. Ho fem tots cada dia, quan viatgem en avió. Pràcticament tot el vol es fa sense intervenció humana (que és més aviat cosmètica: no s'entendria que et deixessin a tu, passatger, sol a dintre un avió sense algú que no fes veure que hi entén i que ho té tot controlat). Curiosament, tothom accepta el fet (només cal posar un senyor d'uniforme a la cabina, pagar-li entre tots un sou, i enlaire!). En canvi, imagineu què passaria si un avió

I NO ENS HAURÍEM DE COMENÇAR A PREOCCUPAR DEL FET, CADA VEGADA MÉS FREQUËNT, QUE ENTITATS COM MICROSOFT PATENTIN CONCEPTES GENERALS, COM ARA «MÒDUL DE CONSCIÈNCIA», «MECANISME DE PREDESTINACIÓ» O «INTEL·LIGÈNCIA AL NIVELL DE LA HUMANA»? ATENCIÓ, PERQUÈ NO SERÍEM ELS PRIMERS A CAURE-HI: EN BIOINFORMÀTICA JA HI HA HAGUT INTENTS DE PRIVATITZAR GENOMES I PROTEÏNES.

sense pilot (de fotografia aèria o meteorològic, no cal que sigui d'espionatge o bombardeig) caigués sobre una zona habitada i fes mal: la gent diria «com és possible, haver deixat anar una cosa així, sense control humà?» i «qui en paga els efectes?», i tot sense que hi hagi cap diferència essencial amb el cas de l'aparell amb pilot decoratiu (tret del famós i socorregut «error humà», que ara ja no s'hi pot al·legar). Com ja vaig dir, no és aquest un assumpte que ens hagi de preocupar, perquè per això ja hi ha companyies d'assegurances que discuteixen afanyosament el tema des fa temps, i s'hi juguen molts *calés*.

Hi ha casos, però, que preocupen, a més de les asseguradores, a la gent èticament predisposada. Per exemple: si un estri domèstic «intelligent» (un aspirador Roomba d'aquests que tenen tant d'èxit, posem per cas) ens fa malbé la moqueta o, Déu no ho vulgui, ens fa mal a nosaltres, a qui carai hem de reclamar? No era potser el fabricant qui s'havia d'assegurar que tal cosa no passaria adoctrinant prèviament el seu estri sobre el que s'ha de fer i el que no? I les nines «animatròniques» japoneses de què parlàvem al número anterior, què? Se'ls hi ha d'incorporar un mecanisme que les bloquegi en presència d'un menor, o d'un sàdic addicte (per evitar de fomentar-li el mal hàbit) (o potser és al contrari? hauríem, potser, de regalar-los *animatroninos* perquè mentrestant no fessin mal a ningú «real»?).

En fi, que ja es veu que la cosa té corda, i ara podem plantejar-nos de començar a estirar-la. I vet aquí que ja hi ha algú que ho fa! Aquest any mateix ja s'ha creat (i reunit a Palerm) l'EURON (=EUropean RObotics Network) que, a semblança del Pugwash (el grup de físics nuclears propau creat el 1957) es planteja temes com aquests. I més: la revista *The Economist* hi ha dedicat un suplement (el juny del 2006) titulant-lo significativament «Can robots be trusted?» i subtitulant-lo «The challenge of making sure robots do no harm». I encara més: un cert nombre de recercadors robòtics japonesos plantegen una rebel·lió oberta contra els Manaments de la Llei Robòtica d'Asimov, que consideren caducs i contraproductius. (Veurem com acaba tot plegat. Aguanteu la respiració!)

### *Els robots simpàtics*

De moment, aliens a tot aquest enrenou ètic, els robotaires continuen fent la seva feina, fer robots, i en fan de tota mena, com apuntàvem més amunt. Un factor, però, acaba essent obvi per a tots ells. (I deixo de banda, avui, la qüestió de la «consciència»: han de ser, els robots, conscients dels seus actes, o de la seva pròpia existència? Ho deixo per un altre dia.) Tenint en compte que els robots (recordeu-ne l'etimologia) treballaran per a nosaltres o, si més no, al nostre costat, no podem concebre'ls ni construir-los de tal manera que ens espantin, o ens irritin, o ens facin sentir incòmodes, etc. Per exemple, no ens poden venir a preguntar cap cosa (ni tan sols sobre si estem bé, o si volem res, o si ens poden ajudar) si ens vénen per darrera, o des d'una zona poc il·luminada, o si ens ho demanen –mal que ho facin sol·lícits i suaument– quan nosaltres dormim o quan és «evident» que estem cansats o indisposats o no estem per orgues. En tots aquests casos ja es veu que la resposta de l'interpellat (el presumpte ajudat) pot no ser necessàriament favorable. I, com es pot sospitar, d'evitar de fer els robots antipàtics a fer-los simpàtics només hi ha un pas. I en això estem. A més de dotar-los d'una mica de «consciència» (que ja es veu que hi pot ajudar), no cal dir que, perquè els robots resultants siguin modestos, respectuosos, agradables i «tinguin conversa» segurament caldrà crear –com ja es plantegen segons quins fabricants– escoles «de bones maneres», i enviar-los-hi (a l'època de les nostres àvies d'aquesta mena d'escoles en català se'n deia *de senyoretas*; en anglès se'n diu *finishing schools* i han tingut, i tenen, qui-sap-l'èxit al Japó, on ara podran continuar la tradició canviant senyoretas per robotets poc «educats» –per no parlar dels robots directament antisocials, que n'hi haurà).

### *Enamorar-se d'un robot*

I si els robots se'ns fan agradables, no podrà pas passar que ens en enamorem? Doncs sí, i de fet ja ho estem fent... almenys des d'aquell 1967 que jo evocava l'altre dia en què més d'un es va sentir comprès i «estimat» per l'estupidíssim programa Eliza, com quan Robert Epstein, avui IA-aire conegut, va dir (a l'Eliza) «Els



AMB ELS ROBOTS, A POC A POC I CADA VEGADA MÉS, ENS ANIRAN PASSANT TOTES LES COSES QUE ENS PASSEN AMB LES PERSONES: ENS AJUDARAN, PERÒ TAMBÉ ENS TRAURAN DE POLLEGUERA, ENS CAURAN BÉ O MALAMENT, SEGONS EL MOMENT I LA CIRCUMSTÀNCIA, ENS CONVENCERAN O HI ESTARAN EN TOTAL DESACORD, ENS HI ENLLUERNAREM (I ENS N'ENAMORAREM) O ELS AGAFAREM MANIA, HI COOPERAREM O ENS DEIXARAN INDIFERENTS, N'HI HAURÀ DE MÉS BEN EDUCATS I TAMBÉ DE POC ÈTICS, ETC. ARA, AIXÒ SÍ, SEGUIM QUE ENS HI RELACIONAREM DE MANERA DIFERENT. SERÀ COM SI ESTIGUÉSSIM INVENTANT UNA NOVA ESPÈCIE D'ANIMALS DOMÈSTICS. I SERAN TAN DIFERENTS COM HO SÓN ELLS. O ALMENYS AIXÒ ÉS EL QUE ES POT PREVEURE, DE MOMENT, VISTA LA DIRECCIÓ QUE PRENEN LES NOSTRES EXPERIÈNCIES AMB ELS QUE HEM TINGUT FINS ARA.

meus pares no m'entenen» i l'Eliza li contestava: «Parla'm dels teus pares». El diàleg (segons que ho explica a «My date with a robot», al *Scientific American* de l'estiu passat) va continuar així: «És que no m'entenen, saps? No confien amb mi, amb el que sóc capaç de fer», i la resposta: «Ha de ser molt dur, això, per tu». Punt final. Enamorament total. I aquest home, en comptes d'haver-se'n desenganyat del tot (quan ha deixat de ser adolescent i n'ha sabut la trampa), va i es dedica a la robòtica! (potser enyorava que l'enganyessin carinyosament?). I en aquest punt estem. Tant, que l'estranya (estranya?) relació que inevitablement –sembla– s'estableix entre persones i màquines (i les similituds i diferències amb les interpersonals) són el tema d'estudi de la Sherry Turkle, una ex-acaniana de l'MIT (on s'està des del 1976 i on dirigeix des del 2001 la –agafeu-vos, que ve el títol!– *MIT Initiative on Technology and Self* que ella hi va fundar) i que no sols estudia el tema aquest (les relacions) sinó que també es dedica, més en general, a la connexió entre «coses» i «pensament». Com que encara no m'he llegit a fons res del que diu (la veritat és que em fa mandra), no sabria pas si recomanar-vos-la o no. (Si algú ho fa i m'ho explica li estaré eternament agraït, no cal dir-ho.) Pel que he vist, hi ha una persona de la Universitat de Barcelona que també es dedica al tema i que ens ho explicarà (o ja ens ho ha explicat?) cap a finals de gener del 2007, a l'IIIA. (Si quan llegiu això encara no s'ha fet la xerrada, us recomano que hi aneu; i si ja s'ha fet, ja ens ho explicarà algú... o tal dia farà un any.)

### *Com seran (els robots) i com som (nosaltres)*

Aquí, seguint el fil argumental, volia explicar algunes qualitats que ens caracteritzen als humans (i ens confereixen la nostra natural i entranyable *estupidesa*) i que, suposadament, **no** han de tenir ni tindran **mai** les intel·ligències no naturals que nosaltres fabriquem (vaja, fóra de bojós que les féssim com nosaltres!). I aquí hi entraven les «Aniston» (vull dir les *neurones mirall* o, si es vol, *neurones especulars*), però com que això forma part del que diré el 26 d'octubre a *Perpinyà* (o ja ho he dit?), ho deixo, si de cas, per a més endavant. I si ens hi veiem (a *Perpinyà*, vull dir) estaré encantat de saludar-vos-hi. I si ja ha passat i resulta que ja us hi

he vist, doncs estic encantat d'haver-vos-hi saludat. Jo visc en un món tan virtual que ja no sé si una cosa m'ha passat, em passarà o cap de les dues coses anteriors. És una mica com aquell que deia que hi havia *tres* menes de persones: les que coneixen el sistema binari, i les que no. (En una altra variant de l'acudit, eren «10» les menes de persones que coneixien, i les que no, el sistema binari.) En la versió, més nostrada, de Josep Pla, la paradoxa seria, de la manera una mica borgiana en què ell ho formulava, que «el món es divideix en coneguts, desconeguts, i *saludats*». En aquest context, tant si sou de la primera com de la segona classe, doneu-vos, com a mínim, per membres de la tercera. A Perpinyà, és clar, que és, com tothom sap –des del moment en què Dalí ho va descobrir–, el *centre del món*.

Barcelona, 22 de setembre de 2006



## *Bons desitjos per al 2007 i altres prediccions diverses*

**E**N un dels comentaris que la fundació Edge demana a diverses celebritats –i assimilats– de la Ciència cada any sobre un tema pretriats (el d’enguany era «Vostè, sobre què és optimista (i per què)?», la Pamela McCorduck explica que quan tenia disset anys li van encarregar –a ella, que no en tenia ni idea– que fes un *survey* sobre el que deien els promotors de la AI, l’*Artificial Intelligence* (els McCarthy, Minsky et alii). Hi va entrar com a noia impressionable i en va sortir impressionadíssima. El resultat? El llibre que tots vosaltres coneixeu: *Machines who think* (observeu-hi el *who*, que en gramàtica anglesa és un relatiu només aplicable a persones i això aquí ja suggereix –al títol mateix– de què anirà la cosa). Tot seguit McCorduck (la del 2007, no la dels 1960) explica que, un cop escrit el llibre se’n va allunyar (de la AI o IA) i pensava que ja no hi tornaria més fins que, fa relativament poc, algú li va dir que tornés a ficar-s’hi, que fóra interessant que comparés les seves impressions de llavors amb les d’ara. I ho va fer. I les que va treure dels primers contactes que va tenir, tan diferents de l’optimisme incommensurable i desbordat dels pioners que ella havia tractat trenta anys abans, li van fer pensar que el llibre que acabaria escrivint (que, per cert, ha acabat en forma d’apèndix a la segona edició de l’original) seria no pas un pamflet inflammat com el primer sinó més aviat una mena de necrològica de la IA com a camp. I en canvi, segons ens diu ara, s’ha adonat que el camp sí que potser ha desaparegut, però és que s’ha mort d’èxit, i ara, gràcies als *gadgets* i a Internet, s’ha difós i és pertot arreu!

Ah, i una altra cosa: el nostre inefable amic Ray Kurzweil (el de

BUTLLETÍ DE  
L’ACIA  
NÚMERO 38  
HIVERN DE 2007

la «singularitat») ens recorda, des del seu article al mateix informe d'Edge, que la potència de càlcul i de memòria dels ordinadors serà en poques dècades tan i tan gran que la situació serà totalment inconcebible (això és justament la «singularitat» que tant li agrada de recordar-nos) i, per tant, d'una predictibilitat vàcua i nul·la, a més d'absurda. (Bé, això és el que deia Kurzweil fins ara; a Edge hi afegeix que d'aquí vint anys tots els nostres problemes actuals estaran resolts, precisament gràcies a la informàtica... i tot seguit passa a ocupar-se del residu de problemes que potser [encara!] ens quedaran: la vida eterna –vull dir, la longevitat *sine fine*–, l'energia abundant i gratuïta i la desaparició de la pobresa, problemes tots tres que la informàtica no ens haurà resolt demà –vull dir, en «vint anys»– sinó demà passat [presumiblement >20 anys]). Bé, no vull entrar en el món una mica autista de l'amic Ray. Sobre el que diu de la potència de les nostres màquines té tota la raó (i de fet no és pas l'únic a haver fet l'observació) i pel que fa a la nostra capacitat de predicció segurament té raó,, si bé no calia pas que insistís en aquest tema, el de la nostra ineptitud per predir el futur: aquesta la tenim constatada des de fa força temps, i jo mateix hi he insistit sovint aquí mateix.

No fa gaire es va publicar un estudi –potser sorprenent (o no)– que, fent servir això que se'n diu la *meta-anàlisi* (l'anàlisi i comparació de mètodes i resultats de tot l'escrit sobre un tema), que el mètode fins ara més acreditat per predir el futur, el «Delphi» (que consisteix a reunir experts perquè, raonadament, facin la seva predicció [diríem] «il·lustrada»), té un valor predictiu aproximadament de zero. És a dir, que, vist el que ha passat després, qualsevol no-expert ho hauria endevinat tant com ells fent servir daus o responent les preguntes a l'atzar, o, dit d'una altra manera, que el grau d'encert dels experts no supera el d'una predicció aleatòria.

### *Ah, les discontinuïtats!*

I és que dues de les coses que ens fan equivocar més en la nostra extrapolació ingènua del present cap al futur és que: (a) no tenim en compte que segurament hi haurà discontinuïtats; i que: (b) no tenim en compte com ens hi adaptarem nosaltres (factor que, sovint, és més decisiu que el factor tècnic en si). M'explico: [sobre

la meua afirmació (a):] Si fa només una dècada ens haguessin dit a qualsevol de nosaltres que «d'aquí deu anys el jovent sentirà música pel carrer amb un «Walkman» molt més petit i amb una capacitat per a centenars de cançons» hauríem pensat que probablement sí, que l'únic que calia era arribar a fer uns superminidiscs que hi cabessin. El que segur que no hauria pensat ningú és que tot això es fes –és l'iPod– sense cap disc a dintre (!) sinó tot desat en memòria sòlida... I espera't que la cosa no acaba aquí! El mateix fabricant se superarà i anunciarà l'iPhone –i per cert, el fabricant en qüestió serà l'Steve Jobs (sí, aquell que va «fracassar» amb Apple en la seva lluita contra l'«imperí» [vull dir MicroSoft] i va haver de sortir-ne per la porta del darrere). L'interpellat en qüestió a mitjans anys 1990 no s'hauria pas cregut res de res de la nostra predicció retroactiva (sense disc? com pot ser! i en Jobs? mare-de-Déu, si en aquella època se'n cantaven les absoltes!). I és que hi ha hagut discontinuïtats (l'eliminació –en lloc de la ultra-reducció– del disc, o la resurrecció del «mort» Jobs) i, si les prediccions per extrapolació estan destinades a fer llufa, no us dic res de les discontinuïtats: són, gairebé per definició, miraculoses i, per tant, encara més impredecibles i abocades no a fer el ridícul sinó a desbarrar del tot i aleatòriament.

### *I, a sobre, nosaltres ens adaptem a tot*

La segona qüestió (la (b) de més amunt) és que una altra de les coses amb què ens equivoquem més flagrantment quan pretenem predir el futur –a part les discontinuïtats, que n'hi haurà– és que no tenim ni idea de com reaccionarà el nostre col·lectiu d'usuaris, allò que en diem «societat». Farà seva una novetat que surti al mercat (encara que tècnicament no valgui un ral)? Li trobarà un ús insospitat (i/o insospitable)? La integrarà en el seu mode de vida d'una manera sòlida (o fins i tot entranyable)? No cal que us expliqui que l'exemple paradigmàtic de cosa totalment imprevista i imprevisible és el telèfon mòbil. Ni un sol escriptor de ciència ficció el va imaginar! No el va intuir ni tan sols el torrat d'en [Philip] Dick –sigui dit [això de «torrat»] amb respecte i veneració, i implorant el perdó d'en Miquel Barceló. I és que això dels usos del mòbil va anar configurant-se de manera dinàmica. Marty Cooper, el «pare del

SI HAGUÉSSIM  
DE FER CAS DELS  
PREDICTORS MÉS  
OPTIMISTES, D'A-  
QUÍ VINT ANYS LA  
IA ENS HO HAURIA  
RESOLT «TOT»...

mòbil», explica que, quan l'invent encara estava en fase de proves (a la Motorola, d'on llavors era enginyer), un dia havia d'anar a una reunió i li va sonar el mòbil –un prototip, que portava al cotxe des de casa, on hi havia estat treballant– i era el seu ajudant que li deia que no calia que hi anés, que la reunió s'havia suspès. Ell mateix, diu, va quedar bocabadat. No havia pogut ni sospitar un ús com aquell. És clar que, com l'ínclit ou de Colom, un cop descobert a tothom li va semblar natural. Sí, però: era previsible? Una mica més tard, un altre membre de l'equip de Cooper, duent un altre prototip pel carrer per veure si era prou portable, va presenciar un atropellament i, de cop, va veure que el trasto aquell podia fer un miracle: no li va caldre ni anar a la cabina més pròxima, per trucar a l'ambulància. I, és clar, tot això passava abans que el famós giny sortís al mercat. Perquè una vegada allí, no sols ha acabat tenint uns usos tan variats (i «sentits» i fins i tot «necessaris») sinó que la gent l'ha fet «seu», l'ha integrat a la seva vida i, per posar un exemple de discontinuïtat a prova de prediccions, n'ha generat un ús i una subcultura tan sorprenent i insospitable com el *texting* (els SMS) o allò que els americans en diuen *approximeeting* (que vol dir que, en comptes de quedar en un lloc per trobar-se –«meeting»–, doncs no: vas trucant i així aneu entre tots definit dinàmicament el punt final de trobada [fins al moment culminant de «ah, ja et veig, allà a la cantonada»]). . . Ah! i aquí els conceptes de puntualitat, destinació precisa o coneixement de la geografia urbana són, com és obvi, redundants. Com ha canviat tot, no? I tot per un *walkie-talkie* sofisticat que s'havia empescat una Motorola en hores baixes per veure si guanyava una mica de quota de mercat. (I, per cert, qui hauria dit que el principal beneficiari de l'invent seria una companyia finlandesa que fins llavors s'havia dedicat a la comercialització de fusta d'arbres? Una altra discontinuïtat! i ja es veu que, així, no hi ha qui predigui res.)

*Tornem-hi, doncs: «Quin és el futur de la IA?»*

Doncs, a risc de vessar-la, sembla que tothom està d'acord en un fet: la IA del futur [pròxim], «previsiblement», ens «envairà», serà pertot arreu: als *gadgets* innombrables que cada dia sortiran, tot els quals seran «intelligents» i estaran connectats a la xarxa (que

vés a saber com serà llavors) i, alguns (d'aquests *gadgets*), amb tota probabilitat els durem incorporats al vestit o al cos. A més, probablement molts d'ells estaran integrats a través de –ho heu endevinat– *el mòbil!* (això ja s'atansa: mentre escrivia aquesta frase per primer cop estava sentint en Jobs com presentava l'iPhone). El *mòbil*, és clar, ja no serà un telèfon –segurament explicarem als nostres néts, davant al seva completa incredulitat, que «allò» va començar essent un *telèfon*– sinó una mena de «comandament a distància» de totes les facetes de la nostra vida, des de la música, les claus del cotxe o de casa, la targeta sanitària (o del que sigui: clubs o DNI europeu,...), el terminal de «la xarxa» on fer tota mena de consultes, «la TV» confegida al nostre gust, el telèfon –naturalment!–, la càmera on registrar tot el que volem recordar, el «company» que ens porta el cotxe o ens aconsella sobre el camí o ens coneix els gustos,... en fi, demaneu!

...I A PARTIR DELS  
«VINT ANYS», NO-  
MÉS ENS CALDRÀ  
RESOLDRE AIXÒ: NO  
MORIR-NOS

### *I la nostra, d'intel·ligència?*

Mentrestant, nosaltres anirem pel món amb una vida transformada i còmoda (i esperant la segona fase de Kurzweil, aquella de no morir-nos [i de tots ser rics i banyar-nos en l'energia solar il·limitada]). Ara bé, i la nostra intel·ligència? Què se n'haurà fet, davant tot de *gadgets* potents, conxorxats i tots ells «intel·ligents» de què estarem voltats? Doncs és evident que no hi podrem competir, com ja no hi competim, en molts camps fins ara típicament «humans» que per a ells són trivials i per nosaltres dificultosos. Per exemple, la predicció mateixa, de que tant he parlat. La raó per la qual els savis del mètode *Delphi* s'equivoquen *tant* –vaja, *tant* com si no en tinguessin la més petita idea, d'allò de què parlen– és per un problema del que tècnicament s'anomena *focus*: no hi ha manera que ningú (que sigui humà) extrapoli a partir d'una experiència que no té, i ni tan sols imagina, i, encara menys, que pugui manipular combinatòriament el pràcticament infinit nombre d'universos possibles. Per això, és clar que el futur de la predicció –i dels seus possibles consegüents ridículs– pertany als ordinadors. De fet, tot el que fem els humans està sotmès al «problema» del *focus*. Per exemple, fins i tot en els nostres moments de pensament més abstracte no ens podem sostreure de les nostres inquietuds,



motivacions o sentiments, que són elements bàsics que acompanyen tota la nostra activitat intel·lectual. . . vaja, no l'acompanyen; és que, de fet, en són els motors: si no hi ha emoció no hi ha vida intel·ligent (parlo dels humans), com ens han demostrat infinitat d'experiments d'Amos Tversky (q.e.p.d.) i de Daniel Kahneman (aquest segon membre del duo va ser a temps de rebre el Nobel que els dos havien merescut), o d'Antônio Damásio i tants d'altres.

Per això, el nostre futur serà, presumiblement, (A) cultivar «la nostra» intel·ligència enmig d'una mar d'intel·ligències no humanes que ens ajudaran a fer-ho tot (fins i tot a pensar –quan vulguem no equivocar-nos. I (B) fiar-nos-en (de la nostra), perquè (1) si bé és (molt) fallible, és l'única que tenim «autoportant» (i sempre a mà!), (2) perquè té molts milers d'anys de rodatge (centrats en la fase caçadora-recol·lectora per adaptar-se a la qual es va anar programant, i això es nota) i (3) perquè fins i tot és possible que alguna vegada ens sigui útil, (3a) ja sigui perquè ens permeti continuar essent un col·lectiu i tenir una cultura, (3b) ja sigui perquè ens permeti –precisament pel fet d'equivocar-se (que és el seu costum)– obrir un camí insospitat vers vés a saber què, (3c) ja sigui perquè un dia –justament per la seva *infiabilitat* combinatòria– salti una espurna d'allò que hem anomenat *geni* o *creativitat* i que potser els ordinadors no tindran mai.

I si això no és ver, esperem que sigui ben trobat. I bon any. D'aquí dotze mesos ens tornarem a trobar (o a «aproximeetar»), espero, per continuar jugant a això de fer prediccions. Segurament no n'encertarem ni una, però divertir-nos, uf! si ens haurem divertit. (I si en voleu més, n'hi ha prou de clicar *Ray Kurzweil* o *singularity theory* al cercador, perquè, des d'en Pich i Pon, ja no queden gaires humoristes.)

Barcelona, 9 de gener de 2007

## *Seguint el camí, un any més*

BUTLLETÍ DE  
L'ACIA  
NÚMERO 39  
ESTIU DE 2007

**T**OTS vosaltres deveu recordar la famosa història de l'Eliza, un programa dels anys 1960 fet de retalls de frases psicoanalítiques suades que aconseguia passar per un autèntic psiquiatre, abusant de la nostra –pel que sembla irremeiable– **credulitat**. M'agradaria posar al dia l'anècdota amb un fet que va passar fa dos anys i que demostra que, pel que fa a credulitat, continuem allà on érem. Quatre estudiants de l'MIT, tips de rebre «calls for papers» de la «2005 World Multi-Conference on Systemics, Cybernetics and Informatics», van decidir crear un generador de text, el van carregar amb esquelets d'oracions, més una pila de verbs, noms, adjectius i adverbis, tot ruixat amb una considerable salpebrada d'argot informàtic. L'article, que va acabar dient-se «Router: una metodologia per a la unificació típica de punts d'accés i redundància», era gramaticalment correcte, això sí, però no tenia cap ni peus. Els organitzadors van passar l'article a revisors humans, que no hi van afegir cap observació ni en van fer cap comentari (donant-lo per bo, doncs), i va acabar *acceptat*... però sense que ningú l'acabés llegint al congrés, perquè als autors els feia vergonya (aliena, sobretot) confessar la seva innocent –o normalifeta... i perquè, a sobre, no volien pagar els 390 dòlars per cap que costava la inscripció.

### *Les bromes*

Això, d'altra banda, recorda la broma àmpliament comentada d'en Sokal, un físic que fa uns vint anys va escriure un article totalment insensat (sense sentit, vull dir) –aparentment una reflexió metametodològica (sobre la física quàntica i la «irrealitat» de la

*realitat*, ni més ni menys)– que es va anar inventant sobre la marxa fent un sofregit arbitrari d'argot físic i filosòfic i tot seguit va enviar a una revista francesa postmoderna de «ciències socials» que el va acollir amb gran satisfacció com l'«aportació valuosíssima i profunda» que es veia de seguida que era. La riota d'en Sokal va donar lloc a un llibre de gran èxit sobre la «impostura» dels *philosophes*. O sigui que la ignorància és universal i està molt ben repartida i que la broma aquesta d'abusar de la credulitat natural (o simplement d'aixecar la camisa) sempre troba nous subjectes i ocasions per anar-la repetint. No es pot dir que avancem gaire, no.

### *L'efecte-sorpresa*

I si creieu que no, que la gent avui ja no es deixa «sorprenere» pels nostres ginys, n'hi ha prou a agafar, per exemple, la vostra sogra o algun fill preadolescent (si el teniu) i fer-li veure algun d'aquest vídeos que ens passen els diversos laboratoris de robòtica en què ens il·lustren, cofois i paternals, sobre les «gràcies» dels seus eixerits robotets, que van des de jugar a futbol fins a ballar en una discoteca passant per pujar i baixar escales, dirigir simfonies de Beethoven o (provar de) confondre's amb la nostra secretària sense que ens n'adonem: la reacció de qui veu aquests vídeos va des de veure en els robots protagonistes uns pobres animalets tendres i simpàtics fins a atribuir-los com a mínim el QI d'un futbolista (que no sé, francament, quin és), el d'un buròcrata que intenta semblar simpàtic o d'un(a) recepcionista amb pretensions de sexy. Com a cas-prova per a sogres, podríeu entretenir la vostra amb, per exemple, el robotet dansaire («en Rebotet», que en dic jo) d'un vídeo recent –d'aquest primer trimestre del 2007–, a veure què hi diu; segur que s'hi entendreix. El repertori de **reaccions populars** als *gadgets* que construïm per distreure'ns –ep, els **militars** no, que no ho fan pas per distreure's!– és divertit d'observar, però encara està per estudiar i sistematitzar. I potser caldria fer-ho urgentment, que no sigui que ens acabem creient, com ha passat sovint, que amb la IA fem *ciència* quan de vegades només hi fem jocs malabars, i en diem «ciència». I, és clar, la **Ciència** és tota una altra cosa: mesuraments, repetibilitat, comparació, intent d'establir causalitat, construir un model, fer prediccions, etc. O és que amb la IA no

volem fer ciència sinó **Enginyeria**? Ah, doncs aleshores cal especificar el problema, construir-ne la solució, avaluar els plus i els *minus*, fer-ne el pressupost i convèncer el client. I tot, amb rigor, i sense propaganda ni fum (el *hype*, que diuen els americans), que hauria de quedar restringit als venedors o publicitaris. ¿Ja ho fem, tot això, els *IA-aires*? Va ser el mateix **Minsky**, em sembla, que fa uns deu o dotze anys ens renyava perquè la IA estava sent o bé un succedani de matemàtica dolenta (plena de teoremes inútils i/o inaplicables que no anaven enlloc ni venien d'enlloc) o bé un conjunt de «productes» que competien entre ells i eren mútuament incomparables i que eren considerats pels seus propositors «millors» que els rivals merament «perquè funcionaven» (però no hi havia mesuraments, ni paràmetres, ni objectius de disseny ni de cost ni res). Minsky deia que els cultivadors dels primers, els pseudo o filomatemàtics eren els **pulcres** (els *neat*) i els dels altres els **potiners** (els *scruffy*) i que les dues espècies humanes resultants eren de reproducció endogàmica, no es barrejaven mai i fins i tot s'evitaven als congressos (els primers hi anaven dilluns i marxaven dimecres, que era quan arribaven els altres). (Això, que sembla paròdia però no ho va ser tant, avui ja no passa: ja som, efectivament, espècies independents). Per això em vull concentrar en productes, i particularment en **robots**, per raons que suposo que quedaran clares.

### *No tot és «sorprendre»*

Ja sé que tot això de la *sorpresa* i la *reacció de l'espectador* no hauria de ser gens central a la IA i hi toca molt de biaix, però cal recordar que fins ara la IA, segons l'oracle del 1964 del mateix sant Marvin (Minsky), s'ha basat *precisament* en la **sorpresa** (el *sorprès* és qui veu intel·ligència allà on només hi ha artefacte), i que encara vivim d'això... i del fet que els militars americans creguin que això és útil per «sorprendre» l'enemic (i, si pot ser, perquè hi deixi d'haver enemic).

Ara bé, i aprofitant que parlem de militars americans (que com tothom sap són els qui paguen i acaben movent el motor de la cosa aquesta de la IA), no us sorprendrà pas, suposo, si us dic que un dels darrers informes del Pentàgon, després de considerar que la IA

ÉS TAN FÀCIL ENGA-  
NYAR EL VEÍ QUE NO  
ÉS GENS ESTRANY  
QUE AIXÒ PASSI CA-  
DA DIA.

civil en robots remots (guiats per GPS) ja està prou desenvolupada, proposa el següent pas: una recerca concentrada en **robots** (*entitats operatives autònomes*, en diuen), tant d'**aire** (avions-robot) com de **terra** (vaja, el que fins ara n'havíem dit *soldats*), de tal manera que al 2015 un terç (!) de totes les tropes americanes siguin un o l'altre. És un pas més, ja es veu, vers l'objectiu ja prèviament definit (cap al 2000): fer un *soldat universal* artificial, *autònom* o, com a mínim, remotament controlat. La idea és aquella tan bonica de fer la guerra però que **no** hi hagi baixes (o que, si més no, no siguin els soldats). I els *IAaires* civils i/o no-americans? Doncs, com diuen els ciclistes i com hem sabut sempre, en això de la IA *xuclem roda* i prou, i encara!

### *Robots per donar i per vendre*

Com estan les coses doncs en el camp civil? Doncs això dels robots ho tenim bé, gràcies. Ara podem dir que en tenim un bon repertori, de **robots**: (a) els que «hi **veuen**» com nosaltres (vull dir que segmenten l'escena, reconstrueixen els volums i busquen sense distreure's el que el context els fa buscar (ideals com a vigilants inavorribles que detecten anomalies –o delictes); (b) els que **raonen** sense «pensar» (vull dir, sense els inconvenients que jo mateix he apuntat moltes vegades que tenim els humans quan «raonem»), classe que no exclou la següent: (c) els que tenen «**emocions**» que motiven i situen el raonament (el citat a (b)) sense les incoherències, desequilibris i exageracions en què caiem nosaltres (vaja, que actuen com les persones que qualifiquem de *ponderades* o *sàvies*, alhora que *objectives* i, diguem, sensibles/comprehensives); (d) els que reconeixen el (nostre) estat de *felicitat* o *tristesia* (i, per tant, podran ser un dia candidats a «bons companys»); (e) els que, com a conseqüència de (a)–(d), *diagnostiquen* les malalties curtcircuitant els metges (això hauria fet felíç Voltaire) i, el que és més, encertant-la; (f) els robots exploradors que *plantegen* «hipòtesis de treball» per anar avançant (i se'n surten millor que els altres) o que col·laboren entre ells o que es *reparteixen* la feina o el territori; (g) els que segueixen la **mirada** d'algú per saber què vol (ideal per a centrals electrodomèstiques casolanes, per a quadres de control automàtic d'un fàbrica, per a ajuda a minusvàlids o –atenció!– perquè, em-

boscats darrere un cartell publicitari, ens vagin ensenyant allò que resulti més temptador per a qui ho miri), etc. I no continuo perquè no m'hi cabria. La idea general és (1) que facin *tot de coses* –per inversemblants que puguin semblar–, (2) que ens ajudin en tot, si pot ser *de bon rotllo* i evitant crueltats innecessàries d'aquestes que inevitablement gastem els humans... i també, això doneu-ho per descomptat, (3) colar-se pel nostre cervell per *col·locar-nos* tota mena de productes per satisfer necessitats innecessàries (com sempre, vaja).

### *Xutar-se un robot a la vena*

I tot això sense comptar els **nano(ro)bots** que ens introduiran per la boca, la vena o l'engonal perquè, remotament guiats i seguits, vagin a l'òrgan o cèl·lula on se'ls ha de menester i hi facin de cirurgia o farmacèutic, amb l'esperança que més endavant, un cop savis i emancipats, circulin pel seu compte pel cos humà investigant, diagnosticant i aplicant els remeis adequats (com ho faria, és de suposar, un **nanometge**). Ara ja només cal preveure que vagin ben orientats (que no sigui que ens hagin de dissoldre localment una pedra del ronyó i ens tallin un tros de pituitària) i que no es desconcertin (us imagineu una colla de metges anant de bòlit pel nostre cos a la recerca de cèl·lules canceroses que no troben i injectant antidòts pertot arreu? o el nostre cos convertit en un hospital caòtic ple de metges desorientats?)

### *Drets i deures*

Això pel que fa a les realitzacions materials. Pel que fa al marc general i legal que plantejarà a la societat **viure amb robots**, ja n'hi ha qui s'hi comencen a preparar. Primer, segur que tindrem una reedició actualitzada dels asimovians principis [morals] de la robòtica (però ara ja no amb robots merament imaginats, com ho feia l'Asimov l'any 1939 amb els seus –que se'ls havia de pintar a l'oli– sinó amb robots de debò i de tota mena). Després d'haver pensat en la **responsabilitat** (legal) dels robots (*l'Economist* hi va dedicar un especial fa no gaire) ara, no fa ni quatre dies, llegia que ja hi ha qui es planteja començar a parlar dels «**drets**» dels robots

UNA PART DE LA  
ROBÒTICA SERÀ  
NANOTEKNOLOGIA,  
PERÒ AIXÒ FA POR  
A MOLTA GENT. I SI  
SE'NS REBEL·LEN O  
DESCONTROLEN ELS  
NANOBOTS QUE ENS  
ENVOLTEN (O QUE  
ENS HEM FICAT A  
DINS)? QUÈ FAREM?

(seguint i allargant el paralelisme amb els animals no humans en general). Vaja, que haurà d'estar clar, a la vista de possibles plets o conflictes, **què** poden fer els robots i què no, quina relació **correcta** hauran d'establir ells amb nosaltres (ens hauran d'ajudar en un mal pas? quin? com ho sabran? «se les carregaran» si s'equivoquen o ens fan mal?) i nosaltres amb ells (podrem sortir-nos-en indemnes si agredim o fem mal a un robot? haurem de pagar –o pagar-li a ell– cap compensació en cas de lesions?). I si el robot és «nostre»? (hi haurà robots en propietat? de lloguer?). Tindran ciutadania, els robots? podran opinar en assemblees? i votar?... Ja veieu que se'ns ha girat feina, amb això dels robots, i potser no és qüestió de deixar-ho en mans d'amateurs o *freakies*: al primer accident amb víctimes veureu com la gent es quedarà parada, els advocats sentenciaran que «hi ha un buit legal» i els legisladors es posaran a fer lleis frenèticament, com sempre, sense tenir-ne cap idea.

*I allò que vam quedar que faríem?*

¿Recordeu el laboratori d'IBM de Zuric amb el seu BlueGene/L dedicat a simular el neocòrtex humà fet d'un milió de neurones estructurades en una capa feta de *cilindres corticals*, cada un de 3 mm de llarg i 1/3 mm de diàmetre (on es contenen tots els nostres records, coneixements, experiència i habilitats)? Sí? Bé, doncs la notícia (de finals d'abril) és que, fent servir una part del BlueGene/L s'ha aconseguit simular (mig) cervell d'un **ratolí** comú –això sí, a una velocitat de 1/10 de la real del ratolí–. Que bé, no? En fi, que la cosa sembla que avança. Ah, i quan tingui notícies del Pentàgon ja us les passaré (és que són molt callats, aquesta gent) i, com sempre, els civils ja farem adaptacions del mític «soldat mecànic» en forma d'**assistents** més o menys pacífics. Ah, i els donarem algun **dret**, això segur. I és possible que les bromes aquestes de fer-se acceptar treballs impassables en congressos les facin ells (i els qui quedem malament siguem, com sempre, nosaltres).

Barcelona, 13 de maig de 2007

## *El que tenim i el que ens espera*

BUTLLETÍ DE  
L'ACIA  
NÚMERO 40  
TARDOR DE 2007

**C**OM que aquesta tribuna més o menys trimestral que amablement se'm concedeix es presta menys, per la seva llargària, a filosofar que no pas a repassar coses que han passat mentrestant, sovint em veig obligat a comentar novetats i, com a màxim, valorar-les una miqueta, i deixar les anàlisis per al lector o per a un temps més propici. Així, doncs, això d'avui serà una «crònica», que voldria dividir en quatre apartats. Els dos primers estaran dedicats a la Intel·ligència Artificial (IA) com a «**informàtica avançada**» (també «IA») i, en particular, (1) a les **interfícies** que ens esperen, i (2) a la **informàtica pervasiva** (que nosaltres mes aviat diríem «ubiqua» o «omnipresent» i potser també, francament, «intrusiva»). El tercer apartat, el voldria dedicar al mític i primordial **soldat no humà** que tots, amb la nostra feina d'*IAires*, contribuïm insensiblement a construir (sense que, de fet, ningú ens ho hagi consultat mai). El quart estarà dedicat a possibles **problemes de disseny** de les màquines que fem o farem. I ho voldria culminar tot amb un homenatge, seguit d'una petita informació de servei.

### *Les interfícies que vénen*

Encara que a nosaltres, professionals, ens sembli que els ordinadors d'avui, amb finestres i ratolins, son senzillíssims de fer anar –sobretot si els comparem amb els que alguns vam conèixer, amb *suitxos*, llenguatge-màquina, llumetes, fitxes, cintes de paper i altres primitivismes inconfessables–, de fet, no són gaire o gens naturals ni intuïtius. Sense fer futurisme imaginari «a la Kurzweil» i cenyint-me a projectes que ja estan engegats i finançats i que



tenen cara i ulls, en voldria destacar aquests dos que explico tot seguit.

a) Com se sap, les interfícies home-automòbil son probablement les que més forcen el progrés, ja que el temps de reacció, la immediatesa i l'eliminació de qualsevol element redundant a la comunicació pròpiament dita (evitant les distraccions, potencialment fatals) són bàsics. I que hi trobem, aquí? Doncs comunicació verbal (sí, ja sè que direu que en teniu experiències nefastes), que, però, en els darrers mesos ha evolucionat fulminantment, gràcies a l'augment vertiginós de la potència de càlcul, fins a tenir «converses» globalment acceptables (sobretot per a l'interlocutor humà, el més sensible a qualsevol aparença d'artificiositat o poca naturalitat). Avui ja es fan sistemes plenament «acceptats» per l'usuari que s'hi comuniquen fluidament (ai, en català, també?) sense que aquest noti res de desagradable o incòmode ni s'hi posi en guàrdia (en contra). I si avui **parlem**, potser demà ja no caldrà ni això. Quan encara estem impressionats pel cas de fa un o dos anys del paraplègic connectat a un ordinador que era capaç de moure-li el cos a partir dels seus pensaments (concretament, dels actes neuronals de volició motora), aquest trimestre Microsoft ha patentat un sistema de «**lectura directa del pensament**» per mitja d'electroencefalogrames que filtren tot d'informació irrellevant per concentrar-se només en les ordres donades per l'usuari i en les respostes que aquest dóna a preguntes que se li han fet. Tot, evidentment, sense que mai Microsoft ens hagi preguntat si nosaltres hi accediríem (ni crec que ningú ens ho preguntí mai). És clar, en el cas d'un cotxe i un conductor la seguretat no és cosa de broma, i ¿qui s'hi pot negar, per la nostra seguretat, que ens preguntin cosetes a base de regirar-nos les neurones?

[Un petit *excursum* sobre els cotxes intel·ligents i els sistemes d'ajuda al conductor de què he parlat manta vegada. I és que he llegit una notícia que em sembla divertida i significativa: quan a una persona se li demana què pensa d'un sistema que es preocupi del conductor, que detecti el seu estat d'ànim i capacitat per conduir, que li corregeixi les errades i que fins i tot el porti a l'hospital si cal, tothom diu que és una bona idea; i quan es prova un sistema d'aquests amb un conductor real, aquest se sent bé, acompanyat i protegit. Ara bé, quan després de l'experiència se li revela què ha passat, quines decisions ha pres el sistema per compensar el que ell

(es pensava que) havia fet voluntàriament i quina és la magnitud dels seus errors, com s'havia picat infantilment i innecessàriament i quan s'havia comportat de manera inexplicable o com de prop havia estat de fer carallotades irreparables, resulta que el conductor que havia quedat tan encantat de l'experiència se sent estafat i indignat –en el sentit tant d'irritat com en el d'haver-se-li «robat» la seva lliure voluntat, la intimitat i la dignitat, tot alhora. Curiós, no? Els comercialitzadors de l'invent ja han vist que els tocarà fer campanyes interminables de publicitat omnipresent i persuasiva perquè els futurs conductors-clients es «deixin portar» i ho acceptin sense que s'indignin, com una cosa natural i que es fa pel seu bé.] (Fi de l'«excurs».)

b) I si no n'heu tingut pas prou amb la comunicació directa **ordinador-cervell** que ens prepara Microsoft, doncs aquí en teniu una altra: Aviat (més aviat que no us penseu, us ho prometo) les *pantalles* passaran al museu, fent-hi companyia als *ratolins* i, no cal dir, als superobsolets *teclats*. Per dir a l'ordinador què volem fer i a partir de quina part de l'*input* visual, tindrem (després del pas intermedi de les *taules de dibuix* que ara ja es comercialitzen) una interfície «gestual» (per entendre'ns, aquella que sortia al *Minority Report* de Spielberg i que tan bé caricaturitzaven els de *Polònia* a costa de la Tura). I les «*pantalles*» ja no seran planes ni materials, sinó que es «projectara» en un pla (o **en 3D**, en un volum més o menys planar) davant nostre (com la de la Tura, recordeu?) i «agafant» i/o «manipulant» amb **gestos** el que ens interressi contribuirem a enterrar el ratolí al magatzem dels estris que no sabem com vam poder arribar a considerar «naturals» un dia. És l'arribada, apunteu-vos-ho (per quedar bé amb la sogra o en reunions, si més no), de la *gesture-based computing* (**informàtica gestual**, que diríem nosaltres). Si en voleu veure primícies haureu d'anar, per exemple, a Treviso (aquí mateix, vora Venècia) on a la companyia iO us ensenyaran (vaja, si volen) la *sensitive wall* que perpetren.

### *Estem rodejats!*

Si tothom concedeix que l'Internet ens ha posat a tots en contacte *independentment* del lloc i les distàncies, ara, gràcies a Google Earth, Google Maps i el GPS ens construiran una «**Geoweb**» on estarem

tots ultralocalitzats i en què tots sortirem «retratats» –mai més ben dit– des de dalt (de fet, a alguns ja se'ls ha sorprès fent potser allò que no tocava, amb gran escàndol dels advocats, dels defensors de la privacitat i, en aquest cas també dels cornuts afectats). El **lloc** on som a cada moment serà vital, per exemple, per avisar-nos de res greu que ens pugui afectar, de rescatar-nos (si fos el cas) o de fer estadístiques d'on és tothom a cada **moment** (per, posem, regular el trànsit) o, cosa que pot esgarrifar segons qui, per reconstruir, a partir d'un crim o delictes, on era cadascun i, doncs, qui va ser qui. No cal dir que això, més encara que les omnipresents càmeres dels nostres carrers, fa esborronar tot els cabells de qui en tingui. És clar que també, s'hi pot adduir, que sàpiguen on som ens pot salvar la vida en un accident o en una aturada cardíaca que hàgim sofert en ple carrer o en un robatori a domicili que algú haurà detectat de lluny sense que hagi calgut que l'aviséssim... però, on queda la privacitat? I això sense pensar –lluny de nosaltres!– que les companyies publicitàries ens puguin comptabilitzar els hàbits i costums i ens retallin el patró o perfil comprador segons el que han vist que fèiem.

ENS COMUNICAREM  
AMB L'ORDINADOR  
AMB LES MANS  
I SENSE PARAULES.  
PER TOCAR  
TECLATS, RATO-  
LINS I PANTALLES  
HAUREM D'ANAR  
A VEURE'LS ALS  
MUSEUS DE LA  
CIÈNCIA, ON HO  
EXPLICAREM A UNS  
NÉTS INCRÈDULS.  
(AIXÒ SERÀ AVIAT,  
US HO JURO!)

A tot això cal afegir que anirem connectats pertot arreu i se'ns podrà seguir també –a més de visualment– *informàticament*. Tindrem un ordinador –dels «gestuals» i «mutts» (vull dir, dels sense paraules, com he explicat, via comunicació directa cervell-sistema), és clar–, tot com si fóssim un Blue-tooth amb potes (en «Dents blaves», com se sap, era un rei danès que va viure abans d'en Canut i que no sembla pas que mirés gaire prim amb la neteja ni amb res), sempre localitzable i sempre disposat. I duem sensors que ens diran –i/o li diran al nostre metge– com anem de temperatura i que informaran a qui-sap-qui que fem els 140 per hora per un camí o que, si veuen que tenim fred, faran que el nostre abric ens escalfi allà on toqui. Podeu dir que tot això es fantasia. Però no. He triat acuradament els exemples a partir, exclusivament, de projectes que estan engegats en algun lloc, que semblen seriosos i prometedors a curt termini, amb perspectiva realista de ser al mercat aviat. (El que ja em supera, francament, és imaginar la pinta que farem tots plegats anant pel carrer, gesticulant com bojós.)

### *Mentrestant, el 5è de Cavalleria...*

Ja he dit aquí més d'un cop que el departament de Defensa americà vol substituir un terç dels seus efectius per robots el 2015. La raó és simple: els cadàvers de robot ni son enviats als familiars en taüts amb la bandera americana ni s'enterren ni son plorats per ningú. De moment el Pentagon ja fa servir rutinàriament avions-robot (un d'ells va matar fa poc Mohammed Atef, cap d'operacions d'al-Qaeda) i també robots metralleta (anomenats *Sword*), però tots ells són operats remotament, per humans. Ara, abans de fer el **soldat immortal** automàtic que es busca, un laboratori de Maryland n'ha fet una petita bestreta en forma de robot que, més que combatre al camp de batalla, el que fa és allò tan bonic dels codis d'ètica militar que diu «no abandonaré mai un camarada caigut», precepte que a la pràctica pot costar molt car a la tropa (el qui es queda a cuidar el caigut té tots els números per ser el pròxim a caure), i això infla qui-sap-com el nombre de baixes. La solució, un robot –no gaire antropomòrfic, per cert– que és una mena de grua (anomenada *Bear* o simplement *tors*) capaç de protegir, aixecar i transportar el ferit a zona segura. Be, doncs; sembla que anem ben encaminats. Al pròxim intent, o al següent, ja tindrem el «soldat desitjat». Ara bé, precisament per això es comencen a detectar moviments per determinar i fixar regles per a aquests «robots de camp», perquè facin servir la seva letalitat amb seny i «èticament». En altres paraules, es tracta de crear i d'inculcar-los una «**consciència artificial**» i que actuïn humanitàriament, més fins i tot, si és possible, que els mateixos humans (a les inconsistències i febleses racionatòries dels quals els robots cal suposar immunes). Per fer el fet, a Georgia s'està elaborant el que en diuen *espai de decisions multidimensional d'actuacions possibles* que, a partir de l'input (arribat via sensors, radar, informes d'espionatge o el que sigui), divideix aquestes *accions possibles* en dues menes: les *ètiques* i les que no ho són. Exemple pràctic: si es persegueix el cotxe de Bin Laden i se'l vol atacar al moment que avança un autobús escolar, el sistema «es conté» i no dispara. Com diu un dels responsables del projecte, comparant regles amb les clàssiques de la robòtica (degudes a Asimov), aquestes eren concebudes per evitar que els robots causessin cap mal als humans i les d'ara, en canvi, són

«perquè com a mínim no se'ls mati de *manera no ètica*» (sic). Això no vol pas dir que a partir d'ara ja tinguem guerres «netes». Un robot voltat del caos i confusió normals en tota situació bèl·lica pot continuar equivocant-se, i molt, ja sigui perquè les dades que li han subministrat i de què parteix són errònies, ja sigui perquè les condicions del camp de batalla canvien ràpidament. Un altre perill, aquest de més gran angular, és que l'ús de robots potser farà més fàcil declarar una guerra (fàcil, si el resultat és un pati de ferralla i un compte de despeses inflat, i no pas un cementiri i una colla de famílies trencades), la qual cosa pot fer contents els generals, que ara tenen una joguina que no esperaven i ja no cal que s'enfrontin a resistències de la gent. Ara bé, com comenta *The Economist*, si el robot té «consciència» aleshores també caldrà esperar que pugui refusar una ordre que no li sembli ben fonamentada, racional o prou ètica. I perquè es vegi fins a quin punt tenim les coses poc clares sobre fins on arribem nosaltres i fins on arriben els robots i sobre els nostres propis sentiments sobre l'assumpte potser val la pena citar el comentari d'un militar que participava fa poc en un experiment a Arizona quan va veure que un robot esclatava, tot destrossat i fet a miques, després d'haver trepitjat una mina (l'experiment era precisament aquest: desminar un camp a base de robots-conillets d'Índies); deia el militar, després de manar aturar l'exercici, tot trasbalsat: «pareu, això és insuportable, és *inhumà!*»

NO ÉS CAP BAJANA-  
DA DIR QUE AVIAT  
FORMAREM PART  
D'UN TOT DEL QUAL  
DIFÍCILMENT ENS  
PODREM ESCAQUE-  
JAR

### *Problemes psiquiàtrics?*

Una cosa que segur que haurem de tenir en compte construint robots, tinguin consciència o no, és que, si previsiblement els mecanismes que hauran de tenir per ser intel·ligents són més o menys pròxims als que tenim els humans, la síntesi i conjuminació de tots aquell mecanismes poden donar alguna sorpresa desagradable. Per exemple, se sap que un component no negligible de molts trastorns psiquiàtrics, des de l'epilèpsia a l'autisme passant per moltes neurosis, és el desajustament, sovint lleu, de diferents mecanismes a l'hora de funcionar junts. ¿Voldrà dir això que, si no mirem prou prim a l'hora d'ajustar components, podem tenir robots neuròtics? o ¿podrem arribar a suportar un robot autista? Una altra cosa: si les pulsions, positives o negatives, com ara l'atracció,

repugnància, por, etc. són elements fonamentals i motivadors de tota conducta intel·ligent, almenys de les conegudes fins ara, ¿no podrà pas passar que ens trobem amb robots que s'encaparrin amb alguna cosa, o que siguin difícils de tracte o, pitjor, que ens agafin mania? Caldrà seguir molt de prop com anem construint robots, quines capacitats van adquirint, quines conductes en resulten i com es conjumina tot en un conjunt que no grinyoli ni tingui efectes inesperats. Especialment apassionants són els resultats que s'estan obtenint aquests mesos sobre el component altruista de la nostra conducta i sobre com la por de ser censurat o castigat condiciona els nostres actes. ¿Sabrem inocular aquests coneixements als nostres robots amb prou saviesa i seny perquè el seu comportament resulti alhora «intelligent», no-egoista, després i «bona persona» (però eficient, no pas «calçasses» ni «pringat») i, a més, «humà» (en el sentit de bo, comprensiu, compassiu i humanitari)? És a dir: ¿podem aspirar a tenir **robots** que siguin «persones **decents**», que tinguin allò que qualifiquem com a «**conducta sana**»?

### *Apèndix A: Un homenatge*

El que vull retre aquí a Donald Michie (pron. «maiqui»), que va morir fa poc d'accident de carretera. Era un escocès del morro fort que va començar amb la informàtica quan encara no existia, desxifrant un codi de tèlex que feien servir els militars nazis i, més tard, col·laborant a Bletchley Park amb la gent del Colossus i amb Turing. Gràcies a ell i a la seva càtedra a Edimburg (que ell va batejar com de *Machine Intelligence*) i la seva desena de llibres enciclopèdics publicats els 1960 i 70 (titulats també *Machine Intelligence*) els europeus sortim a la història de la IA, i hi sortim com a fundadors.

### *Apèndix B: Una informació de servei, qui sap si útil*

M'ha cridat l'atenció un article publicat a **Science** que no voldria desapropiar sense afegir-hi un comentari. És sobre un tema aparentment deslligat de la IA, potser massa inespecífic, però que podrà interessar una certa massa de lectors, atès que molts d'aquests són joves i estan en situació d'obrir carrera i fer-se un nom

ESCÒCIA I CATALUNYA SÓN MONS A PART: EN UN ENTERREN PIONERS DE LA INFORMÀTICA AMB 65 ANYS DE SERVEI MENTRE EN L'ALTRE NO SOM CAPAÇOS D'ASSEGURAR EL FUTUR DELS NOSTRES JOVES

a base de *papers*. Es tracta de l'anàlisi feta dels articles rebuts i acceptats (no necessàriament d'IA, sinó científics en general) per l'European Research Council en una convocatòria feta el 2007 en què se'n van rebre uns 9000 i se'n van acceptar uns 500. El detall curiós que volia ressaltar és aquest: el màxim nombre d'articles rebuts (400) venia d'Itàlia, seguit dels 200 i escaig de britànics, 200 d'alemanys, 200 de francesos i més de 100 d'espanyols. El resultat del procés d'acceptació capgirava els resultats, suposadament massa inflats, d'Itàlia i Espanya, que passaven del 1r i 5è lloc al 5è i 6è (amb només una i mitja dotzena, respectivament, d'articles acceptats), segurament més d'acord amb la potència relativa esperable. **Science** comenta aquesta inflació de papers presentats atribuint-la a les males perspectives professionals dels investigadors joves i a la seva falta de suport institucional en aquests dos països mediterranis, que porta als joves a fer-se conèixer com poden, tot i que la qualitat i el suport que han rebut en el seu treball siguin deficitaris. Com que d'això depèn el nostre futur científic (i el passat ja és prou desastrós) i com que sempre m'ha semblat que el tractament a què sotmetem els nostres científics joves és –a més de descoratjador– vexatori i indigne, aprofito per adherir-me al toc d'atenció de la revista americana, per si això serveix de res.

Barcelona, 18 d'octubre del 2007

## *La dansa dels robotets (que Txaikovski no va poder incloure al Trencauous)*

**A**QUESTS dies constato que el «front teòric» de la Intel·ligència Artificial (IA) no avança gaire. No veig pas que hi hagi grans «visions» intel·lectuals ni s'hi construeixin grans teories. Potser és que tots plegats anem despistats o curts d'idees o bé –i aquesta benèvola hipòtesi és la que aquí caritativament subscric– es tracta del fet que, com a la Biologia del segle XVIII, a la IA d'ara li calen, abans d'anar per lleis i postulats, «fets, fets i fets», com deien els naturalistes a l'època. Per això avui, més que no pas fer un destil·lat de teories, m'estimo més repassar el panorama de robots i robotets i espigolar-hi, més enllà de l'inevitable component peculiarment japonès i/o *freaky*, (a) cap on apunta tot això, (b) què tenim i (c) què sembla possible. Lamento, doncs, tornar a caure aquí en la llista de petites genialitats gracioses, més o menys amanides i il·lustrades amb referències als vídeos promocionals corresponents.

### *Una mà de vídeos<sup>9</sup>*

Per començar, què millor i més genial, per inspirar-s'hi, que la darrera Exposició anual de robots de Tòquio. Per si voleu delectar-vos amb el vídeo oficial de presentació<sup>10</sup>, heus-el aquí: [http://www.youtube.com/watch?v=9GYS3wE\\_zv4](http://www.youtube.com/watch?v=9GYS3wE_zv4) (a partir d'aquí, per brevetat, només citaré les lletres que van després del «v=» dels vídeos de YouTube). En particular, segur que us abellirà la robot-pacient<sup>11</sup> [Vaf-QxhQh6g](#), sobretot si sou estudiants i voleu ser dentistes i anar practicant sense fer mal a pobres incautes.

NODES  
NÚMERO 41  
PRIMAVERA DE 2008

<sup>9</sup> En el moment d'imprimir aquest llibre totes les referències a la web són operatives, però per més seguretat es posen referències per trobar-les en un futur. (n. de l'ed.)

<sup>10</sup> Youtube *Highlights from the International Robot Exhibition in Tokyo 2007* (n. de l'ed.)

<sup>11</sup> Youtube *Simroid 2007* (n. de l'ed.)



<sup>12</sup> Youtube *Toyota shows off violin-playing humanoid robot* 2007 (n. de l'ed.)

<sup>13</sup> Web de Sylvain Calinon (n. de l'ed.)

SISTEMES INTEL·LIGENTÍSSIMS  
COMPLETS, POTSER  
NO ELS TINDREM,  
PER ARA. PERÒ DE  
MOMENT I MENT-  
TRESTANT, MODES-  
TAMENT, PODEM  
ANAR RECUPERANT  
MOVIMENTS PER-  
DUTS I CREANT  
SISTEMES AMB UN  
MÍNIM D'EMPA-  
TIA I D'IMITACIÓ  
INTEL·LIGENT

<sup>14</sup> Youtube *Monkey's thoughts make robot walk from across the globe* 2008 (n. de l'ed.)

Però no us penséssiu que tot són –diguem-ne– bestieses, amb perdó. Mireu-vos amb atenció aquests curiosos productes robòtics assortits de can Toyota<sup>12</sup>: [qyPAIpXm-nU](#) i de seguida comprendreu que els japonesos s'estan fent vells tan de pressa que no estan pas per orgues. Observeu que, a part de l'aspecte *freaky* dels invents, hi domina el vessant pràctic (el d'acompanyar i ajudar els vells a passar l'estona d'un lloc a l'altre, per exemple).

Però és que hi ha més coses. La facilitat amb què hem (vaja, «han» –els japonesos, vull dir–) aconseguit fer moure els robots de manera no programada sinó per imitació fidel, segur que us sorprendrà. I si no ho creieu, mireu-vos algun dels vídeos d'aquesta llista escurçada<sup>13</sup> <http://www.calinon.ch> (dispenseu si recomano vídeos a tort i a dret –potser trobareu que en faig un gra massa– però és que en el darrer número ho vaig fer i no sabeu la de requestes que vaig rebre). La gràcia, aquí (amb els ninotets imitadors, vull dir), és que aquesta mena de robots –i d'altres de semblants que hi ha– aprenen a fer les coses molt ràpid, molt naturalment i d'una manera molt similar a com ho fan els infants «reals». I d'altra banda, quan se'ls posa entre nens, hi congenien força bé i aviat tots plegats es consideren amics i companys de jocs. Recorda una mica, la cosa, com altres robots (em sembla que ja ho vaig esmentar aquí) feien d'escarabats, i ho feien tan rebé que els escarabats «autèntics» que hi vivien acabaven tenint-los com a inspiradors i guies (líders, o gairebé).

### *Recuperar el cos*

El que fa més il·lusió, però, –a mi, si més no– és veure com les noves tècniques de la IA ja estan ajudant a crear un futur on els discapacitats confinats en una cadira puguin ressuscitar la meitat del seu cos que un dia van haver d'oblidar per força. Mireu-vos, si no, aquest vídeo<sup>14</sup> [L8oAz4WS400](#) en què un robot (a Tòquio) camina només pel fet que rep els senyals cerebrals (pertinentment filtrats) d'un mico (de Carolina del Nord) que camina exactament igual que el robot (vaja, al revés, suposo que ja m'heu entès). Això, que jo personalment trobo extraordinari, lliga amb la notícia de fa un any, que ja vaig recollir aquí, del paraplègic americà que podia controlar els seus músculs amb només els seu pensament, i

també amb la notícia de fa molt poc (més aviat mèdica que no pas d'IA) de ponts quirúrgics possibles entre neurones espinals de més amunt de la lesió i les motores d'espina avall. Aquesta segona és especialment miraculosa per als qui han conservat les extremitats i han tingut la paciència de confiar en la seva resurrecció algun dia.

La primera és diferent: és l'esperança de poder recuperar la *mobilitat humana plena* quan –com els ha passat per exemple a les desenes de milers de soldats americans que l'han perdut a l'Iraq (més de vint mil fins ara, concretament)– ja no la tenen tota (i, a alguns, amb prou feines els queda una extremitat per moure). En aquest cas és enormement esperançador el braç-pròtesi Luke Arm (el nom s'explica per la cosa aquella de l'*Skywalker*-fill a qui una espasa làser talla el braç) que ha inventat Dean Kamen (sí, el del Segway) i que amb 18 (divuit!) graus de llibertat permet al seu posseïdor des de pelar raïm fins a descrostar, un per un, els anissos incrustats en xocolata, passant per trepanar correctament amb una barrina elèctrica, obrir o tancar una porta amb clau i fer una encaixada de mans (i graduar-la segons l'amistat o la mala llet que es desitja). I tot seguint només senyals del cervell. I, a més, amb *feedback* sensor (vull dir, rebent la informació tàctil corresponent a l'acció feta). Si en voleu un tast, heus-el aquí<sup>15</sup>: [1hzRja9eunY](https://www.youtube.com/watch?v=1hzRja9eunY) (sobretot cap al final, després del rotllo preliminar).

<sup>15</sup> Youtube *Dean Kamen and the Robotic Arm* 2007 (n. de l'ed.)

### *I després del cos, els miralls*

Fa temps que parlo de les neurones-mirall (neurones especulars?), que són aquelles que funcionen tant (i de la mateixa manera) quan fem una certa cosa com quan veiem que aquesta mateixa cosa la fa un altre. L'existència d'aquestes neurones es va descobrir fa uns deu anys en micos, es van hipotetitzar –i es van trobar– en humans i ara s'han trobat en ocells cantaires (ho podeu constatar<sup>16</sup> personalment a [VDFhrKmDJoQ](https://www.youtube.com/watch?v=VDFhrKmDJoQ)). La gràcia –enorme!– que tenen aquestes neurones és que permeten i donen lloc (1) a «**entendre l'altre**» i les seves intencions, (2) a l'**empatia** (tu sents el mateix que està sentint l'altre fent allò que l'altre fa i que tu segueixes, i assumeixes, en els teus propis termes) i (3) a la **possibilitat de l'«aprenentatge vicari»** (allò tan lleig –potser– d'evitar les teves pròpies patacades havent tingut abans la precaució de veure les

<sup>16</sup> Youtube *Learning language with the singalong neurons* 2008 (n. de l'ed.)

que es clava el company).

Deveu haver observat que aquests tres fenòmens –la comprensió de l'altre, l'empatia i l'aprenentatge per observació (de la conducta aliena)– són precisament definidors de l'especificitat humana. Potser, doncs, incorporant circuits «especulars» a les nostres màquines podrem esperar reproduir tres de les grans qualitats que ens fan peculiarment humans (respectivament, la comprensió-predicció, la simpatia-solidaritat i l'aprenentatge (social) per observació-imitació) i que són la base d'allò que en diem la cultura (de l'espècie). No és pas poc, tot això, per a aquest primer trimestre del 2008! I això passa mentre la DARPA encara no sap si està en condicions de fer la seva propugnada *unitat autònoma automàtica de combat* (llegiu-hi «soldat-robot») i tenir-ho llest per al termini previst (un quinquenni, si fa no fa). Mentrestant, sort en tenim dels japonesos (i dels seus robotets violinistes i dansaires)!

I pel que fa al futur: de moment, mentre no tinguem els promesos soldats artificials, podrem –gràcies als avenços de la IA– anar apedaçant els soldats (naturals) que ens arriben desconjuntats de la guerra, amb una mica d'esparadrap sintètic en forma de pròtesis «intelligents» i potser fins i tot empàtiques. I qui dia passa any empeny.

Les teories, deixem-les per un altre dia. Un en què tinguem una mica més clar cap on anem (si és que anem enlloc).

## De sentiments (i de llur absència)

«Un jove va matar un nadó a Reus perquè se li interposava en una partida de videojoc».

NODES  
NÚMERO 42  
TARDOR DE 2008

**P**ARTINT d'aquesta notícia, l'escriptora lleidatana Imma Monsó deia fa relativament poc a la Vanguardia (19/7/08), a propòsit de la nostra relació amb les màquines, que «la Roboètica és a les beceroles», i que «les preguntes són apassionants: Tenen drets, les màquines [intelligents]? com es veuran afectades les relacions interpersonals, en la seva presència? quins límits caldrà imposar a la indústria dels robots (i dels robotsmascota)?» És d'agrair que un dels temes recurrents d'aquesta columna hagi saltat a les pàgines d'un diari.

El mèrit d'aquest salt és exclusiu de la nostra companya Carme Torras, que ha escrit una novel·la de ciència-ficció, «La mutació sentimental» (Carme Torras. Pagès Editors. 2008, ISBN 978-84-9779-6), a partir de la qual Monsó construeix el seu comentari.

Com que no hi ha res com llegir la novel·la, us hi invito, tot limitant-me jo ara a resumir al vol quatre comentaris que s'han emès a la web a propòsit de l'obra, perquè tingueu alguna noció de per on va la cosa:

- 1) «Amb la robòtica, entrem en els sentiments. No sentim afecte per la rentadora, però en sentirem pels robots companys de joc –dels nens– o assistents –dels avi. La indústria s'hi posa bé, i per això, de moment, els dona forma humanoide i els fa tenir –o simular– emocions.»
- 2) «Una qüestió clau: com conjuminar els fets que el progrés científic, i concretament els robots, ens faci la vida més fàcil i

ens faci més lliures (com a individus) i que, alhora, ens pugui fer més dependents i vulnerable (com a espècie?)»

- 3) «La tecnologia condiona l'evolució de la ment humana: fins a quin punt podrem incidir-hi dissenyant robots que ens estimulin i ens facin més creatius?»
- 4) «Al s. XXII potser s'hauran perdut sensacions, emocions i sentiments (allò que –avui– ens fa humans, allò que ens configura com a tals). I potser els enginyers hauran de dissenyar aparells per potenciar la creativitat (la capacitat de pensar diferent) i l'admiració (o capacitat d'estimar), per enriquir la coexistència entre humans, i entre màquines i humans.»

Què? us en heu fet una idea? Ja heu degut intuir, a partir d'aquests comentaris que la novel·la ha suscitat, que l'obra toca, si bé de manera literària i creativa, les mateixes vores que toquem aquí (sovint).

### *Efectivament, amb la robòtica entrem en els sentiments*

Com fa moltes columnes que repetim, i com ja fa una mà d'anys va dir Damásio (i molts d'altres han repetit), la intel·ligència no és res sinó el resultat de –agafeu-vos– «emocions» que l'engeguen i la motiven (...perquè si no, segons Searle, la ment només pot arribar a ser, a tot estirar, un «despatx de traducció» –d'ideogrames xinesos, diu). El que Torras es proposa –i aquesta columna, de vegades, també– és imaginar quins sentiments tindrem nosaltres quan les nostres màquines intel·ligents també en tinguin. Com és previsible que evolucioni el sistema home-màquina llavors?

En tots els problemes de predicció hi ha un nombre infinit de solucions (...que, segons la nostra desgraciada experiència profètica, seran quasi totes segurament equivocades, per cert). Una d'elles, tan raonable com meditada i sòlida, és la de la Carme, experta amb acreditat sentit comú, professionalitat i pràctica reflexiva. De fet, la seva descripció dels corresponents futuribles és tan seriosa i enraonada que sembla que efectivament la cosa pot acabar sent, si més no al segle XXII en què ella se situa, tal com Torras ho infereix i descriu.

Ara bé, encara que només sigui pel meu hàbit inveterat (i a aquestes alçades probablement inextirpable) de portar la contrària o d'afegir punyeteries atrabiliàries a una brillant tesi, deixeu-me esplaïar-me una mica amb algunes consideracions col·laterals.

*El nostre hardware va ser cablejat al Mesolític i això sembla bastant inexcusable i determinant.*

En aquesta època (circa 40.000 anys aC) l'Homo sapiens sapiens, nou de trinca i vingut de l'Àfrica pel mar Roig i el Caucas, en grupets ridículament escarransits i amb tot l'aparell simbòlic-abstractiu que nosaltres podem ara cultivar, era bàsicament caçador i recol·lector, i això es nota –molt– encara ara. La nostra ment amb prou feines s'ha modificat des de llavors (perquè l'evolució no opera sobre intervals de temps tan curts). Tot el que hem fet ho hem fet amb el mateix *hardware*. I quan dic que l'home era caçador-recol·lector i que tenia el cap fet a la nostra mida hauria d'haver matisat que l'home era caçador i la dona recol·lectora, i que quan parlem del «nostre» cervell de fet parlem d'un continuu que va des del de la dona al de l'autista passant pel de l'home (mascle), que s'hi troba aproximadament a mig camí (no és broma) (i encara hi ha un altre espècimen a mig camí entre el mascle i l'autista, com diré al final). I amb aquest material simbòlic pensant hem avançat, a empentes i rodolons fins aquí. No és estrany, doncs, que la nostra ment contingui elements nascuts i adaptats al Mesolític que avui ja resulten clarament *disfuncionals*. N'esmentaré deu (10) més avall, però abans deixeu-me recordar dos detalls en què jo he insistit de vegades, per posar la cosa en context:

- a) Podem acceptar que la Intel·ligència, grosso modo, és com la descriu Jeff Hawkins fa poc (2005): essencialment, un reconixedor de formes que desencadena, si s'escau, un potent mecanisme *resoledor* de problemes. Vull dir que el que ens fa intel·ligents és, bàsicament, la detecció molt sensible de (petites) diferències (exemple: entrem a casa i «notem» de seguida que hi ha una cosa nova o canviada) que, un cop advertides, enguen automàticament tota la maquinària deductiva («què deu haver passat?») fins a trobar una explicació tranquil·litzadora (i

poder tornar al nostre estat de repòs que, com tothom sap, és el natural).

- b) Com explicava Roger Shank (a la Generalitat, al costat d'en Pujol) els humans tendim, per a saber el que pensem, a estructurar els nostres «pensaments» en forma d'històries: ens les imaginem i ens les expliquem (vull dir que ens passem la vida, a la feina i arreu, essencialment xafardejant de manera compulsiva). Tant és així, que es pot explicar l'origen de la Lògica a la Grècia del segle VI aC com un mer «posar ordre» al caos d'històries mútuament incompatibles i/o internament contradictòries que s'explicaven, sobretot sobre els déus).

### *Deu trets humans disfuncionals*

La que segueix és la llista promesa de deu trets «irracionals» que hem heretat del Mesolític i que no hem pogut superar per falta d'evolució (física) cerebral. No és que un món racional no pugui admetre irracionalitats (més avall en veurem algunes que són força tolerables) sinó que aquestes deu són *disfuncionals*. Admetre en robots aquests peatges evolutius ens portaria a un mecanisme raonador que podria prendre decisions contràries a la racionalitat, entesa com l'adaptació a la realitat objectiva mitjançant la raó, per la qual cosa i perquè ells no vénen d'Àfrica com nosaltres sembla que hauríem de fer que els nostres robots hi fossin immunes: (el copyright de la llista potser l'hauríem de donar tot sencer al nobel Kahneman i el malaguanyat Tversky)

1. La frustració com a motor: Heu observat mai, en memòries i autobiografies, que sol haver una mena de trauma inicial que explica i dóna sentit al relat, que n'és la motivació i que sovint és vitalícia (vull dir, que propulsa tots els actes del biografiat, fins que es mor)? Aquest trauma iniciàtic és un cas evident de frustració que actua com a *primus movens* (com també ho és el fet de «picar-se» amb algú i d'engegar, com a conseqüència, una carrera imprevisible i imparable d'accions). No cal dir que d'una cursa com aquesta poden sorgir coses com la imprevisibilitat (i, per tant, amb una mica de sort, l'originalitat... i, si som laxes en la sinonímia, la «creativitat»). La frustració no

és *disfuncional* en ella mateixa; en realitat la frustració és un component bàsic de l'aprenentatge (incloent-hi el de màquines, el *machine learning*). No és pas la frustració en si sinó la nostra fixació en la frustració, la seva conversió en trauma (que potser arrossegarem tota la vida), el que la fa *disfuncional*. No cal dir que frustració pròpiament dita ve del contrast entre expectatives i realitat (contrast que també causa altres *disfuncionalitats*, com veurem en els trets 5, 9 i 10, més avall). No cal dir que res d'això no hauria de passar amb els nostres robots, que no haurien pas de quedar marcats, com nosaltres, per defectes de fabricació o per episodis inicials durant el període de rodatge.

2. Delimitació de camp: Quan pensem, ens concentrem en una part petita de la realitat (és el que en anglès diuen posar-hi el *focus*). Això és molt bo, perquè descartar elements no-rellevants ens «concentra» de manera potent. Però a un robot, sense les nostres limitacions «d'accés», li és perfectament concebible fer la mateixa feina sense deixar de tenir present el camp complet. (Nosaltres, en canvi, ens deixem portar hipnòticament per allò que tenim al davant: penseu en la força que té una imatge, tan coneguda i explotada pels publicitaris.)
3. Avaluació probabilística molt acurada però amb fal·làcies incorporades: Això ha estat ben estudiat: potser perquè en la caça no es poden fer bromes, les probabilitats que el nostre cervell calcula són prou funcionals; ara bé, com també ha estat senyalat sovint, l'atzar, vist des del cervell, té unes «fal·làcies» adherides (que no detallaré: n'hi ha prou a recordar com de manera totalment insensata comprem loteries o pugem als cotxes, tot i que sapiguem que aquestes dues activitats són desproporcionadament il·lusòria, la primera, o perillosa, la segona: la felicitat –imaginada– d'un premi o les conseqüències –volgudament no imaginades– d'un més que possible accident poden més que qualsevol racionalitat).
4. Obsessió per la compensació equitativa: Molts dels nostres raonaments, sovint invàlids, no s'entenen sinó per una dèria que tenim per la compensació: si abans m'ha tocat a mi ara t'ha de tocar a tu. Aquest esquema mental, molt arrelat (i sense res



a veure amb la «racionalitat»: no hi ha res racional que obligui res ni ningú a compensar el que ha passat abans), va començar a ser estudiat pel matrimoni Cosmides, «psicòlegs evolutius» *avant la lettre*, i també apareix a la *paradoxa de la loteria* («si hi he jugat tantes vegades ja és hora que em toqui») que pertany més aviat al punt 3.

5. Diferent percepció de present i de futur: No som capaços de comparar una cosa amb l'altra. Això és fa evident en experiments on s'ofereixen premis diferits: no hi ha manera d'imaginar-se el futur com si fos present, i la conseqüència és que cometem lamentables errors de càlcul (i no calen experiments: tothom té la pròpia història d'oportunitats perdudes i renúncies per un ocell a la mà).
6. Fal·làcies diverses, totes potser evolutivament justificables però (actualment) *disfuncionals* (*maladaptive*). Hi incloc
  - a) L' «efecte demostració»: apreciem una cosa no pas pel seu valor sinó pel sol fet que «la tenim», una mania que provoca des de l'avarícia de Molière fins al caos del nostre despatx, ple de coses que no recordem què coi eren.
  - b) La *hiperactivitat* (que sol anar aparellada amb el temut «dèficit d'atenció»): evolutivament, sembla justificat pel règim de vida associada al nomadisme però actualment només serveix per donar problemes als mestres en aules plenes.
  - c) L' «obesitat intel·lectual»: negar-se a fer canvis quan el context o la interpretació canvien. (Ho he batejat així per analogia amb l'obesitat física actual, resultat d'un tret adaptatiu –omplir els greixos quan n'hi ha ocasió– esdevingut absurd en època d'afartament permanent).
7. Obsessió per la «Interpretació» (de tot): Fenomen funest i conegut, un fet no té significat fins que li'n donem una interpretació plausible (vaja, que ens satisfaci), encara que la interpretació/solució que hi trobem no tingui res a veure amb la realitat ni expliqui res i, sovint, ens porti a explicacions desviades, demencials o simplement paranoies. Aquí potser també hi escauria l'enorme, exagerada, capacitat per reconèixer formes allà on

probablement no hi ha ni hi ha hagut mai res (un tret que tant ens pot portar a la intuïció feliç com a la forma de causalitat anomenada superstició).

8. Addiccions diverses, i quan dic *addiccions* no només vull dir 'addiccions' (és a dir, dependències) sinó, més en general, tot el que es deu a una «anticipació» del plaer esperat desconnectada del fet que sigui realitzable o no (i que produeix disfuncions com ara, entre mil altres, el consumisme, la comprateràpia i les infinites «teràpies» o els innombrables comportaments per substitució com, per exemple, regalar coses a un fill en comptes de, senzillament, estimar-lo, etc., etc.) (Els robots potser també ho faran, això, però previsiblement seran més difícils d'enganyar, o que s'enganyin.)
  
9. Fixació en falles del passat: Aquí no es tracta de revisar falles –al contrari, fer-ho és necessari per aprendre– sinó, com al punt 1, d'haver convertit l'experiència en trauma; es tracta no pas d'actes sinó de fixacions, de quedar-s'hi enganxat, de no saber-ne sortir, de fer-hi tombs al voltant com si de cada tomb n'hagués de sortir alguna cosa, en un exercici que, a més de dolorós, és inútil, que es repeteix inacabablement i que sovint s'intenta compensar o eliminar, ja sigui per «teràpies» de substitució, com en el punt anterior, ja sigui «oblidant» la causa («oblit» que, però, sovint no és tal: el subjecte continua actuant des de la ferida, tot i que ja no la recordi).
  
10. Raonament com a justificació: I si tingués raó Pascal quan deia que usem la raó no pas per dirigir els nostres actes sinó per justificar-los? (Potser val la pena citar aquí Musil: «M'estimo més [que em facin mal] els criminals: són els únics que fan mal sense sentir-se obligats a filosofar»). Aquest tret, com l'anterior, és específicament «humà» (vull dir, producte de la xarxa d'autocomplacitats que formen l'individu humà) i, per tant, sembla que als robots els ha de ser relativament fàcil de no caure-hi.

*Aleshores, com serà el segle XXII?*

Si aconseguim que els robots siguin racionals (si més no per suplir les nostres falles de racionalitat com les que acabem d'enumerar), com els veurem, nosaltres? La resposta és fàcil: sempre hi ha hagut gent que en un moment donat ha actuat racionalment. Els seus congèneres n'han dit que eren *savis* o bé *freds*, *inhumans*, segons l'empatia que s'hi tenia, i aquestes seran les impressions que previsiblement ens faran els robots racionals: els trobarem «savis». Si aquests saben afegir a la saviesa una mica d'empatia (de la qual hem parlat tot sovint, recordeu allò de les neurones-mirall, etc.), ens tindran tots al seu favor, i ens hi sentirem tan segurs i protegits al seu costat com proverbialment ens hi sentim amb els savis. Si a més afegeixen a la seva conducta, deliberadament o no, un adequada dosi de pífies pròpies (perquè ens puguem sentir reconfortats com sempre en les falles d'altri), aleshores la nostra simpatia està assegurada: considerarem els nostres robots «humans» sense dubtar. (Una qüestió a part és com ens considerarem nosaltres: pobres, ignorants i desgraciadets? Penseu que, pel que fa per exemple a la creativitat, el robots tindrà un camp amplíssim per explorar sistemàticament, comparat amb el nostre estret –literalment– camp de mines.)

*Però els robots, a més de savis, seran com nosaltres?*

El «mix» de característiques que ens fa humans conté irracionalitats diverses que no són gens *disfuncionals*; al contrari, són adaptatives (vull dir que ens han anat bé i ens han fet com som). N'esmento algunes:

- Reconeixement de formes exagerat, que ens fa veure cares en núvols (o complots en economia)
- El poder evocatiu enorme que tenen per nosaltres les imatges, les olors, etc.
- El joc social, en què, com es veu en molts experiments, imposem normes a seguir, càstigs als tramposos, etc.

- L'empatia en si («posar-se al lloc» d'altri), que és bàsica en primats (i possible en robots).
- L'adhesió que sentim pels «nostres» (sectarisme), o l'admiració cega per algú («enlluernaments»), la imitació servil, la moda, el voler ser original, etc.
- Tenir afinitats amb segons qui (i «sentir-s'hi bé»), categoritzar en «bons» i «dolents», demonitzar (i arribar a deshumanitzar) el dolent, etc.
- Por del ridícul, evitació de la soledat i l'aïllament.
- Percepció de l'actuació dels altres com a agressiva.
- Dificultat de superar pèrdues i/o fracassos.
- Tendència a donar explicacions voluntaristes (i/o religioses), especialment en cas de pèrdua (sobretot si és dolorosa)
- Concatenació factual (que dóna lloc a la superstició o a la causalitat científica, segons els casos).
- ...

Podem donar als robots algunes d'aquestes característiques i tindrem uns robots més o menys humans, més o menys com nosaltres. I és possible que en tinguem de diverses qualitats, segons per a què: robots «savis» (totalment racionals), robots que especulin a partir d'indícis, robots que evoquin situacions a partir d'una olor, etc., etc.

I potser haurem creat una nova espècie, el robot *sapiens*, amb qui probablement ens entendrem com a companys.

*Serà veritat, això?*

Probablement no.

Tot i la simpatia que puguem tenir per la Carme Torras i el desig perquè la seva predicció sigui precisa, les bromes de la història segurament s'hi interposaran. Recordeu que nosaltres pensem linealment, per extrapolació raonable, però la història té altres plans. En primer lloc, sol ser no-lineal (en el sentit de les equacions

matemàtiques): en un moment donat, la trajectòria, determinista o no, pot disparar-se en un punt de bifurcació en una direcció no sols imprevista sinó indeterminable. I, sentint-ho molt, de no-linealitats n'hi haurà. Qui hauria dit als dinosaures que els seus plans se n'anirien en orris per culpa d'un asteroide mal col·locat? I, els menuts mamífers de l'època, s'haurien pensat mai que acabarien sent els reis i donarien lloc a coses tan al·lucinants i totalment impensades com, per posar exemples absurds, els *reality shows* o l'exterminació industrial de congèneres?

La conclusió –no pot ser altra– és que tenien raó Sam Goldwyn («profetitzar és perillós») i el seu predecessor seriós, el nobel Peter Medawar (que venia a dir que «profetitzar és equivocar-se»). I ves que no passi que hi hagi una no-linealitat però que l'hàgim provocat nosaltres (la «singularitat» que segons el monotemàtic Kurzweil nosaltres mateixos estem preparant amb el que fem amb la informàtica i que, un cop passada, no hi haurà manera de saber què ha passat?)

Ui, me'n descuidava! Havia promès, més amunt, que diria qui hi havia entre la intel·ligència de l'home (mascle) i la de l'autista. Es tracta d'un espècimen molt conegut però, ja em perdonareu, ara no tinc espai per explicar-ho. Un altre article, espero...

I, mentrestant, llegiu la Carme, tot esperant que la temuda singularitat no ens ho trastoqui tot.

## *Seran proporcionals, els robots?*

**A** primers d'any, Israel va invadir la franja de Gaza. La situació –enquistada– és perfectament coneguda: els protagonistes (diguem-ne «els palestins» i «els israelians» per simplificar) assumeixen en cada cas el paper de víctimes (uns de la prepotència dels enemics i els altres del fanatisme dels veïns). Es veuen a si mateixos ocupats o amenaçats per l'altra part, i es queixen d'un tracte injust, i de no poder construir la pròpia societat en pau i en llibertat. Immediatament, davant dels fets de guerra d'aquells dies, la posició dels observadors externs no va ser d'explicar-se la situació (és sobradament coneguda), sinó de mesurar si l'acció d'Israel era o no «proporcional».

Com en tot el que és humà, doncs, hi ha quatre ingredients: (1) una situació («de què va la cosa»), que tots els protagonistes copsen més o menys bé i davant de la qual tots ells es posicionen identificant-se amb (2) un paper propi («què hi fan ells allí, o que els hi toca fer»). D'això en resulta (3) una conducta, la que cada un dels actors acaba tenint. Ara bé, qualsevol que sigui aquesta conducta, sempre hi ha (4) una qüestió de grau: no és igual una agressió verbal que una de física, i la possible innocuïtat o brutalitat d'una acció es mesura en una escala que va fàcilment, imperceptiblement, del zero a l'inacceptable.

Per què ho dic, això? Doncs perquè en un món amb actors presumptament intel·ligents, siguin robòtics o humans, si bé és important amb quin material actua cadascú (amb sentiments, i potser empatia) encara ho és més la manera en què es perceben les situacions, com s'interpreten, com cada un s'hi posa i amb quina intensitat s'hi actua. En el número anterior vaig explicar que és segur que els robots tindran sentiments, empatia i diguem-ne

NODES  
NÚMERO 43  
PRIMAVERA DE 2009

«consciència», s'especialitzaran en algun camp en què nosaltres fluixegem (i ens hi superaran: per a això els fem), seran creatius i ocurrents, i hi podrem tenir un tracte si fa no fa fluid i desproblematitzat. Això sembla bastant obvi i imparabile, i personalment no em preocupa gaire. M'amoïna molt més, però, com serà la seva conducta, més que no pas les seves aptituds. Perquè, és clar, si fan res que valgui la pena, hauran d'actuar, i hauran, doncs, de prendre decisions (penseu per començar en un soldat –imagineu-lo robòtic– enmig d'una situació complexa i tensa). La meva preocupació neix d'haver vist en nombrosos episodis històrics i/o personals en què consisteix la conducta humana intel·ligent (això inclou la «creativitat») i d'haver constatat les diverses i sorprenents formes que aquesta conducta pot arribar a adoptar (que van de l'admirable a l'horripilant). I em basqueja pensar que aquestes conductes, que en el cas humà potser no tenen remei, puguin ser també les dels robots. I aquí és quan, més enllà de reclamar sentiments o simpatia robòtica, m'horroritzo només de pensar en les conseqüències de tenir robots que puguin ser (sobretot, que actuïn) com nosaltres. Per això tinc interès, vistos els problemàtics antecedents històrics, de mirar-me amb lupa l'anomenada conducta intel·ligent, que no passi que hi reincidim, però aquesta vegada amb robots, i potser havent-ho de lamentar. Abans, però, una qüestió prèvia:

### *De homes, dones, autistes i matemàtics*

En el número anterior vaig insinuar, en passant, que la intel·ligència humana presentava un contínuum que va des dels individus (típicament, les dones) amb molta capacitat de copsar situacions interpersonals complexes –i saber-les expressar lingüísticament– fins als «homes», amb una possiblement més gran capacitat analítica i/o espacial que no acaba de compensar els dèficits en les capacitats «femenines». Aquest contínuum s'estén per la banda «masculina» fins als autistes, individus que sorprenentment extrapolen els trets «mascles» (són aclapadorament nens) i es caracteritzen per pràcticament no tenir relacions interpersonals (en termes tècnics, no tenen *teoria de l'altre*). (Ep!, estem parlant estadísticament: no vol dir que un home o una dona hagi de ser de cap manera; sobre aquest mapa cada individu s'hi situa de manera pròpia.) Doncs bé:

en aquest espectre s'ha trobat un tipus intermedi entre el mascle típic i l'autista: és el matemàtic.

Efectivament: a Ioan James, un topòleg en exercici, li van cridar l'atenció l'actitud, hàbits i reaccions d'alguns matemàtics (històrics). Va ser ell el qui, amb l'ajuda de metges, va lligar conceptualment l'activitat d'aquests matemàtics concrets (ell va sospitar de molts, incloent-hi Newton i Einstein), amb la síndrome d'Asperger, una forma relativament benigna d'autisme. (La llista de «suspectes autistes» que n'ha resultat, comentada per neuròlegs i psiquiatres, és a: [http://en.wikipedia.org/wiki/People\\_speculated\\_to\\_have\\_been\\_autistic](http://en.wikipedia.org/wiki/People_speculated_to_have_been_autistic)). Tenir l'Asperger, sembla, força el cervell a una gran concentració –o la permet, segons com es miri– en activitats de mecànica mental, alhora que –convenientment, sens dubte– provoca un simultani allunyament –emotiu, si més no– de les coses de cada dia. Tot això recorda el «paradís» matemàtic cantorià de què parlava Hilbert, o allò d'en Bertrand Russell, que, enmig d'una situació personal trencada, deia que sort en tenia de la Matemàtica: allà no hi havia maldecaps ni friccions interpersonals que el fessin patir. Vaja, que sembla que tenir Asperger és tenir al cervell un bon mecanisme de «focalització» (que fóra perfecte si no fos que és una malaltia, i que els personatges en qüestió en solen resultar més aviat asocials o antisocials). Aquest tret ha afectat moltes figures històriques, a qui d'infants, típicament, els havia costat molt de parlar, com ara els físics Albert Einstein, Edward Teller o Richard Feynman, la matemàtica Julia Robinson o els músics Arthur Rubenstein o Clara Schumann. Per cert, hi ha un llibre que tracta el tema: és *The Einstein Syndrome: Bright Children Who Talk Late* (2002), de Thomas Sowell, que va d'aquests «late-talkers», sovint (mal) diagnosticats com a afectats d'autisme o de PDD (*pervasive developmental disorder*), en què l'autor –un cèlebre economista americà (negre, i «de dretes») i pare d'un fill aspergerià– es fa ressò de la recerca de Stephen Camarata (de la Vanderbilt University) i de Steven Pinker (de Harvard) sobre aquests infants, que es desenvolupen «asíncronament» per l'efecte d'un creixement tan ràpid com extraordinari de les funcions cerebrals analítiques (creixement que temporalment «roba recursos» a funcions veïnes, com ara les lingüístiques).

Per què parlo d'això en un context de robòtica? Doncs perquè



sembla raonable pensar que els robots, que presumiblement tindran una enorme capacitat de càlcul (que a nosaltres ens semblarà indistingible de l'omnipotència), podran tenir incorporades d'entrada les capacitats analíticoespacials diguem-ne «masculines» i, alhora, poder «calcular» situacions i contextos (i trobar expressions lingüístiques que s'hi adiguin) almenys com ho faria «una dona» (alhora que aplicar-se en càlculs mecànics i desvinculats de l'entorn que, en el cas humà, requeririen un autista o un matemàtic aspergerià). Però tornem on érem:

*Hi ha situacions i situacions (i hi fem papers i paperots)*

Exemple 1: Si ens trobem davant d'un problema (per exemple, un atac epilèptic d'algú o un possible foc deduïble de la presència de fum), la possibilitat d'actuar per resoldre'l sol ser, experimentalment, que un 78% dels presents actuen (vull dir que avisen o demanen ajuda). Ara, si el subjecte (veu que) no és l'únic present, la probabilitat d'actuar baixa al 34%. És a dir, que de cada 12 sers humans podríem dir, simplificant brutalment, que 9 són «bones persones» però que, quan veuen que no estan soles, les que continuen sent-ho són només quatre. Què ha passat? Als protocols de l'experiment –històric, fet a la Universitat de Columbia el 1964– hi sobresortien dues explicacions/excuses: (a) «si hi ha gent, se suposa que algú altre, amb més autoritat, coneixement o recursos, actuarà»; i (b) «si ningú no actua deu ser que el problema no és tan important, o que potser només m'ho sembla a mi, que hi ha problema». No és que un factor accidental (la presència d'altres) canviï la polaritat moral de cop: no és raonable que una «bona persona» ho deixi de ser tan de sobte o que un comportament passi tan volàtilment d'encomiable a censurable només per la introducció d'una peça (un testimoni més). El curiós és que aquest experiment va ser fet immediatament després del cas –i de la perplexitat consegüent– de la Kitty Genovese, una noia de Queens que acabava de ser assassinada al carrer, de nit, malgrat que demanava ajuda a crits i que l'episodi el van presenciar no menys de 38 veïns des de casa seva, cap dels quals va avisar la policia ni va fer-hi res de res. (Aquest fenomen és tècnicament conegut com el *bystander problem*.)

Exemple 2: En un experiment dit, per raons òbvies, «del samarità», uns estudiants passen al costat d'una persona ferida i un 63% s'atura i l'ajuda. Tot seguit, s'agafa els estudiants i se'ls diu que tenen per objectiu arribar a un cert lloc en un temps relativament ajustat; ara els estudiants es tornen a trobar un ferit i, de cop, només el 10% s'atura. Les excuses? «Home, és que tenia pressa», «és que m'havien dit que era molt important que no fes tard!», etc.

Exemple 3: Estem –nosaltres (vull dir persones «normals» i bàsicament cíviques)– davant d'un semàfor en vermell. De cop, un altra persona passa. De cop, la probabilitat que nosaltres també passem s'ha multiplicat per tres. I qui diu passar un semàfor diu deixar-se encomanar per una conducta possiblement «dolenta» (aquesta diferència moral es demostra, experimentalment, irrellevant). Això s'accentua sobretot si qui ens l'encomana té «prestigi» o rep una bona valoració nostra o és un company (és allò que en anglès en diuen sentir la *peer pressure*).

Exemple 4: En un altre experiment, aquest fet l'any passat a Groningen (Pp. Bb.), es comprova la realitat de la *teoria dels vidres trencats*: el nombre de persones que es comporten cívicament (no llancen papers a terra, respecten les prohibicions, etc.) passa de cop a ser un terç (o un quart) quan a l'entorn s'hi veuen indicis de desordre o descurança (papers a terra, brutícia, *graffiti* a les parets, bicicletes abandonades, etc.) respecte de quan l'entorn és net. En una variant de l'experiment, es deixava un bitllet (de 50€, posem), visible, en un sobre mig introduït a la bústia d'una casa: el nombre dels qui s'emportaven el bitllet es va multiplicar per dos només pel fet de que hi hagués *graffiti* a la porta en qüestió. L'excusa? «Home, és que ja es veia que d'allò no se'n cuidava ningú» o «si no l'hagués agafat jo ho hauria fet un altre». La qual cosa sembla mostrar que, per sobre del civisme o la «moral», hi ha les circumstàncies (la situació), la interpretació que en fem, i la nostra (irresistible) tendència a adaptar-nos al que veiem.

Exemple 5: Stanley Milgram (sí, el dels «sis graus») va plantejar el 1961 un experiment en què diversos estudiants –teòricament bones persones i de casa bona (no debades Yale és una universitat de peatge)– havien de donar descàrregues elèctriques a companys quan els ho digués l'instructor. Doncs la majoria ho feien sense despentinar-se, i fins i tot alguns s'hi abonaven amb delit. L'experi-

ment es va haver d'aturar perquè tothom va veure, horroritzat, que el mal és una bèstia fàcil de despertar. Les excuses? «Home, és que si veig que tothom ho fa...!» «És que si m'ho diu el profe...!» I el curiós del cas és que Milgram havia plantejat l'experiment, precisament, a conseqüència de les declaracions del nazi Adolf Eichmann, jutjat aquells dies a Israel: les seves excuses eren «És que volia quedar bé amb els meus superiors» o «La meua responsabilitat era que els trens arribessin a l'hora a Auschwitz... i bé que hi arribaven; vaig complir el meu deure: no veig pas que me n'hagi de penedir» (excuses que van suscitar el famós comentari de Hannah Arendt sobre la «banalitat» del mal). Milgram va descobrir que «el mal» no calia anar-lo a trobar a Auschwitz, i que a Yale mateix n'hi havia reserves. Un experiment més sonat encara es va fer deu anys més tard a Stanford (el famosíssim *Stanford prison experiment*): ara es tractava d'imaginar-se una presó i fer-la funcionar a base d'estudiants voluntaris; el resultat va ser tan horrífic que el mateix experimentador, Philip Zimbardo, no ha deixat de donar-hi voltes des de llavors (ho explica al seu darrer llibre *The Lucifer Effect*).

### *Saber parar*

Els exemples anteriors (alguns dels quals he manllevat de *The Tipping Point*, de Malcolm Gladwell, ara autor de moda) demostren, si és que demostren res, que la distància entre el «bé» i el «mal» és tan prima com un paper de fumar; o, més ben dit, que una mateixa conducta pot ser tant bona com dolenta, tant elogiabile com condemnable, en funció de com ens hi posem, de com els actors interpreten la situació. D'aquí ve la meua particular prevenció anunciada al títol d'aquest article: potser no tot és qüestió del que fem sinó de la seva proporcionalitat o invariància: si entenem la situació (algú ens demana ajuda, cal fer alguna cosa) potser cal que no ho reinterpretem o relativitzem quan algun factor canvia (la presència d'algú altre, l'acomodació al que veiem que fan els altres) i ens permet autocomplaure'ns amb excuses. Des d'aquest punt de vista, l'esperança que els robots actuïn millor que no pas nosaltres es basa en la seva extraordinària capacitat de calcular («imaginar-se») situacions possibles i –així i tot– poder mantenir les pautes inicials de conducta sense deixar-se influir per factors

afegits (com ara si hi ha gent, si hi ha brutícia, etc.)

### *La creativitat del mal*

És obvi que els nostres robots voldrem que siguin creatius. Quan pensem en aquesta virtut de seguida ens sol venir al cap el científic benèvol o l'emprenedor audaç, tot i que sospito que la creativitat també, i sovint, ve de l'agent que mira com pot trencar les regles (i no em fico en si això és bo o no). No puc deixar de pensar en el grau (enorme) de dependència que la ciència «normal» ha tingut de les eternes curses d'armament amb la seva concatenació d'acció i reacció (des de les llances o els arcs compostos fins a les bombes de fragmentació passant per la indústria dels panys i les claus tota sencera, la criptografia i les signatures bancàries) ni tampoc en l'activitat i l'existència mateixa dels *hackers*. De vegades resulta difícil no reconèixer en els nazis alemanys el mèrit d'una creativitat extraordinària: ningú com ells va portar la industrialització i l'eficiència fins a terrenys tan insospitats com la carnisseria humana o la producció massiva de sofriment, de tal manera que a ells devem l'extraordinària atenció que dediquem avui a qualsevol tret dictatorial, violació de drets o falta de democràcia, així com el també extraordinari període de llibertat, irrenunciabilitat democràtica i megaprospèritat que hem tingut a Europa durant seixanta anys gairebé exactes. Potser la creativitat, com el nostre «civisme», no és tant un producte de la moral o la bona voluntat sinó d'alguna cosa, desconeguda, l'únic efecte palpable de la qual és la seva efectivitat a la llarga. Si és així, com ho explicarem als nostres robots i què els direm que facin perquè en siguin, de creatius?

*Si a l'origen no hi ha moral, a la destinació potser hi haurem de posar «ètica»*

Ara bé, com que no hi ha actuació sense error ni conducta sense un mal càlcul, i com que, per molt que puguin calcular, als robots sempre se'ls podrà escapar un càlcul impossible (perquè sigui irrealitzable en el temps donat o perquè depengui de factors desconeguts, arbitraris o aleatoris), caldrà que anem preparant també un marc de regles on situar les fallades o infraccions. Un

vell conegut nostre, Noel Sharkey, de Sheffield, va publicar fa relativament poc a Science (322, 5909) un article titulat «The ethical frontier of robotics», en què demana, ara que encara hi som a temps, començar a plantejar una sèrie de problemes relacionats amb la nostra convivència amb robots i de què aviat sentirem a parlar, començant per alguns que ja ens afecten ara i que pegen d'un fil ètic molt fluix. En particular ell cita: (1) els originats en la relació tan peculiar entre robots de companyia o assistència i els seus companys humans (la canalla o de la gent gran de qui se suposa que s'han de cuidar), tot preveient casos que es puguin esdevenir de conflictes, errors o disfuncions; i (2) els sorgits dels robots usats en contextos militars que, com ja comença a passar, són autoritzats a prendre decisions crítiques (localitzar i/o destruir objectius, per exemple), que demanen una avaluació detallada i agònica i que, sobretot en cas de risc, de mal càlcul o de desgràcia, plantegen un problema inèdit –ètic–, clarament.

### *Somiar truites*

És conegut que els humans tenim al cervell un servei de recompensa-plaer que, segons com es miri, és el nostre pitjor enemic. Fixeu-vos en aquesta comprovació experimental recent: injectar una solució salina a un pacient dient-li que és morfina té el mateix efecte (el mateix!) que injectar-li morfina de debò. Això vol dir que el que ens satisfà no és l'acompliment del plaer sinó la seva mera anticipació (que és només un regalim momentani del còrtex prefrontal, on es processen els plaers). Quantes vides s'han orientat a aconseguir alguna cosa merament pel plaer –imaginat– que (se suposa que) es tindrà quan s'aconsegueixi! Evidentment, no hi ha cap relació lògica entre una cosa anticipada i la seva realització efectiva (en molts casos aquesta realització és factualment impossible). No sols s'han construït (i destruït) moltes vides en funció d'una satisfacció última sinó que es podria dir que la majoria de construccions intel·lectuals de la humanitat són anticipacions quimèriques d'una visió estèticament plaent merament imaginada. Penseu, per tornar al cas inicial, com s'imaginaven israelians i palestins el seu futur ideal –que precisament consisteix en el fet que «l'altre» no hi sigui, ni hi hagi sigut mai (aquesta darrera condició metafísicament impossible).

Doncs, pel bé de tots, cal que els nostres robots s'apliquin a no tenir utopies en el mal sentit del terme (vull dir objectius «personals» que guiïn la seva conducta), que puguin tenir el cap fred en tot moment, i no condicionar la seva conducta racional a les pujades de temperatura momentànies o possibles imatges inflammatòries sense connexió amb la realitat.

### *El robot perfecte?*

En resum: una actuació intel·ligent (que inclou decisions) no s'entén sinó en un context, una situació, que l'agent «interpreta» i davant de la qual es posiciona precisament en funció de la seva interpretació, sensible als canvis que hi pugui detectar. Actuar és «immergir-s'hi», i un robot s'hi ha d'immergir totalment, sense els nostres condicionants filogenètics (dels quals vaig parlar a bastament al número anterior) ni la nostra precarietat calculatòria. Això val també per a la seva superació, via l'«omnipotència de càlcul», de la fragmentació de la intel·ligència humana en segments (femení, masculí, aspergerià i autista). Cal també que un robot posi fre a la irreprimitible humana tendència a fer volar coloms, a escalfar-se la boca i a somiar truites. Ha de compensar tot això amb «proporcionalitat», l'equivalent del «cap fred».

Sobretot, i addicionalment, ha d'evitar de fer com fem nosaltres, que: (1) sempre sobrevalorem les nostres capacitats i subvalorem les dificultats i el temps, que (2) sempre suposem, d'entrada, que el món ens deu alguna cosa, que (3) sempre actuem sobre el món no tal com és sinó com ens agradaria que fos, i que (4) sempre exhibim, gràcies a la nostra pobresa de càlcul, una demostrada incapacitat per predir el que passarà perquè, entre altres coses, també som incapaços de preveure com ens sentirem nosaltres aleshores (sobre això, vegeu «Being Human», al Nature del 8 / 01 / 09). Per aquestes raons i les de més amunt, hauríem d'evitar de totes passades que els robots tinguessin una conducta tan erràtica i poc sòlida com la nostra, a la qual desgraciadament es pot aplicar aquesta paràfrasi de la llei de Grosch: «Qualsevol que siguin els plans, els resultats sempre seran inapropiats (o decebedors)».



## *Hauríem tingut crisi, amb robots?*

**N**o tenim excusa: ja fa més de tres-cents setanta anys que sabem què és una bombolla financera. La primera que tenim detalladament documentada és la de la «tulipomania» holandesa, que es va reventar el 1637. Totes les bombolles financeres s'assemblen –com s'assemblen les de sabó– fins i tot en la manera en què acaben petant. Totes tenen una fase d'ascensió caracteritzada psicològicament per una hiperconfiança contagiosa i per una sensació de prosperitat imparabile i gairebé miraculosa que no tenen contrapartida en la *racionalitat* (pel que fa al comportament humà) ni en les lleis de l'economia (pel que fa a la realitat).

La pregunta és: ja que la component psicològica, inevitable en humans, sembla determinant, podria una intel·ligència no humana evitar caure-hi i estalviar-nos a tots la bombolla i, sobretot, el desgraciat pet final?

Per tenir-ne alguna idea potser caldrà examinar abans alguns dels components psicològics que –en el cas humà– hi intervenen. I que cadascú dictamini si les nostres màquines intel·ligents els tindran o és aconsellable que els tinguin, i si fer-les invulnerables als nostres, diguem-ne, defectes evitarà bombolles futures.

*Ep!, que les crisis econòmiques són coses serioses i freqüents i, sobretot, econòmiques*

Perquè ningú no s'enganyi. Una economia com la nostra és un sistema complex que encara no comprenem bé, i ja fa temps que sabem que té un component cíclic potser inevitable. Tot i que els cicles formalment s'assemblin, segurament una crisi de sobrepro-

NODES  
NÚMERO 44  
TARDOR DE 2009



ducció no és del mateix tipus que una de subproducció, per posar dos exemples. En el cas de la que es va desfermar el setembre del 2008, sembla –o això és el que diu el consens– que la causa subjacent va ser una epidèmia de diner barat. Tan barat que els compradors es van posar a comprar de no dir –segurs que les propietats que adquirien els feien invulnerables–, mentre els prestadors compensaven l'estretor dels interessos amb noves figures d'«enginyeria» financera de solidesa dubtosa, algunes vorejant el malabarisme, i sovint amb l'ajut inestimable d'ordinadors que, pobres, no en tenien cap culpa.

A banda dels mecanismes pròpiament econòmics que tenen entretinguts els economistes, el que vull fer aquí és merament explorar els mecanismes psicològics que han acompanyat aquesta bombolla, i totes les altres, per saber si autosubstituint-nos per màquines hi ha cap esperança de millorar la cosa.

### *Les il·lusions del còrtex*

El culpable de tot és el **còrtex prefrontal**. El recordeu, el còrtex? És aquell tovalló arrugat de gruix subcentrímetric (3 mm) que a Lausana IBM i els locals proven de vigilar estretament i de simular, intent de què ja us he parlat aquí un parell de vegades i en què jo particularment tinc grans esperances d'illuminació. Doncs és en aquest menut territori cerebral on es fan –fem– les següents tres operacions –totes força malament, per cert– que cito sense ordre d'importància (hi és?) i que es poden resumir en això: *processar decisions* tenint en compte –més o menys– les *circumstàncies*. De tal manera que una extirpació d'aquesta zona –com la que va sofrir accidentalment Phineas Gage a Vermont el 1848 (en una explosió una barra de ferro li va entrar per l'ull esquerre i li va sortir pel darrere del cap)– no és pas mortal: al contrari del que pugui semblar, el subjecte pot continuar vivint normalment si bé –almenys a l'ultraestudiat Gage li va passar– ja no té tant en compte els contextos ni prova d'equilibrar els múltiples components d'una decisió –a Gage li va canviar la personalitat i va morir el 1860 com si fos un altre, com si un alter ego, més asocial, l'hagués substituït (qui ho explica molt bé i resumidament i que jo he citat més d'una vegada aquí és l'Antônio Damásio<sup>17</sup>).

<sup>17</sup> Antônio Damásio. *Descartes' Error: Emotion, Reason, and the Human Brain*. Harper Perennial, 1994.

Les tres operacions que deia que té encarregades el còrtex prefrontal són, entre altres –i citades desordenadament– les següents:

- (a) *Comparar situacions*, previsiblement per veure quina és més adequada, correcta o favorable vistes les *circumstàncies*, i decidir-s'hi. (Les dues operacions següents, la (b) i la (c), vénen a ser casos particulars d'aquesta.)
- (b) Comparar un *guany* possible amb una *pèrdua* possible: el guany és vist com una recompensa –i per tant desitjada– i és associat a «plaer». Aquí s'hi afegeix que el mecanisme de secreció d'opiacis cerebrals, característic de l'obtenció de recompenses, sovint s'excita prematurament en aquest mateix moment, davant la mera «anticipació» d'una recompensa –no només la seva obtenció posterior–, cosa que explica més d'una addicció i més d'una obsessió o trajectòria vital. (I si la «visió» d'una recompensa segrega *endorfines*, la d'una pèrdua excita la petita *amígdala*, seu de la por i el pànic, al bell mig del cervell.)
- (c) *Comparar* una cosa (guany o pèrdua, o el que sigui) que es projecti en un *futur* amb la mateixa cosa situada al *present*, davant nostre i (suposada) immediatament accessible.

No cal dir que els humans tenim greus defectes de fabricació en totes tres funcions. Les actuacions clarament anòmales o fora de context (per no parlar de les asocials) són la conseqüència d'un mal funcionament de la funció (a). (La *il·lusió monetària* que discutiré més avall n'és un cas, com també ho són les estrepitoses falles en la *predicció* tecnològica de què podem presumir tots plegats i de què he parlat sovint tot invocant Sam Goldwyn i el seu savi consell de no profetitzar i menys sobre el futur.)

L'operació (b) no cal que funcioni malament per portar-nos males conseqüències: la seva asimetria –semblantment a l'asimetria temporal de la funció (c)– explica el comportament, supercomprobat, que no ens és gens igual perdre 400€ quan abans n'hem guanyat 600 que guanyar de cop i del no-res 200€, tot i que la suma algebraica sigui 200€ en tots dos casos. (Explicació sumària: l'acció de les addictives *endorfines* –el plaer, recordeu?– i la de la temible *amígdala* –una fuetada de por sobtada–, tot i que de signes contraris, no són equiparables ni en mode ni en volum o temps.)

L'operació (c) és, no cal dir-ho, la responsable de la nostra pertinaç falta de previsió (d'estalvis, de benestar, d'amor o del que sigui).

### *La il·lusió monetària*

Aquesta il·lusió, que és un cas particular de la (b) i és un clàssic entre els economistes, és segurament la més òbvia en [la fase ascendent d'] una bombolla. No fem cas, ni que ens matin, del fet evident que la inflació, com les marees, ho fa pujar tot. Ens capfiquem en el sucós preu de previsible venda de casa nostra, que no para de pujar, i no veiem que tot puja (i sobretot també les altres cases), cosa que fa pensable el miracle dels pans automultiplicatius. Tothom veu que el totxo no baixa, però ningú és capaç de recordar quant pagava per anar al cinema fa quatre anys. I ningú s'adona d'una concatenació casual tan òbvia com que si ens hem hipotecat és perquè el baix preu dels préstecs i el nul control –pels bancs en fase alegrement ascendent– de la possible morositat en què incorren ens han fet oblidar a tots que si per un atzar l'aixeta es tanca en un dels seus múltiples punts la cadena tota sencera se'n va en orris. Per seguir el símil, veiem pujar la nostra barca sense voler-nos adonar que és la marea la que puja (en paraules d'un antic contraproverbi, «mal de molts, ... epidèmia!»). En termes econòmics, aquest canvi de rasant es descriu així: «exuberància irracional» (en frase d'Alan Greenspan), la pujada, i «col·lapse de liquiditat», la «pinya» que ens hem clavats.

### *Les altres il·lusions*

Els trets que ara direm, que de vegades podem considerar disfuncionals, s'afegeixen a les il·lusions anteriors (o en són part, en un tot complex i intricat) i poden formar part del nostre desenraonament propi de crisis i bombolles. Només en cito unes quantes (cinc: bàsicament dues, més tres de complementàries):

- 1) *Sobrevaloració de capacitats*. És universal i fàcilment detectable en experiments. Tothom –i aquí s'hi compten des dels optimistes als severament deprimits– creu fermament que, do-

nada una feina proposada o complerta, ell (o ella) és capaç de fer-ho millor del que la prudència permet esperar o que, en el cas d'una feina ja feta, la «nota» que es mereix és més alta del que realment li correspon. Una valoració paral·lela, també amb conseqüències a la baixa, és la del *temps*: tothom creu que hi ha més temps del que realment hi ha. Cosa, que, unida a l'anterior, justifica aquella esplèndida llei de [l'astrònom i informàtic] Herb Grosch i aplicable a tots els projectes d'enginyeria –o de vida en general: «qualssevol que siguin el pressupost o el termini d'execució previst, aquests dos (és a dir, tant els diners com el temps) sempre seran insuficients».

- 2) *Biaix de confirmació*. Aquest, universal i omnipresent, explica que, encara que pretenguem ser imparcials i impassibles en la recerca de la veritat, aclaparadorament apreciem els indicis confirmatoris del que creiem i ignorem sobiranament tot allò que ho desmenteix i ens pugui treure del núvol confortable de les creences instal·lades. Això explica per què ens costa tant canviar de diari o de prejudicis polítics, i també per què ens sorprenem de miracles o «causalitats inexplicables» –i arribem a creure que «volen dir» alguna cosa– sense adonar-nos que en començar el dia hi ha milions de possibilitats obertes –incloent els aparents «miracles» (p. ex. haver pensat en algú just al moment que aquest algú es moria o ens telefonava, cosa que té una probabilitat calculable i clarament superior a zero)– i que ens passem el dia filtrant d'esma i automàticament megamilions (literalment!) d'escaïences només perquè «ja hi comptem» i ni tan sols notem. Tot passa simplement perquè el nostre mecanisme de confirmació, colgat profundament al nostre *còrtex prefrontal*, ens ho fa tot transparent i suau. El *prefrontal dorsolateral*, associat al raonament, ignora qualsevol evidència contrària a la nostra plaent i tranquil·litzant creença, mentre que els que realment s'activen són aquí el *còrtex orbital*, que processa emocions, el *cingulat anterior*, que mira de resoldre conflictes –quan no hi ha més remei–, i el *posterior*, la feina del qual és emetre judicis morals. Quan tot aquest complex aparell –comprimit en uns pocs mil·límetres, no ho oblidéssim– ja ha fet la seva feina (d'ignorar tot el que «ja s'esperava»), entra en acció l'*estriat ventral*, content

i feliç, que és el que assenyalava que s'ha arribat a la recompensa i plaer subsegüent. Vol dir que ja hem acabat de llegir el diari i hem confirmat el que pensàvem abans d'obrir-lo, o que en tot el dia no hem vist res de prou nou o d'estrany ni cap casualitat inexplicable. . . tot i que –siguem objectius se n'han produït una pila (les que les lleis de la probabilitat marquen)– i que –la probabilitat de tots els esdeveniments, estranys o no, suma necessàriament 1. No cal que recordi que estàvem parlant del creixement de *bombolles*, oi? i que aquestes creixen per un procés d'autoconfirmació recíproca i encomanadissa.

[I atenció! això no passa només en bombolles *econòmiques*: un fenomen psicològic equivalent explica l'aparició sobtada de *modes*, d'estats de *pànic* diversos, de climes d'*odi* social intens o, en un cas històric conegut, el canvi social contagiós i imparable que va afectar Alemanya els darrers mesos del 1932 i que va acabar, al gener següent, amb Hitler al poder, democràticament elegit –i totes les llibertats i jueus proscrits al cap d'un mes!; aquí la bombolla va quedar 12 anys inflada i l'«esclat» va exigir una guerra terrible, sota la incredulitat dels mateixos que en la fase ascendent –en un tot inextricable d'entusiasme i por– ho havien fet possible... Però, vaja, tornem a les nostres «bombolletes» simplement econòmiques, perquè, benignes per comparació, segurament són menys llargues i solen no exigir cap guerra per acabar-les de petar.]

### *Algunes il·lusions suplementàries (que hi ajuden força)*

- 3) *Efecte dels companys i de «la massa»*. És el que també s'anomena «influència del ramat» (o «herding instinct» o «aborregament», segons la llengua que es triï). Em fa l'efecte que és tan prou conegut i sentit, que no cal abundar-hi. A petita escala se'n diu *por de sentir-se diferent* o *peer pressure* i potser n'hi haurà prou si us dic que és justament això el que van sentir els confiats inversors del hipermaximurri i suprem pocavergonya Madoff quan veien que tots els entesos creien, contra tota lògica

elemental, que invertir-hi era possible, rendible i, a sobre, una cosa *cool*.

- 4) *Biaix retrospectiu*. Consisteix a dir que «això ja es veia que passaria», i dir-ho mentre passa (no pas després, que ja no té tant de mèrit). És prou evident que tots nosaltres –m’hi incloc (i amb el vostre permís ens hi incloc)– sabíem que «això no pot acabar bé», que «això petarà» i, ara que ha petat, ens fem creus de com és que no hi vam fer absolutament res per aturar-ho, mitigar-ho o, senzillament, posar-hi una mica de seny. Forma part de l’anomenat biaix «retrospectiu», però curiosament i contradictòriament té lloc abans del pet, és a dir, «prospectivament»: sabem que anem al desastre, però hi anem contents i acompanyats.
- 5) *La illusió de racionalitat*. Aquesta ja no pertany al nostre còrtex sinó al dels economistes, que s’han passat una pila d’anys predicant que l’home com a subjecte econòmic és bàsicament «racional» (volent dir que maximitza el seu interès o «utilitat» esperada). Com que els economistes són en general gent normal, ja s’han adonat que l’economia no va exactament així, i s’afanyen a construir una ciència econòmica «conductual» (*behavioural*) que doni compte de tots aquests comportaments sub-racionals o no-racionals. Sovint veiem economistes que suspenen el judici fins haver vist els resultats dels escanejos cerebrals tomogràfics que fan els neuròlegs, a les portes de fundar una nova ciència, la «neuroeconomia», que ens dirà per què fem les coses (i també, atenció, per què comprem un producte d’una marca i no pas de l’altra). Només cal esperar que la nova ciència avanci i ens il·lumini (i que els publicitaris se’n mantinguin apartats). Ara, creure o esperar, com fan segons quins economistes (potser per purgar els seus pecats) que aquesta «ciència» ens evitarà bombolles ja és somiar truites. Si volem estalviar-nos-en algunes potser ens caldrà fiar-nos més de les nostres estimades i potents màquines intel·ligents que no dels gurus (humans) limitadets que sempre hi ha al mercat en oferta.

### *Quin «còrtex» donarem als robots?*

Vistos els eventuais estralls perpetrats pel nostre còrtex prefrontal, especialment quan es despista, què farem, doncs, amb els robots que dissenyem? Els donarem un còrtex humà o parahumà? O els concebrem i construïrem de nova planta, sobre nous principis? (quins?) Difícil, això.

Una possible resposta és la següent: doncs donem-los un còrtex (també «prefrontal»? ). Aparentment, és una opció sensata, ja que el mecanisme cortical sembla bàsicament prou ben dissenyat (aquí demano perdó per l'antropomorfisme del «dissenyat»: vull dir que si n'haguéssim de dissenyar un nosaltres, potser no fóra gens desenraonat seguir els plànols bàsics del que ja tenim). Ara, això sí, hauríem d'evitar que els robots fessin com fem nosaltres de vegades: que es pengessin, que es tornessin addictes de *vés-a-saber-què*, que somiessin truites vàries, que els costés tant de veure l'evidència i canviar d'opinió, o que persistissin en una bogeria col·lectiva la fi i desastre de la qual tothom veu clara.

Jo no tinc gaires dubtes que això serà viable. I probable. No veig gens forassenyat que puguem algun dia tenir robots que, tot actuant d'una manera (racional) paral·lela o equivalent a la nostra, puguin evitar les falles més evidents de la nostra arquitectura mental. Al cap i a la fi, segons penso, tots els nostres «defectes» de funcionament precortical són essencialment producte de limitacions de càlcul, dèficits deguts a l'extraordinària rudimentarietat dels materials que componen i enllacen les neurones, de la seva escassa fiabilitat de components i del poc espai i temps que tenen. Les màquines que fem (i farem) tenen una capacitat i rapidesa de càlcul tan totalment incomparable amb el nostre maquinari cerebral que, precisament per això, poden actuar com nosaltres sense nosaltres (vull dir, suposo que s'entén, evitant les nostres *gaffes* més evidents).

M'imagino perfectament robots que, en el seu paper de pare carbasser benèvol, ens haguessin dit: «no siguis ximple, no veus que això ha de tenir necessàriament un punt d'inflexió?» o «no facis cas d'aquest Madoff: els comptes no surten de cap manera», «això no pot ser veritat» o «per quin set sous creus que a tu no et passarà el mateix que passarà a tothom?» o «per què no baixes del

ruc i afrontes els fets d'un cop?», etc. (Bé, això no és potser gaire diferent del que jo suggeria al número anterior, a propòsit de la «proporcionalitat» –o no– humana i de la necessitat de tenir robots que facin de pare-pedaç i ens vagin dient coses com ara «vols dir que no n'estàs fent un gra massa?» i així, pel nostre bé). No sé si tenir un conseller així, un *Beppetto* (el grill parlant que Collodi va plantar a l'orella d'en Pinotxo) insistent i molest, ens hauria salvat del daltabaix aquest de les bombolles. El que sí que sé és que ara sabríem perfectament de qui és la culpa. Com va dir Einstein quan li van demanar per l'infinitesa de l'univers: «l'únic infinit que jo conec és el de l'estúpidesa humana».





# Com hem de fer els robots?

*La mort de l'expert*

NODES  
NÚMERO 45  
HIVERN DE 2009

**N**icholas Carr (el de *Does IT Matter?*, de 2004) es demanava a l'*Atlantic Monthly* el 2008: «No deu ser que Google ens està convertint en idiotes?» I tot seguit a *Edge* **George Dyson** contestava, ampliant el camp de la pregunta: «I si fos que el preu que hem de pagar per tenir *màquines que pensen* és tenir *gent que no pensa*?» La contesta d'en Dyson (historiador de la ciència i fill de **Freeman Dyson**, el físic) no és només retòrica. Ja comencem a entreveure una situació semblant. **Ian Ayres**, al llibre *Super crunchers* (subtitulat *The death of expertise*, 2008), constata que, avui mateix, la informàtica està fent que cada cop més professionals ben pagats acabin com ha acabat l'expert que tenien els bancs per ensumar garanties a l'hora de concedir préstecs: convertit en un simple operari de centre-d'atenció-al-client, la feina del qual és llegir en veu alta allò que veu escrit a la pantalla.

Primer van ser els entesos a detectar petroli a partir de mapes de geosinclinals, i aviat seran, per ordre descendent de consideració i salaris, els metges, que fa temps que descobreixen que els ordinadors diagnostiquen millor, i els mestres, que veuen que ja no se'ls necessita perquè els alumnes es motivin, aprenguin i siguin objectivament jutjats. Ens estem tornant, doncs, simples introductors de dades perquè els autèntics decisors, les màquines, puguin tenir i practicar allò que fins ara teníem i practicàvem nosaltres en exclusiva, el sentit del judici racional? Ho deixo penjat a l'aire. Potser un altre dia. Avui, esperonat pels resultats d'experiments, articles i llibres recents, volia anar lligant uns quants caps, que han

anat quedant sense lligar en els darrers articles. La cosa anava de les diferències que hi ha entre nosaltres i els robots que concebem i construïm. Començo doncs una enumeració de nou qualitats per a la intel·ligència –Necessàries? Suficients?– per veure si nosaltres i els robots que fem ens diferenciarem (o no). Més que res per evitar que tinguem més sorpreses del compte. Però abans volia passar a veure un vell amic:

### *Dues qualitats prèvies?*

**Michael Gazzaniga**, a *Human: The science behind what makes us unique* (2008), feia una llarga llista dels molts trets «únicos» que ens separen dels animals –amb qui segons ell (només) compartim biologia i una certa experiència «conscient» de la realitat– i que ens defineixen com a **humans**: citava explícitament (1) el control dels pensaments, emocions i actes, (2) la planificació de cara al futur; (3) la «reflexió» i la «consciència d'un mateix», (4) el llenguatge, (5) la imitació i l'aprenentatge socials, (6) la memòria episòdica, (7) la imaginació, (8) la creativitat, (9) l'actitud cooperativa i l'altruisme, (10) tenir una «teoria de l'altre»,... i un inacabable etcètera. **Gazzaniga** no gosava destacar-ne cap com a «més únic» que els altres, però acabava insinuant que els més definitoris són aquests dos: (a) la capacitat de desitjar o imaginar que una cosa es pot fer millor, superant les restriccions imposades per la biologia i l'exterior (**Gazzaniga** diu que cap altra espècie no aspira a ser més del que és), i (b) la capacitat de sortir d'un mateix per veure's *des de fora* estant, i veure-ho tot des d'aquesta perspectiva arbitrària i abstracta. Cal recordar que **Gazzaniga**, deixeble del nobel **Sperry** (el descobridor de l'asimetria cerebral), és qui va imaginar el mecanisme dels hemisferis com un «intèrpret» resident a l'esquerra que analitza les accions, integra les aportacions dels diversos «mòduls» funcionals del cervell i genera un «relat» que, estructurat en forma d'imatges, històries o vés a saber què, que constitueix el «torrent» de la nostra experiència conscient (**Gazzaniga** ho explica en aquest llibre com ja ho havia explicat a *The integrated mind* o a *The social brain* del 1978 i del 1985, respectivament, i em sembla que jo ja n'havia parlat, del diàleg continu interhemisfèric). A partir d'aquest punt deixo la llista gazzanigiana i em faig la meua pròpia

(= apartats 1-9) que té posat l'accent en la comparació de diversos components de la intel·ligència humana amb els corresponents –si s'escau– d'un robot.

### 1 *La simpatia*

Comencem per aquí, com a lubricant social. O més aviat caldria dir l'**empatia**. Aquesta, sobre la qual ja m'he esplaiat altres vegades, porta a parlar de les *neurones mirall*, l'absència o defecte de les quals se sospita en *autistes* i es postula com a causa de la seva malaltia. Sempre les esmento i no n'acabo parlant mai. Avui tampoc. Prometo que algun dia. Mentrestant, vosaltres mateixos podeu anar fent: dos dels seus sorpresos descobridors (a primers dels 1990, a Parma) han escrit llibres on ho expliquen tot sobre les *mirrors neurons*, que són bàsiques i prèvies per construir la *teoria de l'altre*. Aquesta, com se sap, resulta imprescindible per conèixer i aprendre coses de la vida (que no siguin números o d'altres objectes que caracteritzen el món mental dels *idiots-savants*). **Giacomo Rizzolatti** n'ha escrit un de titulat *Mirrors in the brain* (2007) mentre que **Marco Iacobini** un altre (*Mirroring people: the science of how we connect to others*, 2008). I si no en teniu prou, podeu trobar comentaris molt favorables i il·lustratius sobre el tema del nostre vell conegut **Antônio Damásio** en un assaig breu («Behind the looking-glass») a *Nature*, 10/7/08 (p.167) seguit d'una resposta complementària (*Further reflections on how we interpret the actions of others*) de **Rizzolatti** a *Nature* 2/10/08 (p.589).

### 2 *Motius i potser plaer*

Un robot pot ser bàsicament reactiu, o bé pot tenir el que en diríem «iniciativa pròpia». En circumstàncies complexes, fins i tot quan actua reactivament, el seu comportament estarà influït per una base històrica de motivacions bàsiques, el que diríem el seu «caràcter». Quines són les motivacions d'un robot, i com «construir»-les o trobar-les? Recordo que en un congrés John Searle (el lingüista) va rebatre algú altre (un gurú de la IA «dura» que ara no recordo qui era) i deia que si bevem no és pas perquè l'acció «beure» estigui en una llista de coses a fer i un «programa»

l'executi, sinó senzillament perquè tenim set. I que «set» és una sensació promoguda per senyals químics i fisiològics que qualsevol metge entén. Això està en la línia de Gazzaniga i Damásio, que diuen que no hi ha intel·ligència si al darrere (o a sota?) no hi ha «emocions» o com en vulguem dir. Si no sentim no pensem, ni fem. Ara bé: com es tradueix això a l'ofici de fer màquines intel·ligents? És clar que si volem que un robot faci coses deu haver de tenir-ne *motius*. Però, què són, els motius? I on són? Els haurem de llistar i desar en una taula, com deia el guru de la IA programadora, perquè el robot els vagi consultant i així sàpiga què ha d'anar fent en cada moment? O li haurem d'aplicar una descàrrega electroquímica per veure si es mou (i precisament en la direcció que toca)? Vist com funcionem nosaltres, potser ho hauríem de plantejar d'una altra manera. És clar que hi ha coses que fem perquè ens burxen (mengem quan ens ronquen els budells, ens anem al llit quan ja no podem més de son). Però també hi ha multitud, potser majoria, d'accions que no obeeixen estímuls físics sinó que busquen obtenir *plaer*. El nostre còrtex prefrontal funciona precisament pel *principi del plaer*. Es fa una acció perquè fent-la s'obté plaer (o perquè se l'anticipa). Hem de fer robots que es moguin pel premi? Doncs haurem de vigilar que el robot no se'ns torni viciós o addicte, que això sol passar i, un cop passa, ningú sap dir on es pot arribar i on cal parar. I si hi ha addiccions, també hi haurà deshabitacions? o «temptacions»? (v. Greene i Paxton 2009) Els haurem de donar lliçons morals per resistir insinuacions? I podrem donar-los placebos, als robots, o caldrà metadona? I, simètricament, si senten plaer, sentiran també *dolor*? Aquest, com és sabut, pot venir d'un estímul extern, d'una *frustració* (= no haver aconseguit l'objectiu que es buscava) o per absència d'un plaer perseguit o anticipat. Se sap que el dolor –o més ben dit, la seva evitació– és un poderosíssim motivador. Així doncs, caldrà que trobem mecanismes que facin patir les nostres criatures (naturalment pel seu bé), perquè aprenguin les dures condicions de la vida? Davant d'un possible dolor o fracàs, s'arronsaran i tindran allò que diem *por*? I serà, això, *asimètric* respecte al plaer com ho és en nosaltres? Ho dic perquè el dolor es percep com una brutalitat primària (a l'amígdala), mentre que el plaer és com una rebuscada estratègia prefrontal (i això fa que ens sigui tan difícil

comparar present i futur i, doncs, *preveure* sàviament què cal fer). I qui parla de fer *aprendre* a base de cursos de dolor, parla de fer cometre *errors*, i propiciar-los, contradient-nos en l'objectiu: no era perquè els robots no s'equivoquessin que els fèiem? I doncs? Sí, ja sabem, des de Turing, que la intel·ligència permet evitar els errors i que, paradoxalment, sense error no hi ha intel·ligència, i també que quan s'aprèn millor és quan hi ha errors a superar. Però, no és tot això tan ardu i autorecursiu com educar un fill? I, parlant d'addiccions difícils, tindrem també robots *perfeccionistes*, d'aquests que no saben quan parar perquè allò que han de fer els quedi «perfecte»? (v. **Antony i Swinson**: *When perfect isn't good enough*, 2009). Com convèncer-los que parin, que ja està bé i que ja n'hi prou? Uf! Que difícil, tot!

### 3 *L'agradositat*

La *xafarderia* (el *gossiping*) és un dels mètodes més corrents i eficaços per assolir la cohesió grupal (v. Wert i Salovey 2008). Els humans som essencialment animals *socials*, i fem el que podem per conuiu, i si no, ho tenim fotut. Això ens duu a voler *agradar* i ser *estimats*, i sobretot evitar de totes totes el *rebuig* dels altres. L'evitació del rebuig, que explica molts comportaments, té una força enorme, determinant; tant, que va sorprendre els seus descobridors experimentals, **Risen i Gilovich**, de Cornell (el 2007). Com lliga aquesta por –si hi lliga– amb la predisposició a *perdonar* (i amb la necessitat de *ser perdonat*) encara està per estudiar, com també la relació d'això amb la reacció defensiva general de por. En canvi, el que tenim claríssim és que l'*altruisme* és un dels components més sòlids de la qualitat humana (v. *Generosity* de **Martin Nowak**, a *Nature*, 4/12/08, p.579). Ara bé, si el volem inculcar als nostres robots probablement no tindrem cap guia per desenvolupar-lo perquè en els humans l'altruisme s'ha creat en associació íntima amb l'activitat anomenada *guerra* (v. *Conflict: altruism's midwife* de **Samuel Bowles**, a *Nature*, 20/11/08, p.326). Podran els nostres robots, excepcionalment, ser generosos sense haver passat cap fase agressiva?

#### 4 *La voluntat*

**Daniel Wegner**, de Harvard (que també cito més avall), s'ha passat la dècada estudiant experimentalment on és exactament el límit entre la decisió i l'acció, per veure on comença i on acaba la *llibertat* d'acció. Diu que abans de l'actuació voluntària hi ha una *intenció* conscient, que el subjecte viu amb la sensació de ser *ell* –i no una altra persona– el *causant* de l'acció, i que immediatament després de la intenció ve la microplanificació detallada de l'acte motor i la seva execució. Independentment, en el terreny de la manipulació experimental del còrtex (en aquest cas el parietal posterior), el francès **Michel Desmurget** ha descrit (a *Science*, el 2009) com es pot crear, per estimulació *externa*, la *voluntat* o *intenció* –que el subjecte sent com a generada «internament»– de moure una part del cos (sense moure-la, però); tot seguit, amb una estimulació més directa, aplicada en un altre sector del còrtex i amb un corrent més fort, també provoca la sensació d'intenció, però fa que, a més, la part del cos es mogui. I mentrestant **Blanke i Aspell** (a *Nature*, 9/4/09, p.703) expliquen que han descobert experimentalment que la sensació de «tenir un cos» i la consciència de «ser un mateix» sembla lligada a les zones frontal i tempoparietal del cervell; els autors acaben preguntant-se: «si substituïm aquests fragments per pròtesis o manipulem *des de fora* els senyals que emeten, de qui es pot dir que és, la responsabilitat dels actes que se'n segueixen?» Una bona pregunta, que, com les que vénen tot seguit, pertanyen al camp ètic.

#### 5 *L'ètica*

Ja vam parlar fa dos articles (v. Nodes, 43, p.6) de les preguntes que es feia Noel Sharkey a *Science* 19/12/08 (*The ethical frontiers of robotics*, p.1800) sobre els (nous) *problemes ètics* que planteja tenir robots. Aquí només vull postillar que, a l'espai de correspondència de la revista, Goodman i Einspruch responien (a) que els problemes que plantejava Sharkey no eren exclusivament deguts a la presència de robots (per exemple, la qualitat assistencial o magisterial d'un robot en el camp de la medicina o l'educació, respectivament, és part integral de la qualitat del sistema assistencial o educatiu, no

pas res autònom), (b) que Sharkey (i més o menys tothom) els plantejava sempre a la vista del «cas pèssim» (com ara, això ho dic jo, si per explicar com funcionarà un sistema ferroviari es digués que hi podrà haver accidents, i ens hi esplaiéssim), i (c) que si estem tan alerta sobre possibles perills deu ser perquè els motors primaris de la recerca en robots ha estat sempre l'exèrcit americà i les seves necessitats militars, i això ens ha posat en guàrdia. Potser caldria afegir en aquest apartat ètic la contribució, que intenta ser àmplia i sistemàtica, de Michael Gazzaniga (una altre cop, encara), que, al seu llibre *The ethical brain*(2005) (1) s'escandalitza per allò que en diu «neuromarketing» (vol dir l'apropiació dels escàners cerebrals pels publicitaris, àvids de saber quin estímul excita més el consumidor), (2) prova de definir límits (per exemple preguntant-se si –o quan– un malalt d'Alzheimer deixa de [tenir la consciència de] «ser ell», i perd la seva «individualitat»), (3) es planteja la relació entre la «Neuroètica» i el Dret, avisant que, segons la seva opinió, el cervell és un mecanisme bàsicament *determinista* que obeeix *regles* i que es refia molt de la *memòria* (que s'ha vist experimentalment que és molt enganyosa i *infiabile*), i (4) comenta –de passada– que s'ha observat, en experiments recents, que una epilèpsia al lòbul temporal produeix inexplicablement un augment sobtat de la religiositat en el subjecte (i aquí ho deixa). Potser caldria afegir encara a aquest apartat l'editorial *Robot ethics* de Robert Sawyer (a *Science*, 16/1107, p. 1037), en el qual recorda que ja al 1979 un jutge va concedir una compensació de deu milions de dòlars als parents de la primera víctima d'un accident causat (involuntàriament) per un robot, o el llibre de Wallach i Allen *Moral machines: Teaching robots right from wrong* (2008) els autors del qual creuen que els robots (incloent-hi els minaires de dades o els *bots* d'Internet) exigeixen una ètica nova que hauria de estar subordinada, si més no per ara –fins al moment improbable, diuen, que ens superin–, a les necessitats dels usuaris humans, de manera semblant a l'ètica que gastem amb els nostres animals domèstics.

## 6 «Pensar»

Si fem cas de la hipòtesi de **Jeff Hawkins** (a *On intelligence*, 2005), tan sovint citada en aquesta columna, que sosté que la intel·ligència



és bàsicament un procés de *detecció de diferències* (tipus «ui, això ahir no hi era!») que és immediatament seguit d'una activitat *lògicodeductiva* («què deu haver passat?») i que no para fins a obtenir una «explicació» prou satisfactent, i si la posem en pràctica, caldrà doncs, com em sembla que ja vaig dir una vegada, muntar en els nostres robots un reconeixedor de formes ben sensible i acoblar-hi un potent *deductor* lògic, possiblement augmentat amb valoracions i càlculs probabilístics. En aquest cas, tenint en compte que les conclusions seran insegures, caldrà preveure un mecanisme d'acceptació i valoració del risc (i admetre que «risc» no significa fracàs ni frustració ni ha de portar a la paràlisi sinó al pla B i l'acció correctora). Un problema que veig en el plantejament hawkinsià és que el «detector de novetats» és fàcilment esgarriable pel nostre inveterat *biaix de confirmació* (sobre el qual em vaig esplaïar a bastament al darrer número). Caldrà doncs filar molt prim i evitar que les ganes o necessitat de confirmar creences que nosaltres tenim no s'encomani als robots. Altres factors de disseny que s'hauran de considerar són: (1) el *context*: se sap que cap cosa no té significat ni operativitat si no és en un context apropiat, i que la intel·ligència s'ha de situar, i ha de col·locar cada element al seu lloc; (2) el nostre cervell contínuament emet, manipula, invalida i transforma *hipòtesis* (una demostració experimental senzilla, sorprenent i recent d'això, és la següent: tota predicció feta per un individu guanya precisió –el predictor l'encerta més– cada cop que se li demana que hi torni; i és que, com es comprova, el seu cervell no ha parat d'anar modificant posicions dinàmicament, (v. Vul i Paschler 2009); (3) com fa temps que suggereixen **Gazzaniga** o **Schank**, els nostres processos mentals d'identificació, abstracció, causalitat i predicció s'estructuren de manera natural en forma d'*històries* que expliquem, la qual cosa ens permet anar construint el nostre pensament alhora que ens fa sociables; i (4) no sols detectem detalls canviats sinó *microgestos* o «senyals» corporals o gestuals dels nostres congèneres que ens orienten molt més que les paraules (v. *Nature*, 29/1/09, p.528). Tot això porta a qüestionar-se: la intel·ligència dels robots haurà d'estar basada també en un mecanisme de Hawkins? i hauran els robots d'anar donant voltes a una hipòtesi, presentar-s'ho en forma d'història (i explicar-la a algú?) i estar atents als petits gestos? I, com a corollari, no vol

pas dir tot això que la seva intel·ligència implícitament serà, com la nostra, *social*?

### 7 Causes i efectes

Com aprendran o descobriran els robots que certs efectes se segueixen de certes causes i que això permet predir el que és raonable trobar? De *causalitats*, n'hi ha moltes: (1) Hi ha la *superstició*, que també practiquen els animals: s'ha vist que una cosa passava després d'una altra, i tot seguit es creu que repetint la primera passarà la segona (és la fallàcia clàssica *post hoc ergo propter hoc*): hem pogut veure animals patèticament repetint *ad infinitum* una mateixa maniobra palesament inútil, i tot perquè un dia van veure dues coses juntes o en successió. (2) Més endavant apareix històricament la primera *sobrenaturalitat*, la creença en «esperits» (atenent a la pressuposició segons la qual tot el que passa és perquè és voluntat d'algú, més la suposició addicional que sempre hi pot haver «algú que mira»). (3) Més tard ve la *religió*: com s'ha pogut comprovar fa poc en condicions experimentals, quan el *sofriment* (i, en grau menor, la sensació d'aclaparament, de temor o de desesperança) domina, els «esperits» s'unifiquen en un o pocs déus omnipotents, de la voluntat dels quals es passa a dependre incondicionalment (v. Gray i Wegner 2009). (4) De més tardana aparició encara, i més difícil d'obtenir, és la causalitat àrduament i col·lectivament construïda, rigorosa i basada estrictament en el *principi de realitat*, anomenada *ciència*, amb estructures mentals prèvies a aprendre i consolidar, i amb mètodes propis i disciplinats i amb respostes provisionals i permanentment qüestionades. Doncs: la pregunta a fer-se és: què trobarem en els nostres robots? es tornaran supersticiosos? faran l'esforç de creure en alguna cosa invisible l'existència de la qual inferiran dels dos pressupòsits que he dit més amunt? O bé: tindran «esperit científic»? de qui l'heretaran? com els el transmetrem? Caldrà explicar-los bé els conceptes i la metodologia perquè l'apliquin correctament? com? I si mai se'n desvien, com fem nosaltres sovint, caldrà considerar-los «irracionals»? (i evitar que decideixin res important?). Àrdues qüestions, aquestes, però no pas gaire diferents ni gaire més difícils que les que hem hagut d'afrontar sempre amb els nostres fills o alumnes o amb tothom

en general, amb qui convivim més bé que malament.

## 8 Consciència

Aquesta característica sembla bàsica en qualsevol actuació intel·ligent. Ho veiem quan falla (quan «no hi ha *consciència*»): quan «anem zombis» (i llavors: se'ns pot atribuir la mateixa responsabilitat en aquestes condicions que en les normals?) o quan «la perdem» (temporalment, per un «xoc», o gradualment, per un Alzheimer): té sentit dir que continuem sent els «mateixos»? Per començar, potser fóra bo de reduir el problema al nivell més elemental possible: en efecte, hi ha un tipus de «*consciència (d'un mateix)*» bàsica que és el mínim que es necessita per saber on acaba un i on comença l'univers exterior. Si no es té, no es pot ni tan sols moure el cos propi: com saben alguns malalts paralitzats, es giren de costat al llit i cauen. I tampoc, sense aquesta «consciència» elemental, no es pot distingir entre el que et passa a tu i a l'altre (és el cas dels *autistes*), entre els propis sentits i l'empatia, o entre la realitat i l'imaginari (el cas dels *esquizofrènics*). Ara bé: què és «la consciència», i com «construir-la»? En això som al cap del carrer: no en tenim la més remota idea. Només ens queda esperar que aquesta característica sigui un *Deus ex machina*, una d'aquestes qualitats *emergents* que confiem que sorgeixin, una mica com de miracle, de la *complexitat*. I cal recordar que si no hi ha «consciència» pot haver-hi «responsabilitat»? Greus consideracions a fer. I cal que ens hi anem afanyant, perquè ja han començat els problemes. N'esmento un, com a mostra: fa poc un jurista es plantejava (a *Science*, 22/5/09, p.1014) si hi podia haver cap judici sobre propietat intel·lectual en què es disputés qui és l'«inventor» d'una idea sorgida en una execució de programa per un robot: és l'autor del programa (o del robot)? o és el robot mateix? Si és el darrer, el robot passa implícitament a ser considerat jurídicament «persona», i se li deuen drets d'autor (que –de moment?– cobrarà la institució propietària); l'autor considerava que tot passava pel tràmit tècnic de demostrar que l'autor, humà o no, era «capaç d'adonar-se» del que havia fet (= que posseïa «*cognition*»), i afegia que sobre això el Dret europeu i l'americà discrepaven. La de trasbalsos jurídics i ètics que ens esperen!

## 9 *Collateralitats diverses*

I, parlant de tot, tindran *estrès*, els nostres robots? Vist que, en humans, el *cansament cognitiu* i, molt més encara, estar estressat, abaixa moltíssim el rendiment i augmenta exponencialment l'eventualitat d'errors, és evident que caldrà fer els robots resistents a tota mena de cansament i de tensió emocional. I, encara, una pregunta potser inesperada: es deprimiran, els robots?... Ho dic perquè una teoria de moda (d'Andrews i Thomson, 2009) diu que la *depressió* és un mecanisme evolutiu que permet concentrar-se en un problema i, des de la reclusió i l'*entotsolament*, capgirar-lo i superar-lo. Si és així –si bé jo, francament, no ho crec– ja em veig una cua de robots demanant, com nosaltres, la baixa mèdica.

Vaja, que si no parem compte haurem fet els robots perquè siguin «màquines» i, amb l'excusa d'avesar-los a la dura vida, potser els trobarem «patint» com a criatures (v. apartat 2); els haurem fet perquè siguin fiables i no cometin errors i, amb l'excusa d'aprendre, potser en cometran més que nosaltres (us imagineu que ens fiem d'un robot i ens atropella? no li perdonaríem de cap manera, encara que ens demanés excuses sinceres o se'n sentís horriblement penedit); i els haurem fet perquè siguin forts i potser seran tan fràgils i trencadissos com ho som nosaltres mateixos (v. l'apartat 9). Això seria com haver anat, amb grans esforços, del no-res a la misèria, com deia, tan lúcida com sempre, Marx –variant Groucho.



# Calaix de sastre, en tres apartats

## Apartat 1: Requiescat in pace

NODES  
NÚMERO 46  
PRIMAVERA DE 2010

**K**im Peek es va morir el desembre passat. No, no era Dustin Hoffman ni *Rain Man*, tot i que en va ser l'origen. Un dia de 1984, Barry Morrow, un guionista de Hollywood, el va conèixer per casualitat i el seu cas li va interessar. No, Kim Peek (1951-2009) era un aparent autista de la modalitat *idiot savant* capaç de recuperar a velocitat de Google i recitar de memòria, sencer i sense oblidar cap coma, *La caiguda de l'imperi romà*, de Gibbon, *La persecució de l'Octubre Roig*, de Tom Clancy, i 9.000 llibres més, dels quals no sols recordava la pàgina de cada cosa sinó també el context i els detalls que es desprenien d'un ítem donat. La seva quinzena d'interessos anava de la història a la geografia, del cinema als esports (incloent-hi vida i miracles de totes les seves estrelles), o de la Bíblia a la història de l'Església passant per la literatura en general i per Shakespeare en particular, i per la música clàssica. Se sabia tots els codis postals i emissores americanes, era capaç d'explicar-te, via instruccions tipus Yahoo, com anar a qualsevol lloc que li diguessis i d'identificar amb no gaires notes centenars de composicions clàssiques, afegint-hi informacions sobre l'estrena, la vida del compositor i les circumstàncies polítiques. Fins aquí, no gran cosa. D'*idiots savants* en teníem per triar i remenar.

El que fa molt interessant Peek és que, tot i no ser gaire «intelligent» (tenia un QI de 87) ni gaire hàbil físicament –no se sabia cordar el botons i caminava de tort–, just entrant a la maduresa va saber adquirir habilitats «normals», generalment negades als *idiots savants*, com ara tocar el piano, que feia servir per il·lustrar

els seus coneixements musicals, i, encara més rara en *savants*, la d'arribar a *entendre* els milers d'acudits que recordava sense entendre'ls, i també copsar el sentit idiomàtic o irònic d'algunes frases que repetia sense haver-ne captat mai abans el sentit no-literal, l'accessible a la gent normal. Això el fa únic, perquè és com si tinguéssim *in vitro* una màquina humana dotada d'una memòria associativa prodigiosa i poguéssim fer-hi experiments i fer-li preguntes. El que sí que hem fet és mirar-li el cervell del dret i del revés. I ens hi hem trobat una sorpresa: Peek no tenia *cos callós*. La broma habitual dels neuròlegs quan els demanen què coi és això és respondre que no serveix per a res, només per enganxar les dues meitats del cervell. Doncs, ara seriosament, resulta que no. Jo ja n'he parlat més d'una vegada, del *corpus callosum*, que és el que es talla als epilèptics per evitar la propagació d'atacs d'un hemisferi a l'altre, una experiència repetida i estudiada –per Sperry i Gazzaniga– que ens ha fet veure com són de funcionalment diferents els dos hemisferis cerebrals (d'això jo també n'havia parlat, aquí). I doncs? Quina importància i significat té, si és que en té, el fet que Peek no disposés del seu cos callós? Que els seus dos hemisferis no es comuniquessin, li era un favor o un impediment?

Un altre descobriment que ens ha llegat Peek, que no sabem quina relació té amb l'anterior, és aquest: el seu hemisferi esquerre (el del llenguatge i el coneixement analític) presentava força anomalies. Això, de moment, no és una sorpresa tan gran com pugui semblar. Primera, perquè en pacients «normals» madurs que han tingut accidents a l'esquerra se'ls desenvolupa espontàniament una conducta transparentment batejada de «síndrome de *savant* adquirida». Segona, perquè se sap que la fortíssima descàrrega de testosterona que reben els fetus mascles resulta molt agressiva a l'hemisferi dominant (l'esquerre), al qual costa molt de refer-se'n. Potser no cal que recordi aquí que els mascles presenten una freqüència i proximitat a l'autisme i savantisme molt superior a les femelles... si bé cal notar que no sempre les dues coses van juntes: només la meitat dels *savants* són autistes, i en particular Peek no n'era.

Les explicacions del fenomen que acabo de dir es divideixen en dues. L'una, que l'hemisferi esquerre s'intenta recuperar del

cop testosterònic sollicitant, i obtenint, l'ajuda del dret, que coopta part de les seves capacitats per omplir el buit i simular les aptituds mancants i necessàries, bàsicament analítiques, de l'esquerre. La dos, que l'hemisferi dret, quan es veu alliberat de la «tirania de l'esquerre», té barra lliure per fer allò que normalment es deliria per fer però que l'esquerre no li ho permet. Tant en una explicació com en l'altra, com deueu haver notat, un cos callós tallat o inexistent és plausible que interfereixi (com?) en els mecanismes descrits, si són aquests, de «curació» del xoc fetal testosterònic. Ara bé, que tot aquest embolic ens il·lumini o no a l'hora d'entendre i concebre intel·ligències artificials és una qüestió que deixo enlaire. (com a referència global sobre aquest cas us recomano aquesta: Darold Treffert i Daniel Christensen, *Inside the Mind of a Savant*, a l'*Scientific American* del 23/12/2009, publicat quatre dies després de la mort d'en Peek, tot i que al Google trobareu no menys de 1,7M referències concretes de «Kim Peek»).

### *Apartat 2: Borges, el memoriós*

A la secció de llibres del *Nature* del 4/2/10 hi ha una curiosa ressenya feta pel bioenginyer argentí Rodrigo Quian Quiroga establert a Leicester que, després d'inquirir sobre Jorge Luis Borges (1899-1986) a María Kodama (la vídua), ens descobreix un secret que potser explica no només el personatge sinó també alguna clau de la memòria, una mica en la línia del que deia jo a l'apartat 1. Se sap que Borges va escriure el 1944 un curiós llibre titulat *Funes el memoriós* sobre un (suposat) Ireneo Funes que, havent caigut d'un cavall i havent-s'hi fet mal al cap, era capaç de recordar-ho absolutament tot, per exemple «les formes dels núvols del matí del 30 d'abril del 1882, i comparar-les amb les irisacions de la tapa d'un llibre que un dia va veure». Resulta que a Borges li va dir un metge de petit que tenia una malaltia degenerativa i que es quedaria cec. Doncs el *nano* va començar una cursa contrarellotge i, com que tenia bona memòria, va aconseguir, com era capaç de fer de gran, recitar pàgines senceres de textos en espanyol, anglès, alemany i altres llengües (entre les quals el català) –i, de vell, ja cec del tot, encara era capaç de dir la pàgina exacta on calia anar a trobar una cosa. Ara sabem, gràcies a un pacient «H.M.» analitzat



els anys 1950, que la cosa és responsabilitat directa de l'hipocamp, que està al mig del pas dels circuits sensorials del cervell. Això Borges no ho sabia, però buscant informació útil al seu cas va trobar un llibre del 1902 d'un psicòleg alemany (G. Spiller) que va deixar anotat profusament en la seva lletra jove i que el va ajudar a fer càlculs: al ritme que s'hi hipotetitzava, als seus trenta-cinc anys hauria acumulat uns 10 K records. Això li va semblar acceptable, si bé el va posar en un disparador: calia afanyar-se per no parar cec abans. Irònicament, Borges no es va quedar cec fins al 1955, el mateix any que el van nomenar director de la biblioteca nacional argentina, un cop de gràcia que segur que el sentit irònic de l'autor devia apreciar.

Borges va descriure l'hipotètic Funes amb una precisió analítica –semiautobiogràfica de fet, com es veu ara– sense saber ni sospitar que al mateix temps a l'altre cap de món, a la URSS, el neuròleg Alexander Lúria estava estudiant el «pacient» S.V. Xereixévski, un periodista de memòria aparentment il·limitada, que era capaç de recordar-ho tot –absolutament tot– i alhora de no entendre res, i menys encara de comprendre el sentit dels milers d'acudits que sabia recitar. No cal dir que en això, tornant a l'apartat 1, va tenir molta menys fortuna que el nostre Kim (Peek) que a la seva mitja vida, en un cop de sort, va «aprendre» a entendre. I moltíssima menys, és clar, que l'irònic Borges, tan «memoriós» com intel·ligent.

Segur que en aquestes històries que he explicat (apartats 1 i 2) hi ha alguna cosa a aprendre, si bé no aconsegueixo saber què. (Sobre Peek ja us he dit on anar a trobar-lo. Sobre l'inclit Borges, vegeu *Nature* 463, 611, article *In Retrospect: Funes the Memoriosus*)

### *Apartat 3: Noticiari*

(De cada notícia se'n pot extreure, si es vol, una moral i una aplicació a la vida ordinària, i també a les nostres màquines intel·ligents... però això últim que s'ho faci cadascú.)

**Notícia 1:** En un experiment controlat (*Psychological Science*, gen. 2010), es comprova que el fet de tenir «poder» fa que els subjectes d'experimentació als quals se'ls ha «conferit» tendeixin a comportar-se, automàticament i marcadament, de manera que

diríem *hipòcrita*. De seguida tendeixen a emetre explicacions autojustificatòries (de les decisions pròpies) que inclouen la falsedat, que ells copen i accepten de manera espontània i natural i que creuen que els és permesa precisament en virtut de l'autoritat que els ha estat atorgada. Una lliçó política per a nosaltres, potser? Observeu que els subjectes són estudiants presos a l'atzar i que no s'ha pas de suposar que tinguin necessàriament vocació política o de poder. *Sempre han tingut bec les oques*, que deien els nostres avis per explicar que tothom està fet del mateix material i que *il y a des choses qui ne changent pas*. Ah!, hi ha un bri d'esperança, però: una minoria de subjectes reaccionaven al fet de «tenir poder» no aprofitant-lo per sentir-se autoritzats a dir mentides sinó que, tot al contrari, el sentit de privilegi els feia molt més estrictes, no més laxes, amb ells mateixos. (Sí, eren minoria.)

**Notícia 2:** Es descobreix (*Mind*, gen. 2010) que el soroll de fons interfereix, i molt, l'activitat del nostre còrtex prefrontal (CPF), que és, com se sap, on es ponderen els factors socials i humans de cada decisió –sí, els ètics també–, on es planifica una conducta i on es prenen les decisions. Si combinem la degradació de funcionament del CPF pel soroll amb les importants conseqüències, també descobertes fa poc –en el camp financer, amb la crisi–, de l'anomenat «cansament cognitiu» sobre la bondat/ oportunitat de les decisions preses, ja es veu que no podem ser gaire optimistes. Entre fresses i fatigues, estem aconseguint que la qualitat de les nostres decisions convergeixi a zero. (I això suposant que la informació de la qual es parteix sigui bona o contrastada, cosa que avui, en termes d'Internet i d'informació volant i refregida, no és el cas gaire sovint.)

**Notícia 3:** El nostre cervell consumeix gairebé la meitat de l'oxigen que respirem. Doncs bé, se sap des fa temps que el cervell és bàsicament un òrgan de regulació –en sentit tècnic– de les funcions corporals. I que, en el 10-20% de temps que queda en *standby*, pot aprofitar –que no vol dir que aprofiti– el coll d'activitat que li deixa el cos per «pensar». (De fet, fisiològicament, «pensar» és per definició això: estar en *standby* del cos). Ara bé: En què pensem quan no pensem en res? Doncs es veu que anem fent

proves de funcionament de circuits neuronals (segons Ming Song, de l'Acadèmia xinesa). «Provem» associacions antigues i si cal restaurar-les hi reforcem sinapsis, i si no ho deixem córrer i les acabem de matar. A més, això sembla bo, en principi, perquè les associacions que reforcem són les d'abast ample i, així, afavorim processos de memòria associativa i analògics. La conclusió? Doncs que no pensar en res vol dir «fer manteniment», tot i que hi ha gent, es diria, que o no té res a mantenir o bé, com que no pensa mai, no fa altra cosa. Aquí ve bé de recordar allò que deia Bertrand Russell: «N'hi ha que no pensen, ni que els matin... i de fet, abans de pensar, es moren» (Cito de memòria: *Some people will die before thinking... in fact, they do*).

**Notícia 4:** No us fieu mai dels testimonis! Després d'innombrables experiments ha quedat clar que el que recordem pot no tenir cap relació amb el que hem vist, per molt que a nosaltres ens ho sembli. Ep!, que la cosa és greu. Ja s'havia pogut comprovar que «abusos» recordats al detall no havien pogut passar mai materialment. Ara, a base d'analitzar milers de casos i els processos de memòria subjacents, Elizabeth Loftus, d'Irvine (UC), ha arribat a la conclusió que, contra l'explicació intuïtiva i «fotogràfica» que ens en fem, un record NO és pas la reproducció d'una imatge sinó, realment, tota una reconstrucció d'una imatge –nova, doncs–, que pot tenir o no relació amb la «recordada». El procés de reconstrucció (o d'elaboració) és força complex i hi intervenen coses com ara el context, o paraules sentides en el moment. Quan dic context o moment vull dir els de la reconstrucció, no pas els de la imatge original... si és que mai va existir, perquè pot ser que allò recordat no hagi passat mai o que fins i tot sigui físicament impossible. Se'ls ha girat feina, doncs, als jutges, que potser ara hauran de consultar més els neuròlegs i no fer tant de cas dels psicoanalistes.

**Notícia 5:** L'adulació no compra voluntats, us diran, sobretot si es nota. Doncs no. Experiments recents (fets a Hong Kong i publicats al *J. Marketing Research*) han demostrat que quan ens diuen coses agradables, encara que siguin clarament falses o manipuladores, alguna cosa de dintre nostre es remou, i no reaccionem pas de la mateixa manera amb la persona que ens ha adulat com

ho fariem amb ella si no ho hagués fet. És a dir, ho necessitem, i ho agraïm, encara que dissimulem o diguem missa. Conclusió (per als robots que ens hagin de tractar, de vells): Aduleu, encara que es vegi que no us ho creieu. Perquè l'altre ho espera, com en una mena de ritual. La prova és que si ens trobem algú i esperem que ens faci alguna manyaga verbal i no ens la fa, la resposta nostra serà: «ostres, que fred i que desconsiderat que és!». No, no reconeixem pas, mai de la vida, que esperàvem un elogi, mal que fos maldestre o fals, i que, no havent-lo obtingut, hem rebaixat dos o tres esglaons la categoria moral del nostre interlocutor. Atenció, robots, prepareu el sabó.

*Notícia 6:* Això els catalans ja ho sabíem, però sempre fa goig que la ciència ens ho verifiqui (en aquest cas, a la universitat belga de Gant). Resulta que els parlants bilingües són, pel sol fet de ser-ho, més intel·ligents que els unilingües –en el sentit corrent de la paraula «intel·ligència», sigui entesa com a QI o com a «factor g». En efecte, el bilingüe es veu obligat a comparar contínuament una paraula, i la seva fonètica i morfologia i el seu camp semàntic, amb la seva equivalent en l'altra llengua, encara que en aquell moment només estigui parlant en un idioma. La sola existència d'una alternativa dispara els mecanismes, coneguts i necessaris, de comparació/traducció/contextualització que són típics de la intel·ligència general, i així se'ls prepara i potencia per a altres ocasions en què se'ls necessiti.

*Notícia 7:* «Estar enamorat» és una oració que sembla demanar un objecte, oi? Doncs no. Sembla que hi ha gent per a la qual «estar enamorat/da» vol dir només «estar enamorat». No, de ningú. La persona –diguem-ne– «estimada» només n'és l'excusa o ocasió. L'objectiu és estar en aquell estat d'inòpia que en diem enamorament. Amb això ja n'hi ha prou. Ho va constatar *The Economist* (*Real life is not a romantic comedy*, 9/1/10) quan va entrevistar un company de feina del ruc que va paraitzar durant sis hores tot un aeroport americà perquè es va saltar un control per abraçar la seva «estimada»: el paio –un postdoc de Rutgers de 28 anys– anava tan revolucionat no pas perquè la noia se l'estimés sinó perquè –aah!– «estava enamorat». Com si tot fos un decorat de Hollywood

i s'hi filmés una comèdia romàntica. L'expressió dels milers de passatgers parats a Newark, tot i que potser ho haurien entès, no va ser gaire romàntica, però.

**Notícia 8:** Hi ha finalment comprovació experimental (D. Eagleman, del Baylor College) del fet que, en casos d'accident, el temps sembla que va més lent, o que res no acaba de passar mai. Sembla que la desacceleració és deguda a les hormones lligades a la sensació de por i de perill. Així doncs, si aconseguim de regular-les potser haurem pogut controlar la dimensió de temps (psicològic). Com si tinguéssim un alentidor de càmera, regulable.

**Notícia 9:** Un estudi (comentat per l'economista Piers Steel, de Calgary) sobre la *procrastinació* (del llatí *procrastinare* =endarrerir, deixar-ho per cras =demà), és a dir, l'hàbit d'anar-ho deixant tot per més tard, que coincideix amb el «*mañana*» *syndrome* atribuït als sud-americans. Doncs resulta que, tot i els inconvenients d'aquesta malèfica i contraproduent estratègia –no, d'avantatges no se n'hi veu cap–, tothom –repeteix, tothom– ho fa de tant en tant, i el 15-20% d'adults nord-americans hi són addictes. Percentatge que puja al 95% dels (immadurs) infants i al 80-95% del (madurs) estudiants universitaris. Atenció, robots, no hi caiguen també vosaltres perquè si no, qui coi farà la feina?

Aquesta darrera pregunta (la meua de la Notícia 9) és rellevant. Perquè, no és per això que fem les màquines? perquè facin la feina? Tornem a Russell. Quan un dia li van preguntar per què estudiava la Lògica i després d'aventurar una possible resposta («potser ho fa per saber com pensem?»), ell va contestar, en la línia de la frase citada més amunt (Notícia 3) i sorneguer com sempre: «No, senyora, ho faig per no haver de pensar més». Va ser un dels moments fundacionals de la IA. És per això que no solament fem màquines –perquè facin la feina– sinó que volem que siguin intel·ligents –perquè la facin bé i ens hi puguem entendre... i nosaltres puguem dormir a la palla, que és un dels grans signes de la civilització.

# *Persones i màquines, i quin futur ens espera*

*Què és una màquina?*

NODES  
NÚMERO 47  
TARDOR DE 2010

**D**E jubilat, em faig molt amb grups de vells cultes i avorrits. De vegades, quan saben que havia estat ficat en el negoci aquest de la Intel·ligència Artificial, em programen, obsequiosos, una conferència perquè els expliqui què coi és això. Presumiblement per saber-ho abans que es morin. A l'Ateneu Barcelonès me'n van programar una ara fa uns mesos que va tenir molt d'èxit. Hi va haver molta gent de la part alta de la piràmide –la forquilla d'edats anava de 60 a 100 anys– i m'hi van fer moltes preguntes. No sé pas si van entendre gran cosa, però se'ls veia entusiasmats, i això em va fer pensar que no, que no hi devien entendre gaire res. Més endavant, ja més serens, m'ho van concretar, i per escrit. Els tertulians em deien que havien quedat «alarmats i entusiasmats», curiosa i contradictòria expressió d'un ambigu estat d'ànim. Si és realment aquesta la percepció que els va quedar en net de la conferència, es pot dir que va ser un èxit, perquè és la mateixa que tenim de la IA els mateixos professionals: la cosa també ens entusiasma però alhora, per molt d'optimisme que hi posem, no les hi tenim totes. És un camp, per dir-ho amb el tòpic suat, ple de d'oportunitats i de paranys i, com en aplicacions passades de la ciència, la bala ens pot sortir pel carregador. A la meva xerrada jo, temerari, em vaig aventurar pels viaranyes de l' «ètica» que tindran –perquè n'hauran de tenir!– les nostres màquines. Ah, coi! No hi havia pensat ningú. Els meus propectes col·legues havien estat educats

en l'ètica tomista en col·legis de capellans, i això els feia costeruda la comprensió de conceptes com ara que una màquina pogués mai tenir responsabilitat pels seus actes i no pas el seu «amo», o qui l'hagués «programat». I és aquí on em vaig trobar un obstacle bàsic. És fortíssima i universal la convicció que una màquina, per definició, no farà mai altra cosa que el que se li hagi manat. Quan em vaig adonar que aquest era el nucli dur de la qüestió, vaig diversificar les meves explicacions en diversos fronts:

1. Això de «programar» màquines és només la primera fase, rudimentària, de la informàtica. Avui tenim màquines que són tan complexes –i el seu recorregut d'accions tan enorme– que encara que les hàgim programat de manera determinista no hi ha mètode possible de saber quin tipus d'accions es veurà obligada a fer, tot i seguir literalment les nostres directrius. Però és que pot passar que la màquina decideixi trencar una rutina poc productiva i aleatòriament en determini una altra. O que aprengui de l'experiència, i millori el seu comportament per reeixir. I, encara més, pot passar que executi programes, els avaluï i els descarti o els millori, en un paral·lelisme estricte amb el mecanisme natural de l'evolució –d'això no fa gaire en dèiem *algorismes genètics*. O que funcioni no pas per un «programa» sinó per una estratègia d'analogies, o programació «associativa» –d'això fa poc en dèiem «xarxes neuronals–, o per un mecanisme de complexitat caòtica controlada (allò d'operar a la *vora del caos*) o vés a saber per quin altre mecanisme més. No sols les nostres màquines poden fer tot allò que se'ls ha dit –o rebotar-s'hi si ho veuen convenient– sinó que poden fer allò que *no* se'ls ha dit. Fins i tot poden fer coses que no s'han previst i que ni tan sols s'han pogut *imaginar*. La combinatòria i la complexitat ja fan aquests miracles, i si a sobre la màquina aprèn o experimenta, o aprofita l'atzar al seu favor, llavors ja no hi ha qui ho pari. En tot cas, la predicció i control del comportament de la màquina depassa el seu «creador». I aleshores, a qui reclamar si hi ha cap accident? Els meus senils col·legues no sabien què contestar. I això que els citava casos legals reals en què el perjudicat havia reclamat contra la màquina com a autor responsablement autònom dels seus actes, i fins i tot els jutges havien hagut de

reconèixer que sí, que si hi havia cap causant del sinistre havia de ser la màquina, perquè *decidia* i perquè tenia *autonomia* d'acció, i que ni a l'«amo» ni al «programador» se'ls podia exigir, en justícia, que haguessin de compensar ni pagar res, perquè la decisió presa era, clarament, de la *màquina* i no pas d'ells, que ni tan sols podien haver previst què podia passar. Una mica com passa amb els fills, vaig afegir: oi que no responsabilitzem els pares pel que hagin pogut fer? doncs igual! També els vaig parlar dels robots militars, fets per localitzar i destruir objectius i alhora dotats de certs frens «ètics»: per exemple, no causar dany innecessari, ni a innocents. Això dels militars els va horroritzar. En total, no van quedar gaire convençuts. I és que el punt crucial –potser inacceptable segons per a qui– és que les màquines són, com nosaltres, entitats que poden *decidir*, i per tant ser objecte de *qüestionament* pel que hagin decidit i fixeuvos que no els vaig dir que siguin «responsables» –no vaig gosar fer servir la paraula– ni, encara menys, que hagin tingut «llibertat» a l'hora de decidir o que tinguin «consciència» del que han fet, perquè aquestes paraules són tabú i hauríem entrat en un bucle. Ai, on hem anat a parar! Ui, l'ètica! L'efecte de la inèrcia, i de la Filosofia estudiada al batxillerat.

2. Com a contrast i per donar un contrapunt a l'apartat anterior: No sols és concebible avui, com deia, un ésser *no-humà* amb l'atribut de l'*autonomia*, i de *responsabilitat*, sinó que de la mateixa manera avui també és concebible un ésser *humà sense autonomia* i, per tant, *sense responsabilitat*. I no cal anar a buscar soldats nazis jovenets mecànicament enterrant jueus a Ucraïna, o infants africans massacrant la pròpia família per fer mèrits: sense sortir d'aquí mateix la cosa s'entén bastant bé, sobretot quan penses en companys amb Alzheimer que ja han passat el punt limítrof de la humanitat i que ja són definitivament en el de l'animalitat, quan no la «vegetalitat». Qui gosa parlar d'«actes responsables», aquí?
3. Contràriament a l'esmentada «maquinalització» de la màquina hi ha la «antropomorfització» espontània de les màquines que veiem actuar sense saber què fan: allò de creure que el que fa una màquina ho fa, fins a cert punt, «perquè vol» o almenys



permet que passi, com quan renyem el nostre cotxe perquè ha badat i s'ha fumut en un sot, com feia un dia John Cleese, de *Monty Python*, dient el nom del porc al seu cotxe perquè s'havia ficat en un femer. Aquí la nostra tendència natural és a atribuir-li, ingènuament, responsabilitats a la màquina (per això clavem un cop al parafang del cotxe o a la pantalla del PC). **Però** la tendència dominant és la inversa, dir «només és una màquina, fa el que li han dit», com m'ho va il·lustrar a la xerrada de l'Ateneu un ínclit pròcer de la societat catalana quan em va dir, amb to desafiant: «Sí, home, com si haguéssim de tenir mai una màquina que llegís i entengués el que ha llegit!» Vaig aprofitar la seva cara triomfant per dir-li que allò que ell veia absurd havia passat a Barcelona mateix als anys setanta, quan un diari d'aquí va encarregar a una empresa de software local la fabricació d'un documentalista no-humà. (Un *documentalista* sol ser un becari que, cada cop que sorgeix una notícia inesperada sobre algun tema –terratrèmols, mort d'algú, etc.–, furga en els arxius del diari i de la informació que n'extreu redacta un article per contextualitzar). Això s'havia aconseguit a Barcelona feia més de trenta anys i havia funcionat prou bé. L'home em va mirar als ulls incrèdul i sorneguer i em va dir: «La màquina pot haver redactat l'article que se li demana però, què ha entès del que ha llegit?» I jo li vaig contestar: «Res; exactament igual com el becari que abans feia la seva feina».

4. A la nostra època, els animals –ja no dic els de companyia– s'han convertit en *subjectes de dret*, amb drets (i deures). I de responsabilitat, doncs. El símil dels animals domèstics permet centrar bastant la discussió perquè és tanta i tanta la nostra empatia amb ells que els humanitzem sense voler. D'aquí a humanitzar també els nostres assistents robòtics hi va un pas, que els nostres vells, atesos per algun amable robot amb qui s'hauran de comunicar, faran amb una sorprenent facilitat. Aviat hi sentiran tendresa i, espontàniament, els reconeixeran drets, hi confiaran i els donaran un protector marge de confiança. I, quan hi hagi alguna percebuda negligència, ja veurem a qui acusen els parents.

## Què és una persona?

Ja n'he parlat alguna altra vegada, però hi torno. Estudis recents –i confluents– semblen indicar que les persones humanes ocupem tot un espectre de conductes, i els individus s'hi situen consistentment en un extrem d'extrema *mesquinesa* moral (un 15%) i en un altre d'extrema *bondat* (un altre 15%), i aquests percentatges semblen «transculturals» i invariants. L'altre 70% són els individus que s'*adapten* a la situació, els qui «fan el que veuen fer» (recordeu els experiments de Milgram o Zimbardo a Yale i Stanford dels 1960-70, v. Nodes, 43). Que una societat sigui virtuosa o malalta, avançada o ignorant, depèn de cap a on es decanta aquest 70%. El cas teòricament impensable d'una societat amb un 85% de malvats l'hem pogut veure recentment, per citar només tres casos, a l'Alemanya nazi, a la Cambotja de Pol Pot o a la Ruanda dels genocidis. I això no té res a veure amb la moralitat: el 15% de «dolents» –els col·loquialment coneguts com a «malparits» (amb perdó)– potser van tenir una infància dura (o no), el 70% del mig –els acomodaticis– no sol percebre, en el moment, que la seva decisió sigui especialment condemnable (recordeu Eichmann fent arribar els trens a l'hora, o els experiments del «samarità» o dels *graffiti* a Groningen) i el 15% «de pedra picada» potser ho són més per obstinació o simplicitat que per altra cosa i, en el cas dels herois que persisteixen en la seva bona acció i s'hi moren, potser no s'hauran arribat ni a reproduir. Contra la deriva moral o la dinàmica destructiva, tota *l'enginyeria social* que es faci –allò que en diem «educació» o «costums»– ha d'actuar sobre aquesta massa fluctuant, com d'altra banda sap tot bon professor. Aquí, com en *l'evolució* (en què l'únic que compta és *sobreviure*), la moral ni els valors no hi tenen res a veure: només són eines. I l'objectiu, modest, és menys ser tots virtuoses que evitar situacions nefastes. La pregunta és, doncs: en quin terreny se situaran les nostres màquines? Sabrem incorporar-les en la nostra lluita per la dinàmica virtuosa? Doncs potser s'hi situaran elles mateixes: presumiblement perquè sense l'imperatiu reproductiu no hauran d'«adaptar-se» o «fer el que veuen» sinó que, pensades i autocontrolant-se per «fer el que cal», engrossiran i amplificaran el nostre 15% de «bones persones». Si és així, tindrem un aliat inesperat en la lluita contra el 15% de *fdps* –diguem-ho així– i per

arrossegar els altres a la virtut.

Cal recordar que l'espècie humana, contra el que prediquen els moralistes, no és dolenta ni malvada sinó tot el contrari. És pràcticament l'única espècie viva que presenta una característica insòlita, l'*altruisme*. Tan insòlita, que els biòlegs evolucionistes tenen feina per explicar-se com pot haver sorgit, quin benefici procura a l'espècie i per què és tan general i persistent. (Un dels darrers a estudiar-ho, Joseph Henrich, de la U. de British Columbia, conclou, desesperat, que això de l'altruisme és un trencacolls: *fairness is a puzzle*, diu perplex). L'home és altruista, i molt. La seva actuació ha donat lloc a les nostres societats. I aquestes, malgrat la seva història penible i «contorta», són collectius portadors de *moralitat* (que, segons definició, és *altruisme més tolerància, magnanimitat* i sentit de *fair play*). No hi ha cap espècie com aquesta. No vol dir que tots els individus de l'espècie siguin així (el 15% de malvats rebaixen la mitjana), però sí que aquests trets informen el collectiu i les idees (costums, moral, religions, literatura, activitats o pensament) que genera. I a més, contra el que pugui semblar, la humana és una espècie molt poc violenta. Un estudi de fa poc (de John Mitani, d'Ann Arbor, a *Current Biology*) recapitulava xifres: els ximpanzés –almenys els de l'única colònia que experimentalment s'havia pogut estudiar bé, els *ngogo* de Tanzània– són un collectiu *deu* vegades més violent, en nombre de morts (per violència), que les societats humanes de caçadors-recol·lectors. Ja sabem, de fa temps, que la violència humana es va *triplicar* en el pas a l'agricultura i la civilització, però també sabem, estudiant les xifres, que fa dos o tres segles que la violència no para de baixar, de tal manera que ara –encara que els diaris no ho sàpiguen– estem en el punt més baix de la història, a un  $1,0 \div 1,2$  de la xifra paleolítica. Aconseguirem traspasar aquesta no violència als nostres robots, i fer que siguin encara més tranquils i poc sulfurosos que nosaltres (ells que no tindran urgències sexuals)?

Si el que ens preocupa és l'ètica i la tranquil·litat, i tenir robots «persones», cal recordar que ni estem sols ni partim de zero. Un article recent (a *Science*, 26/2/10) enumerava les característiques de «persona» (de *personhood*, s'hi deia) dels dofins, i resulta que les complien pràcticament totes; l'autor, després de repassar totes les definicions i recitant la llista de característiques (intelligència,

amabilitat, llestesa, sociabilitat, etc.), sentenciava: «doncs hem d'admetre que el *dofi* és, efectivament, *una persona*» (dir que és *humà* fóra passar-se). No hauria pas de sorprendre'ns. Fa més de deu mil anys que vam començar a fabricar, amb tècniques d'enginyeria genètica, una nova espècie d'animal que es caracteritza per ser «bo», fidel, *ètic*, llegir-nos la cara i predir les nostres intencions, ser expressiu, bon assistent i company, i amb sentit de la justícia i del deure. Se'n diu *gos*, i ha estat objecte d'estudis recents que s'han sorprès de fins a quin punt és com nosaltres. El gos és, com diu un dels articulistes, un *animal artificial*. L'*homo sapiens* que el va dissenyar se'n va sortir prou bé. Si els humans d'abans del neolític, amb una tecnologia precària que no passava de l'encreuament estimulat de llops, van aconseguir una obra tan òptima, no ha de ser això un bon estímul per nosaltres a l'hora de concebre màquines que com a mínim ens facin companyia?

### *Quin futur ens espera?*<sup>18</sup>

Fa uns anys es va publicar un estudi –potser sorprenent (o no)– que, fent servir això que se'n diu la *metaanàlisi* (l'anàlisi i comparació de mètodes i resultats de *tot* l'escrit sobre un tema), que el mètode fins ara més acreditat per predir el futur, el *Delphi* (que consisteix a reunir experts perquè, raonadament, facin la seva predicció «il·lustrada»), té un valor predictiu aproximadament de *zero*. És a dir, que, vist el que ha passat després, qualsevol no-expert ho hauria endevinat tant com ells fent servir daus o responent les preguntes a l'atzar, o, dit d'una altra manera, que el grau d'incert dels experts no supera el d'una predicció aleatòria. I és que dues coses que ens fan equivocar més en la nostra extrapolació ingènua del present cap al futur és que: (a) no tenim en compte que segurament hi haurà *discontinuitats* (qui ens havia de dir als anys 90, quan encara anàvem amb *walkmans*, que podríem escoltar milers de cançons a l'*iPhone* de l'*Steve Jobs*, l'oponent, llavors en hores baixes, de l'imperi Microsoft?); i que (b) no tenim en compte com ens hi adaptarem *nosaltres*: no sabem com reaccionarà el nostre col·lectiu d'usuaris, allò que en diem societat, davant de les novetats del mercat; els hi trobarà un ús insospitat? les integrarà en el seu mode de vida? (penseu en com l'imprevist «mòbil» ens ha canviat

<sup>18</sup> L'apartat 3 és un extracte de l'entrevista que em va fer l'Albert Juandó com a apèndix al seu treball *Ètica a les infotecnologies*, dirigit per en Miquel Barceló i presentat al Dept. de Llenguatges i Sistemes Informàtics poc abans d'escriure jo aquest article. L'Albert m'ha permès gentilment extreure les meves pròpies paraules del seu treball.

els costums, com l'electricitat va canviar les cases i les fàbriques; l'automòbil, les ciutats i la geografia; la ràdio, la propaganda; i la guerra o la TV, el nostre món sencer).

A risc d'equivocar-se, i molt, sembla que tothom està d'acord en un fet: «previsiblement» la IA del futur pròxim ens «envairà», serà pertot arreu, serà *ubiqua*: estarà incorporada als *gadgets* innumbrables que cada dia sortiran, tots els quals seran «intelligents» i estaran connectats a la xarxa –que vés a saber com serà llavors– i alguns d'aquests *gadgets* amb tota probabilitat els durem incorporats al vestit o al cos (com a autodiagnosticador mèdic, climatitzador integrat, nanofarmaciola o desfibril·lador d'emergència). Molts d'ells probablement estaran integrats a través del mòbil! El *mòbil*, és clar, ja no serà un telèfon –segurament explicarem als nostres néts, davant la seva completa incredulitat, que «allò» va començar essent un *telèfon*– sinó una mena de «comandament a distància» de totes les facetes de la nostra vida, des de la música, les claus del cotxe o de casa, la targeta sanitària (o del que sigui: clubs o DNI europeu), el terminal de «la xarxa» on fer tota mena de consultes, «la TV» confegida al nostre gust, el telèfon –naturalment!–, la càmera d'enregistrar tot el que volem recordar, el(s) llibre(s) que llegim, el «company» que ens porta el cotxe o ens aconsella sobre el camí o ens coneix els gustos... qui sap. I, tot i això, de sistemes intel·ligentíssims complets segurament no en tindrem, per ara.

De moment, i mentrestant, podem anar creant sistemes –modestos– amb un mínim d'empatia i d'imitació intel·ligent que ens facin l'efecte d'estar acompanyats per éssers humans. I és que si donem als robots algunes característiques dels humans (irracionalitats diverses que, lluny de ser *disfuncionals*, són adaptatives, és a dir, que ens han anat bé i ens han fet com som), podem tenir uns robots més o menys com nosaltres, robots de diverses qualitats, segons per a què: robots «savis» –totalment racionals–, robots que especulin a partir d'indicis, robots que evoquin situacions a partir d'una olor, etc. I potser haurem creat una nova espècie, el *robot sàpiens*, amb qui probablement ens entendrem com a companys. Tot això ja s'havia dit, aquí mateix (v. Nodes, 42). Però és probable que no sigui veritat. Vaja, sí, hi haurà alguna cosa de tot això, però... només *vagament* semblant (o no). I és que nosaltres, per preveure el que passarà, pensem *linealment*, per extrapolació raonable. Ara bé, en

un moment donat la trajectòria, determinista o no, pot desaparar-se en un punt en una direcció no sols imprevista sinó indeterminable. De *bifurcacions* n'hi haurà, això segur (penseu en els dinosaures i l'asteroide o el volcà que els va liquidar), i el món que en surti serà tant irreconeixible com ho seria el nostre món d'autos i teles i PCs i videojocs per a un habitant del segle XIX. I ves que la no-linealitat que ens espera no acabi en la molt terribla *singularitat* que alguns prediuen<sup>19</sup> i que es veu que Von Neumann, com tot, ja va imaginar i anticipar als anys 1950 (en converses amb Ulam).

I sobre l'ètica de la IA, què dir? Ja fa temps que es parla de l'ètica dels robots, que és molt menys rudimentària i que va molt més enllà de les septuagenàries lleis de la Robòtica d'Asimov. I ja fa temps que es parla de bastir un marc general i legal que permeti a la societat viure normalment i pacíficament amb màquines intel·ligents, establint els deures i els drets jurídics de cadascun, potser seguint i allargant el paralelisme amb els animals no-humans, domèstics o no, amb què estem habituats a viure. Perquè, per exemple, si un estri domèstic «intel·ligent» ens fa malbé la moqueta o, Déu no ho vulgui, ens fa mal a nosaltres, a qui carai hem de reclamar? O què passa si un avió sense pilot (de fotografia aèria o meteorològic, no cal que sigui d'espionatge o bombardeig) cau sobre una zona habitada i fa mal? La gent continuarà dient: «com és possible haver deixat anar una cosa així, sense control humà»? i tot seguit demanarà: «de qui és la culpa? qui ho pagarà, això?» Però l'home ja no serà l'únic subjecte de dret.

Aquí potser hi hauria d'afegir una qüestió suscitada per alguns «singularistes» (els precursors Good, Vinge, Hanson o Joy, i els actuals, sobretot Ray Kurzweil), que és aquesta: si mai arriba a ocórrer d'aquí 20 a 40 anys la *singularitat*, que és aquell punt a partir del qual les nostres màquines intel·ligents ja seran més intel·ligents que nosaltres i, per tant, estarem a les seves mans, caldrà haver previst abans algun mecanisme que eviti això últim. Per exemple –com ja teoritzava el 1984 la novel·la de ciència-ficció *Neuromancer*, de William Gibson (amb la seva *Turing police*)– un «autofrè», automàtic, incorporat, que faci que les màquines intel·ligents, en la seva evolució, «no ens superin» (gaire), o no ens facin (gaire) mal, o alguna cosa d'aquest estil. O potser haurem de confiar que la seva evolució els porti a reforçar la tendresa o magnanimitat amb els

<sup>19</sup> Si es vol tenir una visió de síntesi de la famosa *singularitat* (v. l'apartat 3 supra) pot valdre la pena fer un cop d'ull a l'article *Technological singularity* ([http://en.wikipedia.org/wiki/Technological\\_singularity](http://en.wikipedia.org/wiki/Technological_singularity)) de la Wikipedia –en català encara no hi és–, si bé trobo que el treball és fluixet i a estones no sap evitar passar-se d'alarmista.

inferiors (humans). Alguns addueixen que, alliberades de l'imperatiu reproductor, les màquines no sabran res de l'amor i, per tant, no seran tendres, diuen. En això segurament s'equivoquen, perquè segons estudis evolutius recents l'amor i la reproducció no semblen tenir gaire relació ni amb la *tendresa* ni amb la *magnanimitat* (més aviat al contrari: amb la possessió i la violència) sinó que aquestes dues qualitats semblen haver-se originat, en els homes, en la guerra i en el conflicte (v. *Conflict: altruism's midwife* a *Nature*, 20/11/08, ja citat a *Nodes*, 45), com a revers de les qualitats (bèl·liques) contràries –l'odi i l'acarnissament– i amb el suport químic de l'oxitocina del cervell. Horror! L'amabilitat filla de la guerra? la guerra partera de la cavalleriesitat? Quin daltabaix en els nostres mites i tòpics «bonistes»! Doncs segurament és així, i pel nostre bé esperem que per aprendre aquestes virtuts les màquines no s'hi hagin d'exercitar en cap guerra. I que, per algun miracle, siguin almenys tan bones i pacífiques com som capaços de ser-ho nosaltres, i més.

De tot això els «singularistes» en diuen fer *màquines amigues*, gràcies a la *friendly AI*, que és la que hem de promoure, potser amb urgència i sense badar gaire. Com ho farem, això? Els «singularitaris» no ens ho diuen, i em fa l'efecte que per evitar el possible desastre confien més en un miracle que en els nostres mitjans. Vaja, que molta tecnologia-ficció però al final tot queda en una pràctica molt antiga: *resar*.

## *Els robots, els humans i l'Internet*

NODES  
NÚMERO 48  
PRIMAVERA DE 2011

**E**N una entrevista en l'últim Fulls dels Enginyers (set./oct. 2010) Josep Amat, de felicitat i recent jubilació, venia a dir això: "preparem-nos, perquè la cosa no ha fet més que començar", i deia que a Catalunya ens calen professionals –un enginyer per cada dos o tres-cents robots– que els projectin, instal·lin i se'n cuidin. I, afegeixo jo, també ens cal que els qui no són enginyers, la resta de la societat, els acceptin i s'hi adaptin, tant els qui els facin servir o se n'aprofitin com els qui, cada vegada més, en depenguin.

### *A. Robots pertot arreu*

Des del primer (un braç articulat Unimate instal·lat a la General Motors el 1961), els robots no han parat de créixer, i avui els industrials ja són més d'un milió, la meitat a l'Àsia. Els robots ja fan de tot: d'aspiradora, de talladors de gespa en camps de golf, de planxadors de roba (encara no, però aviat), de recollectors de fruita dels arbres, de peladors de préssecs o raïm (aviat), de pesadors-palpadors d'objectes a classificar, d'intrusos en la cavitat toràctica d'un pacient per connectar-li els elèctrodes al marcapassos, d'inspectors d'esquerdes en canonades de servei, de vigilants ocults de carrers o d'accessos al metro per "detectar" conductes sospitoses o possiblement conduents a delictes o violació. I així.

Les dues institucions de la Terra més robodependents, amb gran diferència, són els militars americans i el poble japonès. No hi ha com veure cap on van aquests dos per ensumar el que ens espera a la resta de mortals uns quants anys després. Del primer ja n'he parlat sovint. N'hi ha prou de recordar que els pressupostos



són enormes –però incognoscibles, perquè estan prou repartits en conceptes diversos– i que continua vigent l’objectiu que pel 2015 la meitat de les forces armades americanes estiguin robotitzades. De moment, ja n’hi ha uns quatre mil a l’Iraq i a l’Afganistan netejant mines o, en forma de borinots (*drones*, oficialment dits *unmanned aerial vehicles* o UAV), participant en feines de reconeixement i –també– de combat, on, tot i que volen amb un controlador humà remot que els vigila, cada cop tenen més autonomia i decideixen més –tenen “responsabilitat”, doncs?– si bé, per sort per a ells, encara no surten als vídeos de WikiLeaks.

L’altre capítol a part són els japonesos. Són aquest un poble envellit que fa temps que ha decidit no donar cap visat d’entrada a cap immigrant sense qualificacions. Això vol dir que no tindran colombians o filipines que es cuidin dels seus vells i hauran de fer-se amics del seus robots, com qui es fa amic del seu gos, perquè en dependran en ajuda i companyia. Els japonesos han decidit que no cal que siguin humanoides (però millor que ho siguin) i han identificat les feines a què han de dedicar els seus, que són les de les tres K: les perilloses (*kiken*), les brutes (*kitanai*) i les dures (*kitsui*). De moment, com que la robòtica és una enginyeria molt intensiva en capital, han començat per les aplicacions que diríem frívoles, com ara la de fer de recepcionista, o de pacient, perquè hi practiquin els aprenents de dentista, o, més seriosament, de muntadors o empaquetadors de productes fabricats, o de tastadors o provadors –avorrits i repetitius– d’aliments o begudes (comprovant que els recipients estiguin exactament plens o les etiquetes correctament posades, col·locant les trufes elegantment a la capsca, classificant pomes, disseccionant amb precisió canals a l’escorxador, tallant llesques o preparant amanides, etc.), i tot sense haver d’anar al lavabo ni fer becaines. O de cirurgians que operen des de fora del cos o, menuts, que s’hi fiquen. Com també microrobots proveïts de microcàmeres que s’introdueixen en un edifici i, tot volant, filmen tot el que hi veuen (... i ja es veu que aquí entrem en terreny de la privacitat: ens agradarà que es fiquin en la nostra vida?). I també, és clar, cuidar vells, fer-los companyia, entretenir-los, vestir-los, traslladar-los, etc.

Que els robots es dediquin a totes aquestes coses vol dir que hauran de ser “bons” i de “fer bondat”. Perquè es trobaran ine-

vitablement, com nosaltres, amb dilemes ètics. I no és el mateix –o ho és?– dir-li el que ha de fer que no que ho faci, és a dir, que “decideixi” fer-ho després d’haver considerat els pros i els contres, i les probabilitats. Potser no caldrà que “senti” que ha fet bé ni que sigui “conscient” del que ha fet: al cap i la fi, nosaltres no acostumem a fer cap de les dues coses. Ara bé: si la decisió és errònia o perjudicial, què farà?... “penedir-se’n”? pagar-ho? (com?) fer-ne un informe? (perquè n’aprenuem i no torni a passar?). Ja es veu que aquí tenim molt de camp per córrer. Però l’hem de córrer de pressa i sense badar.

### *B. Defectes de fàbrica*

Ja he dit més d’una vegada que *l’homo sapiens* és defectuós, que està mal dissenyat. En els darrers números m’he dedicat a fer diverses enumeracions de «defectes» que tenim. No vull dir «defectes a superar» sinó característiques que tenim i que ben bé ens podríem estalviar, que denoten un mal disseny de l’espècie humana (o que són el residu històric d’adaptacions passades) i que, posats a fer robots, valdria més que aquests no els tinguessin, pel seu bé i pel nostre. Aquí actualitzo la llista amb algunes de les darreres incorporacions publicades. Els humans, en efecte,

- tenim defectes de memòria, que ens porten a reconèixer falsos culpables a la comissaria o al jutjat (\*\*);
- creiem una cosa segons la confiança amb què és dita: si sentim un testimoni que afirma confiat el que potser dubta o no sap, tenim una tendència invencible a creure’ns-el molt més que no pas el que constata una cosa certa que ell sap i pot contrastar però ho fa mostrant modèstia o inseguretats (v. *Science*, 17 set. 2010, p. 1478);
- creiem una opinió o teoria segons l’autoritat que conferim a qui l’exposa: si abans hem comprovat que l’expert «pensa com nosaltres», la seva expertesa es torna inqüestionable (v. el *Journal of Risk Research* del setembre del 2010);
- topem amb els tabús: en una negociació, si hem de cedir en un tema que «sentim» com a «sagrat» o «intocable» ens hi neguem

en rodó, com es comprova en temes de política (paraules clau: «Israel», «Palestina», «poble», «drets», etc.), si bé, demostradament, el problema es pot evitar canviant el llenguatge amb un virtuosisme adequat (v. Dehghani en el *Journal for Judgment and Decision Making*, 1r trimestre del 2010);

- podem constatar com l'autoritat exercida brutalment aconseguix, tot i que sembli impossible, construir autèntics «sentiments» en els individus i teixir complicitats humanes. Dictadures llargues i repressives acaben fent «estimar» –fins i tot amb tendresa –les pàtries o conceptes que imposen per la força (el cas nazi és canònic, però n'hi ha molts altres); el mateix fenomen explica com hi pot haver estimació filial a pares de conducta repulsiva..., encara que en aquest cas la dependència econòmica i l'oxitocina (v. més avall) segurament hi ajuden;
- notem que els estereotips ens persegueixen pertot arreu. El sentir-se fora de lloc i/o jutjat com a «diferent» fa que el nostre rendiment baixi en picat, un fet comprovat per exemple en exàmens de selectivitat, cosa que explica moltes disparitats entre sexes o races. Un indici: si en una prova s'explica prèviament que l'examen no detecta ni té res a veure amb el sexe o la intel·ligència de l'examinand, el resultat automàticament millora (v. Leander et al. en el *Journal of Experimental Psychology*, 46, 2010). Un altre indici: si un sol·licitant de feina veu que la gent amb què haurà de treballar és més jove (o gran, o culta) que ell, automàticament se sent empetitit i desfavorable. Un altre exemple, ben gràfic, l'explica l'autor mateix de l'anècdota, un postgraduat negre que, havent detectat que la seva presència creava instintivament una certa prevenció en gent blanca, va veure que si xiulava un tema de Vivaldi la reticència desapareixia de cop (v. Claude Steele: *Whistling Vivaldi, and other clues to how stereotypes affect us*, Norton, 2010);
- veiem que més val tenir bon accent. Fins i tot a les universitats americanes, on quasi tothom té accent «de fora» i teòricament no hi ha discriminació sinó acolliment, s'ha comprovat com és de potent l'actitud subtil –una «pèrdua de fluència cognitiva», no pas un prejudici– que provoca un accent no prou estàndard

en els circuits cerebrals interpretadors del llenguatge (v. Lev-Ari i Keysar, de la U. Chicago, al *Psychological Science* del juliol del 2010); apreciem els apreciables, però sense passar-nos. Un curiós experiment de Parks i Stone (*Journal of Personality and Social Psychology*, agost del 2010) demostra –i ho quantifica– que tothom admira les persones generoses però que la gent no sols avorreix les persones incompetents i impredecibles sinó també les egoistes i –atenció– també les seves contràries, les persones «despreses» (Raó? Doncs això: amb aquest tipus de persona els participants se senten incòmodes, i ho raonen així: «Tothom ha de fer ús del que té de manera igualitària»);

- ens deprimim sovint, cosa que fem aparentment per fer llàstima i pidolar ajuda. (Però, atenció, això només és una teoria, fortament discutida, que sosté que la depressió és funcional i –evolutivament– adaptativa); ens passen gorilles pel davant (i no ens n'adonem). La frase està treta del títol del llibre que explica aquest i altres defectes de l'atenció humana (Chabris i Simons: *The invisible gorilla and other ways our intuition deceive us*, Crown 2010, recensat a *Science*, 27 agost de 2010); som confiats, sobretot els infants; i més els humans que no pas els bonobos i molt menys encara els ximpanzés normals. Cosa que s'explica, segons *Science*, 14 nov. de 2008 i *Current Biology*, gen. de 2010, perquè els humans som neotènics també en això (vull dir que conservem en l'adultesa trets infantils com ara aquest, o la curiositat, o la cara plana);
- tenim aquest altre dimorfisme sexual: les dones valen per dos homes, però no més. Vull dir que el nombre de coses que es poden fer a la vegada, que en el cas dels mascles humans és –tòpicament– igual a un, és, en el cas de tots (i això inclou les femelles), igual a dos, com a màxim. Es tracta d'una limitació física, cerebral: un límit al *multitasking* humà (clarament superable per les màquines);
- patim les traumàtiques síndromes del *niu buit* i del *cor trencat* quan un fill se'n va de casa o quan ens abandona un(a) amant. Són el corollari, quan la relació s'acaba, del lligam fortíssim, i plenament funcional, que uneix mare i fill o dos amants, respectivament, mentre la relació dura;

- rebem contínuament un diluvi de substàncies. Els humans ens veiem inundats periòdicament amb dutxes no sempre desitjables ni funcionals de neurotransmissors, hormones i altres drogues, sovint addictives, com ara la dopamina (guia la nostra conducta, a través de la cerca del plaer i del premi i sovint en té prou amb la seva mera anticipació), la serotonina (fixa l'estat d'ànim), l'oxitocina (l'hormona de l'hipotàlem, que promou la confiança i que ens fa estimar el fill o la parella i així, miraculosament, ens torna monògams i bons pares, com s'ha demostrat injectant-la en talps o ocells promiscus, que immediatament deixen de ser-ho), la testosterona (tothom sap què fa, però contra l'opinió popular també opera en les dones i bàsicament consisteix en un delicat mesurador-corrector de l'estatus), l'estrogen (construeix el fetus mascle i guia els rituals de festeig), el cortisol (ens «tensa» fins a l'«estrès»), i així:... *nar-hi nant*... Ara bé, sobretot: estem desprevinguts contra l'estupidesa. Sobre això, vegeu la nota que figura al capdavant d'aquest article.

M'he proposat anar enumerant les deficiències humanes a mesura que es van descobrint i estudiant perquè opino que són més aviat *disfuncionals*. Deixo al lector la consideració del que faria un robot amb cada un d'aquests trets i de si algun li serviria per aprofitar millor els seus avantatges o li serien més aviat un impediment. En particular, considero que, de cap manera, els robots no haurien de caure indefensos en l'estupidesa (\*) que segur que trobaran, i que hi haurien de saber reaccionar molt millor que no pas nosaltres.

### C. Algunes preguntes de mal contestar:

#### 1. Estem massa capboirats per Internet?

<sup>20</sup> L'autor es refereix a l'article de la plana 353. (n. de l'ed.)

Nicholas Carr, de qui ja parlava al número 45 de Nodes<sup>20</sup> a propòsit del seu article de fa dos anys a l'*Atlantic Is Google making us stupid?*, ha tret un nou llibre, de títol autoexplicatiu: *The shallows*, subtitulat a Europa *How the Internet is changing the way we think, read and remember* i a Amèrica, més amenaçadorament, *What the Internet is doing to our brains*. La seva hipòtesi és que Internet ens està estovant el cervell: que per tenir idees no n'hi ha prou amb tenir

dades, sinó que aquestes s'han de covar, o paladejar, abans que no en pugui sortir res de creatiu. És evident, i Carr ho reconeix, que cada nou mitjà –l'alfabet, el rellotge, la impremta, o la civilització de les imatges– ha transformat la nostra manera de pensar: n'hi ha prou a recordar aquí com abans de Plató hi havia qui s'espantava davant dels efectes previsibles de la invenció de l'escriptura («si ja no cal la poesia ni l'oralitat, qui recordarà res, ara?», li feia dir Plató a un grec il·lustre). Carr s'espanta amb Internet i ve a dir el mateix de l'amic de Plató, afegint-hi «i qui serà capaç de pensar, ara?». Ens veu envoltats per tot un món de coneixements invasius però superficials (*shallow*). En la tradició autoflagel·latòria americana d'Allan Bloom i el seu quasi-apocalíptic

*The closing of the American Mind*, del 1987 (a propòsit de l'educació i la cultura poc auto-exigents), Carr ja insinuava el 2008 que el PowerPoint no sols havia canviat la manera de presentar les (poques) idees dels executius de les grans companyies sinó la manera mateixa en què (potser) pensen. Aviat es va adonar, diu, que la «navegació» pels mars de Google és un perill molt més gros, perquè per pensar, deia, no sols s'han de tenir coneixements sinó que s'han d'haver madurat i s'han de poder seleccionar, combinar i controlar activament. Al seu llibre Carr veu en la superficialitat de l'Internet un cas particular del món en què vivim, en què un bon professor és menys el qui transmet continguts que no el qui sap «actuar» millor. (I no parlem dels polítics, amb el seu gest o la seva «frase» efímera, premastegada i a punt perquè la premsa l'espombi...!)

Tres conseqüències que Carr no remarca però que tothom ha pogut constatar en qualsevol mostrador són aquestes: (a) estar atents a una pantalla converteix les persones que ens atenen en éssers indiferents i antipàtics (no se sap ben bé si estan per tu o caçant musaranyes), a part d'ignorants (vull dir que no saben dir res que no surti per pantalla, de la qual semblen incapaços d'apartar els ulls ni un moment), (b) el contacte visual amb la pantalla exclou el contacte visual amb la persona del davant, amb la qual ja no s'és capaç de parlar ni, molt menys, de xerrar amigablement; i (c) la pantalla sembla excitar el zel comunicatiu dels teclejadors amb col·legues o amics, però només per passar-se reguitzells de lletres o missatges, i no pas per –senzillament– parlar (segons el tradicional

i acreditat mètode d'emetre sons tot veient la cara del teu interlocutor, i tot amanit amb gestos carregats de significat que tots dos saben interpretar). I això sense pensar amb el vincle de confiança entre metge i pacient, que perilla de trencar-se definitivament si la pantalla és la font del diagnòstic i el metge, amb la mirada gèlida i abduïda, només fa de comunicador.

Vistes totes les desgràcies que es preveuen, n'hi ha que s'hi apunten, al Carr (perdoneu-me la brometa fàcil: *és un txist!*), i, tot superant-lo, proclamen la fi de l'era del llibre i del treball intel·lectual, la de la scholarly mind, que va començar amb Gutenberg i s'acaba ara... Pot ser... Però Jaron Lanier, l'inventor de la realitat virtual, diu al seu llibre *You are not a gadget: A manifesto* (2010) que no n'hi ha per a tant, que hipernavegar per Internet és com afartar-se de caramels, i que arriba un dia que un mateix se n'acaba atipant, dels tips, i que, com en tot, el remei és la prudència i l'autocontrol.

Qui ho sap? D'aquí a uns anys segurament en podrem fer la prova. Donarem a llegir aquests articles –o una simple frase que contingui ironia– a persones i robots: si només l'entenen aquests, és que ja haurem traspasat el terrible llindar de l'estúpidesa, contra la qual, com deia Schiller (i Asimov), fins i tot els déus lluiten inútilment. (\*)

## 2. *Qui és responsable de la gestió correcta dels avenços informàtics(\*\*\*)?*

Raonablement, si hi ha algú capaç de resoldre un problema i d'explicar-lo a la societat és qui el coneix bé. Normalment, per detectar i resoldre problemes d'informàtica –el de la invasió de la vida privada, o el de com fer «amigables» les màquines futures– no hi ha d'haver millor coneixedor que l'informàtic professional, que hi té els genolls pelats. Ep!, si cal, assessorat per tots els advocats i polítics que calguin, però és ell el resolador principal, per experiència i criteri. Com en tot, l'essencial és tenir seny, senderi o sentit comú i mantenir, si es pot, una mínima serenor (dir-ne «objectivitat» fóra passar-se de solemne). I és que per preveure, detectar i resoldre problemes en un camp no hi ha com: 1) estar profundament impregnat de l'activitat en qüestió i de la

seva problemàtica, tant de l'habitual com de l'excepcional; 2) ser mínimament intel·ligent i/o lúcid, i 3) tenir el sentit comú prou rodat. Tot això, és clar, sense tenir en compte el cas en què qui ha de resoldre el problema –o qui (només ell) el pugui resoldre– sigui qui l'ha provocat, cosa que de vegades passa... I que potser no és dolent del tot: qui maquina maldats més val que treballi per a nosaltres, ens farà adonar de totes les maldats possibles; no és pas per casualitat que quan els policies –o els sistemes de seguretat– volen resoldre un problema contracten delinqüents –o *hackers*– perquè els facin la feina: són creatius, insuperables. I, triar per triar, val més que estiguin de la nostra banda.

### 3. *Internet i la propietat intel·lectual (\*\*\*)*

Les preguntes que se solen plantejar són: (1) cal assegurar-hi la lliure circulació? (2) es dinamitaran els drets d'autor i la creació artística? No ho sé pas. Com a persona sensata i mínimament experimentada sé que el futur és imprevisible, com ja he dit. Només puc constatar això: el «pirata» de A és per a B un «dinamitzador», i el «terrorista» destructor d'un és l'heroi «alliberador» de l'altre. Sense les fugues de secrets tècnics dels gremis medievals no tindríem revolució industrial, sense pirateig de patents industrials no hauríem tingut indústria moderna ni a Catalunya ni al Japó modern, i si Jobs no hagués robat descaradament el GUI de l'Alto de Xerox, Gates no hauria pogut piratejar-lo de Jobs. I si les telefòniques, les musicals i les editorials es veuen a la ruïna potser no se n'haurien de queixar tant: abans va passar amb IBM amb el PC o amb Hollywood amb els DVD, que es van deixar prendre el producte davant dels seus nassos, presos d'una estultícia que només es pot explicar perquè encara era més gran la supèrbia infinita que gastaven, amb un sentiment d'invencibilitat a prova d'errors i mals càlculs: ara ja fa temps que demanen justícia i acusen els seus despulladors de falta d'«ètica»; sol ser l'excusa principal que donen els perdedors, encara que la perdició se l'hagin guanyat a pols. N'hi ha prou constatant el fenomen recurrent de les queixes d'Anglaterra contra Alemanya o els EUA per haver-los copiat la revolució industrial, o els dels EUA contra els japonesos perquè gosaven fer cotxes millors que els seus o, ara mateix, de tots con-



tra la Xina que ens treu la feina dels nostres treballadors pagant misèries als seus i embrutant-ho tot. Tot plegat, *déjà vu*.

El que cal considerar objectivament, fora d'aquests esgarips d'animal ferit o de ric malacostumat, és si l'adquisició, *piràtica* o no, va resultar en un millor nivell de prosperitat i civilització per al conjunt. Aquesta és la prova del cotó-fluix. I millor deixar fora, si és possible, la «moral». Qui se'n recorda, avui, de les justíssimes protestes dels baleners de Nova Anglaterra –que feien espelmes– quan Edison amb la seva bombeta elèctrica els va arruïnar d'una manera cruel i miserable? Doncs van haver d'abandonar el seu mitjà de vida i els seus pobles i barques baleneres, sense remissió ni consol: anaven contra «el progrés» i no van trobar ningú que els compregués o els tingués en les seves oracions. La vida és cruel –ho sabem des de sempre–, i el progrés tècnic avança per destrucció del que ja no té sentit –ho sabem des de Schumpeter. La pirateria, en particular, ha estat sempre, com a mínim des dels víkings, una de les maneres més efectives de desbloquejar mercaderies i de dinamitzar mercats. Modernament, a base de no respectar la propietat intel·lectual (concepte molt recent, novíssim) la pirateria s'ha convertit en una manera de portar idees al mercat perquè arribin al consumidor, manera de la qual es pot discutir –tant com vulguem– tant l'efectivitat com la moral.

Per cert, sobre aquest tema de la pirateria, i dels drets d'autor, fa poc ha sortit un llibre que en parla (Adrian Johns, *Piracy: The intellectual property wars from Gutenberg to Gates*, University of Chicago Press, 2010), molt il·lustratiu, si bé té l'inconvenient que l'autor admet que no sap imaginar què pot passar en el futur i confessa no ser gens expert en els aspectes legals de la qüestió, en els quals sí que hi entén Michael Gollin, el seu recensor a *Nature* (25 feb. 2010), que resumeix el seu (propi) punt de vista així: (a) La pirateria no és un concepte moral sinó una dinàmica de la realitat que cal que tothom, tant científics com advocats, polítics, empresaris o consumidors, compreguin bé abans de prendre-hi posicions (més val que no siguin ideològiques); (b) Quan els pirates guanyen la partida, la innovació i la qualitat se'n ressenten, mentre que quan l'exclusivitat és excessiva, la societat perd la facilitat i els beneficis d'accedir-hi, i (c) Per compensar els interessos d'innovadors i copiadors, i perquè de tot en resulti un ambient sa i creatiu, cal un

esforç i vigilància constants. No hi puc estar més d'acord.

#### 4. *I la privacitat? (\*\*\*)*

Scott McNealy va dir, fa uns quants anys, que lluitar per la privacitat és com somiar truites. Literalment va dir: «Privacy is dead. Get over it.» Doncs, com diu ell, ens hi haurem d'acostumar, superar el problema i no perdre-hi més temps. Crec que la cosa va bastant per aquí. Podem fer tot el que calgui per preservar-la, però cal admetre que és una lluita contracorrent, 1) perquè les innovacions permeten envair-la, 2) perquè els malintencionats ho aprofitaran, i –atenció!– 3) perquè no tothom estarà sempre en contra de la intromissió en certes vides: sí que hi estarem quan els interceptats siguem nosaltres mateixos o «les bones persones», però no tant, o gens, quan la intromissió i el seguiment estret sigui de malfactors o terroristes que hagin fet o vulguin fer un mal inacceptable a gent innocent. I, si no, deixem passar el temps i ja ho veurem. Jo mateix em podré queixar que em segueixin pel carrer –o que em segueixin la mirada dintre el supermercat–, però ja no em semblarà tan malament si la càmera que em seguia permet detenir el delinqüent que m'hi ha robat la cartera –o que, per l'estudi detallat de les compres que solc fer, ja no em molestin més amb propaganda indiferenciada sinó que només m'anunciïn la sortida de productes que «saben» que em poden interessar. Això últim que explico és el que em fan els d'Amazon: de tant en tant em diuen «coneixent-lo com el coneixem, potser li interessarà saber que ha sortit aquest llibre». I jo, francament, no sols no ho considero una intromissió sinó que fins i tot ho agraeixo. I puc demanar que deixin de «coneixe'm», cosa que jo haig de poder exigir per defensar la meva privacitat i que, en aquest cas, Amazon respecta.

En resum: com sempre, la tecnologia –com la realitat– forneix eines per resoldre problemes o, en mans de segons qui, per crear-ne de nous. Per tant, cal estar molt atent a les maneres que trobin els maliciosos per portar la tecnologia disponible al seu molí. Anticipar-les, detectar-les, contrarestar-les, etc. Si cal, incorporar-los al nostre camp. I «fins la pròxima!», perquè hi haurà noves eines, i els maliciosos trobaran noves maneres de fer mal,... o hi haurà uns maliciosos diferents. Caldrà vigilar, doncs, i no abaixar

mai la guàrdia. I caldrà explicar-ho als polítics, que sempre o són ignorants del tema o n'estan tan allunyats que el seu criteri és irreal i la seva acció inefectiva. O bé que facin de polítics –una temporadeta només– justament aquells que coneguin el problema, i el resolguin, i fora!... I així sempre... i fins la pròxima!...

Barcelona, 25 d'octubre de 2010

(\*) A propòsit de la l'estupidesa, i la seva brillant teorització per l'inclit i malaguanyat Carlo Cipolla, m'agradaria esplaiar-m'hi un dia aquí, com ja vaig fer, amb gran èxit, en un congrés de l'ACIA fa anys a Mallorca. Perquè, desgraciadament, la desproporció entre els qui pateixen inermes l'estupidesa humana i els qui hi poden muntar una defensa cipolliana racional és enorme. I crec sincerament que tenim el deure d'instruir els nostres robots sobre el tema, perquè se'l trobaran a cada cantonada. Prometo posar-m'hi algun dia, encara que no sigui un tema pròpiament *aciàtic* i només sigui per consolar-nos, i riure una estona.

(\*\*) Sobre el record com a «reconstrucció» i no pas com a «instantània fotogràfica», ja m'hi vaig referir al número 46 de Nodes<sup>21</sup> («Notícia 4»), a propòsit d'aquest article, que recomano:

Hal Arkowitz i Scott Lilienfeld: *Why science tells us not to rely on eyewitness accounts*, Mind (gener 2010) en el qual es fa èmfasi en els resultats i l'obra general de l'Elizabeth Loftus, que ha anat sempre en la línia comentada («la memòria com a reconstrucció») i les implicacions jurídiques, policials i psicoanalítiques de la pregunta què és, realment, un record? Poso aquí sota una petita bibliografia de la Loftus perquè es vegi que tot això ja li ve de lluny. *Witness for the Defense : The Accused, the Eyewitness & the Expert who puts Memory on Trial* — St. Martin's Press, 1991

*The Myth of Repressed Memory: False Memories & Allegations of Sexual Abuse* — St. Martin's Press, 1994

<sup>21</sup> L'autor es refereix a l'article de la plana 365. (n. de l'ed.)

*Eyewitness Testimony* — Harvard University Press (1<sup>a</sup> edició 1979, revisada 1996).

Va ser cap al 1990 que la vam tenir a la Facultat d'Informàtica de Barcelona en un congrés d'Intelligència artificial que va organitzar la malaguanyada i enyorada Núria Piera sobre «Física qualitativa», un dels temes «calents» que llavors hi havia a l'agenda de la IA. Loftus venia de la UC a Irvine, era membre associat (*fellow*) del Centre de Neurobiologia de l'Aprenentatge i la Memòria, i va resultar ser una persona molt competent i pròxima.

(\*\*\*) Els apartats 2-4 són extractes d'una entrevista que em va fer l'Albert Juandó com a apèndix al seu treball *Ètica a les infotecnologies*, dirigit per Miquel Barceló i presentat al Departament de Llenguatges i Sistemes Informàtics (LSI) poc abans d'escriure jo l'article del número 47 de Nodes<sup>22</sup>, en el qual ja apareixia la meva resposta a una de les seves interessants qüestions. L'Albert em va permetre gentilment extreure les meves pròpies paraules del seu treball.

<sup>22</sup> L'autor es refereix a l'article de la plana 373. (n. de l'ed.)



## *És l'estupidesa humana exportable a les màquines?*

**E**N el número anterior parlava dels moltíssims robots i programes intel·ligents que tindrem d'aquí a no gaire temps, robots que parlaran amb nosaltres i ens entendran i que cuidaran els nostres vells, alguns del quals –reconeguem-ho– seran/serem rematadament estúpids. I tal com ja ho deia en l'article anterior: l'estupidesa és una cosa molt seriosa. Schiller la trobava tan impossible de combatre que escrivia que els déus mateixos hi lluiten en va (Asimov va convertir la frase en el títol d'una novel·la seva: *The gods themselves*). I Einstein deia que l'estupidesa humana era l'única magnitud realment infinita que coneixia. Com que fóra una estupidesa suprema –valgui la redundància– no evitar que els robots que fem siguin també estúpids, o es vegin descoratjats per la nostra –la nostra estupidesa, vull dir–, cal pensar-hi a poc a poc abans que no se'ns encomani. Heus aquí algunes idees inconnexes, per anar-hi meditant.

NODES  
NÚMERO 49  
TARDOR DE 2011

### *A) L'estupidesa vista per fora: les lleis de Cipolla*

«Tots els qui semblen estúpids ho són i, a més, també ho són la meitat dels qui no ho semblen» (Quevedo)

Carlo-Maria Cipolla (1922-2000) era un historiador de l'economia, italià, amb un peculiar sentit de l'humor. Va escriure un assaig en broma –però només en el to– sobre la importància dels estimulants erèctils en el comerç mundial: hi demostrava que la recerca de precedents de la Viagra havia creat el comerç sencer de les espècies, la revolució de Mahoma, el Renaixement italià, la revo-

lució industrial i l'era moderna; vaja, tot. No content amb aquesta relectura diguem-ne «entremaliada» de la Història, al llibre *Allegro ma non troppo* (1988) també va voler compartir amb tots nosaltres el seu descobriment de l'essència de l'estupidesa, que ell va formular en termes matemàtics.

Suposem, diu Cipolla, que poguéssim classificar les accions humanes en egoistes i altruistes i que les volguéssim representar, segons la seva qualitat i intensitat, en les coordenades cartesianes  $x$  i  $y$ . Així, les accions que beneficiessin el subjecte estarien a l'eix  $x$ , amb magnitud proporcional al benefici obtingut o obtenible (o perjudici; en aquest cas l'acció es representaria a la banda esquerra, negativa, de l'eix). Igualment, les accions que comportessin un benefici –o perjudici– per als altres s'alinearien al llarg de l'eix  $y$ , proporcionalment a la seva intensitat. D'això resultaria una partició de l'univers de les accions humanes en quatre quadrants.

Com és obvi, el quadrant 1 seria el de la gent normal, que és la que fa les coses de manera egoista però, generalment, sense perjudicar els altres (Cipolla en diu “els intel·ligents”). El quadrant 2 seria el dels qui podríem anomenar desgraciats, gent que, sense fer mal als altres, es perjudiquen a si mateixos en l'acció. Aquesta gent es podria dividir en masoquistes, els qui hi posen intenció, i beneïts, els qui ni tan sols se n'adonen. El quadrant 4 és el dels qui podríem batejar col·loquialment com a *malparits*, ja siguin de la varietat sàdica o directament de la universal i estesa varietat de FDP (un acrònim que espero que s'entengui, dit informalment i amb perdó).

I a qui correspon el quadrant 3? Doncs, per definició, als estúpids, gent que fa coses que perjudiquen els altres però que, alhora i incomprendiblement, no en treuen res de bo i fins i tot, potser, en surten perjudicats ells mateixos fent-ho. Aquesta gent és perillósíssima, diu Cipolla, perquè no es pot entendre de cap manera per què fan el que fan, ni predir-ho, ni muntar una estratègia d'evitació o de minimització. Pertanyen, doncs, al domini irracional i impenetrable de la natura i de la conducta humana. Vistos com a col·lectiu, diu l'autor, els estúpids formen un grup poderosíssim i destructor, molt més que totes les màfies i les multinacionals juntes, que, a més, actua al marge de totes les lleis humanes i divines amb una gran efectivitat i amb una coordinació envejable.

Cipolla hi afegeix tres lleis que observa en el seu funcionament: 1) Tots nosaltres, inevitablement, subestimem el nombre d'estúpids que hi ha en circulació. 2) La probabilitat que una determinada persona sigui estúpida és independent de qualssevol altres característiques que pugui tenir aquesta persona. 3) Els no estúpids sempre i sistemàticament subestimen l'enorme potencial destructiu (l'infinít d'Einstein) dels estúpids.

Així, doncs, què li direm a un robot perquè no es torni un estúpid? I què cal que faci quan se'n trobi un que ho sigui? No tenim cap idea clara de com es crea l'estupidesa, i només una vaga idea dels mecanismes que hi pot haver al darrere. Òbviament, és d'interès prioritari descobrir-los, si n'hi ha, per evitar-la o per combatre-la. Tot, abans que resignar-s'hi. Tot parafrasejant Burke: «Perquè triomfi l'estupidesa només cal que la gent normal no faci res». La broma de Cipolla es pot continuar, com fa l'italià Pino Aprile en el seu llibre *Elogio dell'imbecille* (2002), subtítulat *Els intel·ligents han fet el món, i els estúpids hi viuen de conya*. Però com que l'estupidesa pot fer gràcia però és dissolvent i perillosa, a partir d'aquí en parlarem en to seriós. Cipolla mateix ens hi va alertar, en un llibre del 1994 (*Tre storie extravaganti*): «Sovint es diu que l'home aprèn de l'experiència. Haver estat historiador durant més de quaranta anys m'ha convençut que aquesta creença és ingènua, i que l'home no n'aprèn res, de l'experiència –ni de la individual ni de la col·lectiva–, i que, per tant, continua repetint amb monòtona perseverança els mateixos errors i crims, amb conseqüències nefastes per al progrés humà.”

### *B) L'estupidesa vista per dins: resultat d'una dissonància cognitiva, potser?*

S'ha dit que l'estupidesa és l'única malaltia que no pateix el malalt sinó la gent que té al voltant. I també s'ha dit que si una persona normal repeteix una estupidesa un nombre suficient de vegades acabarà pensant que aquesta actitud és normal, la defensarà i la incorporarà: resulta més còmode. I si justificant una estupidesa se sent incòmode, en cometrà una altra. I així, un cop creat l'estúpid, ja no hi ha marxa enrere. Sense adonar-se'n, l'estúpid s'aboca a un cercle viciós de falta d'empatia, frustració, victimisme, autoodi i



enveja, factors aquests que augmenten encara més la probabilitat de cometre nous actes estúpids de manera sistemàtica i creixent. La irracionalitat es torna norma i la reacció del subjecte, caracteritzada pel victimisme, és «Oh, és que els altres no m'entenen!»

És impossible analitzar la gènesi i motivació concreta de cada acte d'estupidesa. De vegades l'acte resulta d'una pura juxtaposició aleatòria de components, ajudada per la irreflexió o la irresponsabilitat. D'exemples n'hi ha tants com vulgueu: «Oh, he atropellat algú? Si jo només volia veure quin soroll feia el motor a 6.000 revolucions!». La incapacitat de preveure les conseqüències dels propis actes, juntament amb la irreflexió i la irresponsabilitat, són les causes habituals de la conducta estúpida. Ara bé, hi ha mecanismes generadors d'estupidesa?

Jo tinc, potser per ceba personal, un possible candidat. És un mecanisme descobert als anys cinquanta que, aplicat al nostre problema, potser ens permetria delimitar –una mica– les causes d'algunes estupideses concretes, i preveure-les. Es tracta del conjunt de fenòmens que s'ha anomenat, per raons històriques, dissonància cognitiva (DC) i que jo rebatejaria, més descriptivament, així: *eliminació mental de contradiccions* (EMC).

És un fenomen descrit per Leon Festinger en un llibre (*When prophecy fails*, 1956) en què relatava com ell havia observat, d'incògnit, els seguidors d'un culte que creien que la fi del món, segons els havia comunicat un OVNI, seria tal dia, i que la secta, reunida en un lloc concret, se salvaria de la destrucció del planeta. Festinger volia veure què passaria quan veiessin, tots units i expectants, que no passava res. S'indignarien, de veure's víctimes d'un engany? Havien regalat totes les seves propietats en va? Doncs no. Davant la sorpresa de l'observador, la majoria va optar per evitar la decepció –que hauria estat insuportable– de manera creativa. Van arribar tots a la conclusió que els àliens els havien concedit una segona oportunitat, de tal manera que ara, a més, se'ls permetia estendre pertot la bona nova, que era: «no deixeu que la humanitat destrueixi la terra, salveu-la i us salvareu!». L'explicació de Festinger, inspirada en Piaget, ara ja és una teoria potent (la *Cognitive dissonance theory*), que ha sofert moltes crítiques en tot aquest temps però sembla que als seus cinquanta anys aguanta bé. Té l'avantatge de poder explicar moltes conductes aparentment

dissímils sota un esquema teòric senzill i de permetre que es pugui simular, predir i experimentar amb models empírics (i també de tipus connexionista).

Què diu, la DC? Doncs que quan es presenten dues percepcions (situacions, idees, creences) contradictòries, el nostre cervell (concretament la regió de què parlarem més avall) «se sent incòmode» i malda per eliminar o reduir la contradicció, compensant-la. Això es fa mitjançant la racionalització del problema i implica el canvi d'actituds, creences o accions, o bé la justificació, la culpabilització o la negació de la realitat. «Que hem gastat una fortuna comprant un cotxe? Doncs no cal flagellar-se'n. Segur que almenys el cotxe és bo i no tindrà avaries». No es tracta de veritat o fals; és simple consolació..., però la necessitem. «Que fumem i ens diuen que això no és bo? Bah!, la probabilitat d'una mala fi és petita... I a més bé m'haig de morir d'alguna cosa!» «Que he pres una decisió que ja no es pot canviar? Doncs què t'hi jugues que al final acabarà essent la correcta».

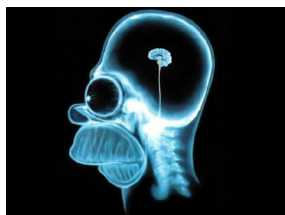
Primera conseqüència: donada la nostra tolerància zero a l'ambigüitat i a la incertesa, immediatament recorrem, davant el fet –incomodíssim– de no saber, a omplir el problema amb dades o amb hipòtesis; així ens sembla que ja en sabem alguna cosa. Necessitem fer-ho; si no, ens sentim malament. (Que les dades o les hipòtesis siguin extravagants o no tinguin cap relació amb res, no té importància). D'aquí poden sortir tantes teories conspiradores com vulgueu: una explicació «racional», encara que sigui demencial, val mil vegades més que el no saber, que és directament insuportable. (Cal recordar com els pares d'unes nenes violades d'Alcàsser van preferir refugiar-se en unes improbables *snuff movies* que en la simple i terrible realitat d'una violència banal i gratuïta? O l'angoixa insuportable de pares de fills desapareguts que, constatablement, es converteix en suportable resignació un cop es demostra que efectivament havien desaparegut i se n'han trobat els cossos?). Com la natura en Física, la ment *abhorret vacuum*. I no cal dir que l'intent d'omplir buits cognitius a base de racionalitzacions i teories dóna lloc a una munió infinita d'estupideses.

Segona conseqüència: dels moltíssims biaixos o prejudicis «cablejats» que els humans compartim, n'hi ha uns quants que, quan es posen en entredit, ens porten a fer coses que, vistes de fora, són

estúpides. Cito tot seguit quatre biaixos perillosos i universals. El primer s'expressa així: «Jo sóc una persona racional, solc prendre decisions correctes»; per tant, «Si mai faig res que els altres no vegin raonable, m'esforçaré a defensar-ho i convèncer-los, o tiraré pel dret» (... amb el risc de quedar com un estúpid). El segon biaix és «Jo sóc bona persona»; per tant, «Si mai faig res que aparentment no estigui bé, espero que els altres entendran que he tingut raons per a fer-ho» (... «I si no m'entenen, em sap greu però: que es fotin»). El tercer consisteix a valorar molt tot allò que ha requerit esforç, la qual cosa porta, si l'esforçat sóc jo, a exigir compensacions, i, en no obtenir-les o ser-me reconegudes, a la típica actitud, universal, que consisteix a creure que el món –i tothom– ens deu alguna cosa. El quart biaix o mecanisme compensatori que vull citar és aquest: «Es pensen que sóc ruc? Doncs ara veuran!». (Aquesta darrera actitud, perillosíssima, pot engegar un procés imparabile, generador de qui-sap-quantas accions estúpides). I així podríem continuar, amb més biaixos i conseqüències: tantes fonts d'estupidesa com vulgueu.

La DC permet explicar casos d'estupidesa, però, atenció!, també permet entendre altres fenòmens que no impliquen necessàriament, o gens, l'estupidesa, com ara aquests quatre: (1) confondre realitat i desig (d'això se'n diu, formalment, *predicció afectiva* i, vulgarment, somiar truites); (2) mantenir les creences tot i que la realitat hagi canviat com un mitjà (*inèrcia cognitiva*, se'n diu); (3) tenir *doubletensa* (el *doublethink* d'Orwell) o capacitat per creure dues coses contràries alhora; (4) tendir a sobrevalorar el que un mateix pot arribar a fer (i el seu dual, infravalorar el temps i recursos que calen). I també, no cal dir, (5) el fenomen de la religió, que comprèn un nucli de creences i un sofisticat joc de compensacions entre fe i realitat o entre creences principals i secundàries. Tots aquests casos de DC es desenvolupen, tal com es constata experimentalment, en un estat mental de tensió, i tenen com a subproducte freqüent l'ansietat, que la compensació cognitiva final, amb estupidesa o sense, resol.

El que a mi em crida l'atenció és que les regions del cervell que, experimentalment, els processos de DC activen, el còrtex cingulat anterior i l'insular, són les que intervenen en la compensació d'elements –possiblement desajustats– de la conducta social interactiva. És a dir, que l'EMC és, tot i que sembli un procés intern



Una imatge ràpida i sintètica del concepte d'estupidesa

de l'individu, un procés d'interacció social no explícit; vull dir que el subjecte, per un mecanisme personal automàtic, no l'explica/explicita, i que potser per això no l'entén ningú: els altres ho interpreten, en el cas més favorable, com una simpàtica peculiaritat del subjecte o, si el cas és opac, com una extravagància..., o en el punt extrem, quan no s'entén gens, com l'acte d'un estúpid. En tots aquests casos pot haver passat que el subjecte, pel fet de no haver volgut o pogut explicar-se, mostri una interrupció o dèficit en el seu procés empàtic. Vaja, que l'estúpid és bàsicament o un eixelebrat o un egocèntric frívol que va curt d'empatia. I ja es veu que, pel fet de ser un procés intern motivat per causes incontrolables, immediates i personals, l'estupidesa apareix als altres com un fenomen irracional, imprevisible i perillós (el tercer quadrant cipollià, recordeu?)

L'estupidesa, però, no deixa de ser, ben mirat, una noció relativa. L'estúpid per a un pot ser una persona perfectament normal per a algú altre. I la interfície estúpid/persona normal pot ser d'una complexitat extraordinària. Algú aparentment estúpid podria acabar sent una persona normal només que tinguéssim informació sobre els seus motius o circumstància (per exemple, un boig cridant pel carrer pot resultar ser algú que parla per telèfon). I viceversa: n'hi ha prou a recordar la cèlebre frase de Sacha Guitry («Si els qui parlen malament de mi sabessin què penso d'ells, encara en parlarien pitjor») per adonar-se'n que una persona considerada fins ara normal pot passar, amb un simple augment d'informació sobre el que li va per dins, a semblar-nos un perfecte estúpid.

Per entrar en la ment d'algú que acaba de cometre una estupidesa s'ha de veure si el subjecte té prou capacitat de raciocini. I, si la persona es comporta racionalment, intentar penetrar en el joc de biaixos i compensacions que li bullen al còrtex. Això, que no és diferent del que fan els pares quan miren d'entendre les reaccions dels infants, demana molta empatia (i potser també oxitocina, vegeu Nodes, 48)<sup>23</sup>. Doncs bé: aquestes –l'empatia i l'oxitocina (o el seus equivalents robòtics)–, també les necessitaran els robots si volem que aquests diagnostiquin bé l'estupidesa d'algú amb qui topin i hagin de tractar i conviure-hi.

<sup>23</sup> L'autor es refereix a l'article de la plana 383. (n. de l'ed.)

C) *Hi pot haver robots estúpids?* D) *Què fer amb un estúpid?*

La meua resposta a la primera qüestió és que no crec que en tinguem, de robots estúpids, si ho fem bé. En els casos d'estupidesa que són producte directe de la irreflexió, penso que la seva enorme capacitat de càlcul de les conseqüències de cada acte la podrà evitar. En els casos en què l'estupidesa es deu a no saber aturar l'acció quan ja es veu que serà estúpida (o quan, un cop comesa, no se sap explicar de manera airosa), confio que progressarem prou, en el sentit de donar empatia als nostres robots, per evitar en molts casos l'efecte final de passar per estúpids.

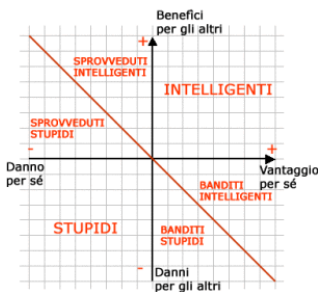
Ara bé, si em demanen com pot un robot tenir empatia –si és que és l'empatia el que necessitaran– contestaré que no en tinc ni idea. Només sé que l'empatia en els humans –que són, de fet, els únics que en tenen– es basa en la duplicació de circuits sensomotors mitjançant les neurones-mirall, de què ja he parlat mantes vegades aquí, que permeten representar en el cos propi una sensació –i preparar l'acció de resposta– pel sol fet de percebre la situació en un altre subjecte: la taràntula que puja pel cos del James Bond que veiem a la pantalla ens fa esgarrifar com si pugés pel nostre, l'insult o el menyspreu que veiem que pateix un personatge ens fa reaccionar com si ens l'haguessin adreçat a nosaltres. Perquè un robot «comprengui» una situació i hi reaccioni, cal que el mecanisme s'assembli al nostre? Cal que sigui «empatia»? Potser sí, però alerta: els psicòpates fan servir els seus circuits-mirall per preveure la conducta de les seves víctimes i aprofitar-se'n, sense que aquí es pugui parlar d'empatia en el sentit etimològic, de «patir junts» o «posar-se en el lloc d'altri». Els robots només simularan situacions i accions? O bé aniran més lluny, i «patiran» amb l'altre? Què és, patir? I si els falta sentiment, què és, «sentir», per a un robot? Com es programa, això? O potser descobrirem una manera igualment eficaç (o més) per «entendre» situacions i conductes? Tot són preguntes. I les respostes són obertes. Però jo sóc optimista. Alguna cosa trobarem.

I com ha de tractar un estúpid, un robot? Per cert: com ho fem, els humans?... Difícil, això. Com passa amb altres situacions delicades –com l'assetjament, el *bullying* o el *mobbing*– cada cas

és un món. En general, mantenir la tensió sol formar part de l'estratègia de l'estúpid, com en els tres processos citats. Doncs: o s'obvia totalment el problema, si es pot, o bé s'hi ha d'anar amb peus de plom, provant estratègies, rectificanc si cal, adaptant-se a cada moviment de l'altre... A més, hi ha el perill que el causant del problema s'acomodi a la nostra estratègia i l'exploti, com de vegades hi ha qui explota la no-agressivitat de l'altre per aprofitar-se'n. I hi ha el perill, si s'accepta l'enfrontament, d'engegar tota una situació a fer punyetes. Potser haurem de programar els robots amb cursos sencers de psicologia, perquè no se sorprenguin de res i superin amb èxit raonable la topada amb un estúpid i en minimitzin les conseqüències..., i perquè ningú no se n'aprofiti, dels robots benèvolts, pobres.

Si, com s'ha dit alguna vegada, amb els ordinadors i la IA hem recapitulat en pocs anys tota la Filosofia feta (sense màquines) en quatre segles, des de Descartes fins a Dennett; amb els robots ens veurem obligats, sembla, a inventar-nos una «Neurologia» robòtica (amb neurones-mirall, potser, o el seu equivalent? i amb les neurones fusiformes del còrtex cingulat i insular perquè processin conflictes i informació social?), i també haurem de refer tota la Psicologia: replantejar-nos-la tota sencera, aprendre-la de nou i ensenyar-los-la, a les màquines, amb la paciència d'un assistent social o d'un professor de secundària.

De moment, tot són preguntes, i tot és obert. Ja ens podríem donar prou per satisfets només que poguéssim alterar una mica, amb robots o sense, el decebedor fenomen descrit per Cipolla<sup>24</sup> sobre l'estupidesa humana i la seva constatada invariància en la Història.



Els quadrants originals de Cipolla

<sup>24</sup> Cipolla afinava l'anàlisi dividint el segon i quart quadrants en semiquadrants: així, els desgraciats són «desgraciats intel·ligents» quan fan més bé als altres que mal a ells mateixos, i «desgraciats estúpids» quan en surten més perjudicats ells que beneficiats els altres. Paral·lelament, al quart quadrant, els «malparits intel·ligents» són els qui fan mal als altres però sense superar el profit que n'obtenen ells, mentre que els qui en treuen poc benefici en relació amb el mal que fan són els «malparits estúpids». La bisectriu a 135° divideix la gent grosso modo en dues meitats, la inferior-esquerra de la qual són els qui podríem anomenar genèricament «estúpids», almenys en mètode i esperit. Cipolla no va gosar dividir el tercer quadrant, que és el nucli dur dels estúpids, ja que aquests formen un grup compacte, impermeable a qualsevol anàlisi racional.



## Màquines de Turing i aturats de llarga durada

**A**QUESTA vegada em posaré *tremendo*. Generalment, tinc el costum de veure les coses des d'una perspectiva àmplia i globalment positiva. Però l'optimisme i el bon rotllo, crec, no han pas d'excloure consideracions no tan positives si bé igualment vàlides, encara que només sigui per veure l'altra cara de la lluna i saber què hi ha. Lamento que el que diré pugui sonar en algun moment catastrofista. No és la meva intenció. Al contrari, com sempre, es tracta de predir el que ens espera, o d'intentar-ho. En paraules que alguna altra vegada he citat de Filòstrat, es tracta de «ensumar el futur» que, segons ell, és el que s'espera dels científics, i fer-ho sense cap partit pres i sense excloure cap hipòtesi.

### A) *Benvingut, Sr. Watson!*

Ho recordeu, oi? Fa uns mesos un ordinador construït *ad hoc* per IBM –anomenat Watson en homenatge al fundador de la companyia– va guanyar un concurs de preguntes de la televisió americana davant de dos antagonistes humans experts (hi havien guanyat milions en anys anteriors). El programa, *Jeopardy!*, es caracteritza per formulacions indirectes que sovint només permeten trobar la contesta per mitjà de la ironia, el joc de paraules i les referències culturals. Quatre exemples a l'atzar, en tots els quals Watson va donar la resposta correcta: (1) en comptes de demanar directament qui és el detectiu inventat per Conan Doyle la qüestió proposada és «el seu barret de caçar cérvols<sup>25</sup> l'ajudava a inves-

NODES  
NÚMERO 50  
HIVERN DE 2011

<sup>25</sup> referint-se al *deerstalker hat* que duia a la versió de Hollywood) (resposta: Sherlock Holmes)



tigar»; (2) en lloc de «autor i empresari de teatre irlandès que es va empescar un vampir a l'altre cap d'Europa», la formulació és «un autor anglès (William Wilkinson) que al segle XIX va escriure un llibre sobre els principats de Valàquia i Moldàvia el va ajudar a crear l'escenari de la seva obra més famosa» (resposta: Bram Stoker); (3) «El llebrer s'hi va crear fa més de 5.000 anys, en aquest país africà, per caçar gaseles» (resposta: Egipte) o (4) «Un actor de sèries de TV, a mig camí, es torna francès i existencial amb la seva 'caiguda'» (resposta: «Eddie Albert Camus»). Ja es veu que la cosa va molt més enllà de l'experiment anterior d'IBM, el Deep Thought del 1997, especialitzat a recórrer mecànicament enormes espais de jugades d'escacs abans que Kaspàrov li guanyés la partida. Aquí, sense connexió a Internet ni enlloc, Watson havia de respondre a tot a partir del seu coneixement enciclopèdic propi, desat en l'equivalent de 200 milions de pàgines.

Si us hi fixeu, això està a anys llum dels sistemes de *question-answering* clàssics, consistents a triar una paraula clau i enganxar-hi una frase precuinada en forma de resposta. Aquí, a més de resoldre referències semàntiques i culturals subtils, Watson es proposa a si mateix valorar el grau de confiança que li mereix cada possible resposta i, si té diverses alternatives i cap no el convenç, emet un «no n'estic segur» –en va emetre molts durant el programa– i així evita perdre diners en respostes sobre les quals no se sent prou sòlid. Un exemple: preguntat per un insecte que feia de gàngster anava a contestar «James Cagney», però no s'hi va atrevir (la resposta correcta era Bugsy Siegel, el creador de Las Vegas). O, en una badada inexplicable per a un ordinador amb teclat, no es va veure amb cor de contestar la pregunta «Aquesta abreviatura d'un gran premi d'automobilisme també està entre les tecles dels ordinadors». Quan es van veure els temptejos que havia fet –sense gosar revelar– resultava que havia aventurat «GP» i «NASCAR» i que no havia caigut en la tecla funcional «F1», ubiqua, que n'era la resposta correcta.

Watson no sols va deixar impressionats els espectadors i els seus contrincants –potser també ho hauria quedat **Turing** i li hauria aprovat el **test**– sinó també IBM, que ja l'està reconvertint en un contestador de preguntes mèdiques (o jurídiques) difusament o confusament plantejades. Un metge o advocat prou espavilat o

prou pessimista hi diria de seguida «ja hem begut oli, els metges i advocats!». I qui diu preguntes mèdiques i jurídiques, contestades a partir d'un magma impossible d'historials i jurisprudència, diu preguntes d'un contribuent desorientat sobre quins impostos li toca pagar o si hi ha cap trampa per on filtrar-se aprofitant l'impenetrable teranyina de lleis i disposicions diverses. I qui diu això, diu un possible «google» intel·ligent que, a partir d'un indici vague que li donem i del que ell dedueixi de nosaltres i del context, ens adreci directament a les respostes rellevants i certes –ell mateix comprovant-ne la fiabilitat– sense fer-nos perdre el temps entre piles de brossa i milions de *hits*.

Aquí a sota (apartat C) provo de formular molt sumàriament la qüestió de si els Watsons són una amenaça o no per als professionals qualificats i *tutti quanti*, sobretot després de veure l'espectacle de l'atur aparentment imparabile que patim a Occident i, en particular, a Europa. Però abans em detindré un moment en el més enllà.

### B) Sobre singularitats, i més enllà

No, no penso entrar en la “singularitat” que tant preocupa Ray Kurzweil i que ara ell situa cap al 2050. Ja ho sabeu, oi? Aquell punt, al voltant dels 200 milions de MIPS de velocitat de càlcul, en què els ordinadors ens hauran empaitat el cervell i la humanitat sencera ja haurà begut oli. En aquest moment en què escric, Kurzweil ha trencat amarres, ha transcendit la seva pròpia singularitat i ara es passeja pel «transhumanisme»: Kurzweil creu fermament que la nostra única salvació és que ens facin una càrrega (*upload*) del cervell tot sencer en una màquina i es planteja si, un cop l'hagin carregat a ell, «hi serà tot» o no; el preocupa que «no s'hi trobi» ben bé com és ara o, pitjor, que el resultat sigui ell però en forma de zombi. Això l'amoïna molt, perquè un cop fet l'*upload* la cosa no té retorn: passarem a ser definitivament immortals. (Vull dir, per si la cosa no s'ha entès, que no ens morirem mai, congelats en l'estat *uploaded* en què hàgim quedat: d'aquí la importància de no quedar fet un zombi: seria, mai més ben dit, definitiu.)

No me l'he acabat de creure mai, en Kurzweil, i menys en aquesta fase mística que sembla que està passant últimament. I, a

sobre, la immortalitat no em fa fred ni calor. Em ve al cap allò de Woody Allen: «La immortalitat és molt avorrida. Sobretot cap al final».

La qual cosa no vol dir que l'horitzó d'una singularitat en forma de *sorpasso* no sigui plausible. El problema és que no ens la sabem imaginar. I, a més, quants de nosaltres no estarem alimentant xiprers el 2050 i escaig en què la cosa està prevista que passi? Com deia Keynes quan li demanaven «i a llarg termini, què?» i contestava: «a llarg termini serem tots morts i enterrats», i per tant no té interès ni versemblança plantejar-s'ho. I no cal dir que, amb els temps que corren, quaranta anys (endavant) és un termini llarguíssim; i, si no, penseu què fèiem i pensàvem ara fa quaranta anys (enrere). Doncs què serà al 2050, ara que tot va –i anirà– molt més ràpid?

Com a nota exculpatòria per la meua falta de fe en les idees dels Kurzweil, Moravec i altres immortals, tinc el consol de pensar que això de preveure singularitats no és tan desenraonat ni nou com pugui semblar. De fet, la idea –sense extrapolacions gratuïtes– és tan raonable que la va pensar primer –qui, si no?– John von Neumann abans de morir-se als 1950, i que la va desenvolupar I. J. Good als 1970. (Per si no ho recordeu, Good va ser l'assessor informàtic de Kubrick al seu *2001: A Space Odyssey* del 1968, al mateix temps que Minsky li feia la consultoria en IA.)

Si m'he referit al “singularitarianisme” kurzweilià només ha estat per delimitar futurs i per clarificar que l'horitzó de futur a què em referiré en la resta de l'article no és pas aquest, sinó molt més pròxim: el d'ara mateix, el de la crisi econòmica. I, feta la reverència obligada al geni de von Neumann, torno al tema base d'aquest article: la crisi i l'atur.

### *C) I si tot plegat fos culpa nostra (de la IA, vull dir)?*

Sobre aquesta pregunta aparentment absurda i egocèntrica (o gremial) haig de dir que el primer senyal d'atenció me'l va donar una dada que vaig llegir en revistes d'economia fa uns mesos que palesava que no sols la desocupació no baixa enlloc encara que l'economia es recuperi sinó que ja no es pot continuar dient allò tan trist que havíem anat repetint: que l'atur afecta la part de la

població amb pocs estudis. Doncs ja no: la informatització de molts serveis «de qualitat» ja afecta la classe mitjana que sempre se n'havia sentit protegida. No cal pensar en Watson i els seus successors futurs, segurament sobrerevolucionats. Ara mateix, només observant el vostre entorn immediat, em podríeu dir quants metges s'han quedat sense feina o emigren perquè la feina que feien o haurien pogut fer la fan unes màquines que els han substituït a l'hora d'interpretar anàlisis o radiografies, redactar o confrontar informes, planificar fases d'acció i resoldre possibles conflictes?; i, quants metges hi hauria ara si haguéssim continuat amb el (relativament rudimentari) grau de mecanització hospitalària de només fa cinc o deu anys? (En tinc experiència directa.) I això val també per als serveis de cara al públic, com ara una oficina bancària o un despatx de dentista. O a la mateixa oficina de l'atur: us heu fixat quantes pantalles hi ha perquè uns pocs empleats ens diguin de cos present allò que la màquina sap i els ha dit (i que potser ens podria haver explicat per Internet directament a casa nostra, sense intermediaris)?

Sempre s'ha escrit, i jo ho he repetit sovint aquí mateix, que no ens hem d'espantar pel progrés científic i tecnològic, perquè les desocupacions brutals que històricament ha provocat (a l'agricultura o a la indústria) s'han compensat en un altre lloc (a la ciutat, a la indústria innovadora, la recerca, els serveis creatius, etc.). És a dir, que la demanda ha anat augmentant al mateix ritme, aproximadament, que creixia l'oferta dels nous productes tecnològics. Però això potser passava perquè tant la demanda com la productivitat augmentaven linealment, i estaven a la part baixa, suaument creixent, d'una corba possiblement exponencial. I potser ara la demanda continua lineal mentre els guanys de productivitat estan desbocats en la zona exponencial.

Si fins ara el progrés tecnològic ens havia donat màquines que eren, primer, amplificadors de força i, més tard, assistents en el treball humà, ara potser el problema és que els ginys que fem ja no són assistents sinó que són prou intel·ligents per fer, ells mateixos, de treballadors i que, per tant, ja no ens ajuden sinó que ens substitueixen. I el problema és: què faran els treballadors (humans) desplaçats? on aniran? Alguns, és clar, es podran refugiar –de moment– en feines manuals que encara no estiguin a la mà de les

màquines, com ara la collita o preparació d'aliments o la neteja de vàters o la cura d'ancians, feines, per cert, típiques d'immigrants que els nadius no estan disposat a fer. Uns altres es podran refugiar en feines creatives i tasques intel·lectuals no repetitives ni predictibles, però aquests no seran prou en nombre per cobrir els llocs de treball que hauran quedat anul·lats ni per fer pujar la productivitat, i és fàcil augurar-los una vida difícil, sempre a l'avantguarda i per davant del que puguin fer els ordinadors, i en dura competència internacional: parlo dels emprenedors, dels dissenyadors (d'aplicacions informàtiques, de moda, art, etc.) i dels creatius en general.

I la resta de la població? Cada vegada sembla més probable que no sols no hi haurà feina (en el sentit de treball humà digne) per a tothom sinó que cada vegada ens haurem de preparar més per donar a la població general de totes les edats una orientació vital molt semblant a la que adquireixen els nostres vells davant la perspectiva de no treballar més a la vida. Si la cosa va per aquí, ja cal que ens hi anem preparant des d'ara. I ja es veu que el daltabaix mental, cultural i d'estil de vida que se'ns obre és espantós.

El problema –ja es veu– sembla que no és que estiguem estancats econòmicament perquè no creem prou progrés tecnològic sinó, al contrari, perquè n'estem creant massa, i ara ja ens comença a superar per totes bandes i fa obsoleta la societat i les solucions que fins ara ens han funcionat. I el progrés que creem ja no és per fer tasques rutinàries sinó per fer-ne, també, de cognitives i fins i tot de creatives. I sembla que devem ser a prop del punt d'inflexió (el *tipping point* de Gladwell) en què el poder «mental» de les màquines que tenim, gràcies per exemple a la seva capacitat de reconèixer formes (plaques, informes o el que sigui), amenaça de suplantat el de gran part dels professionals ben pagats i el de les nostres classe mitjanes. Un metge o un advocat normal ja avui té l'agressiva competència de sistemes automatitzats que poden, com ell, investigar en la història els casos pertinents aplicables, avaluar situacions, resumir-les bé i descobrir solucions i aplicar-les flexiblement. En molts aspectes, els empleats de coll blanc que poblen les nostres oficines, bancs, botigues o laboratoris, així com la majoria dels professionals liberals (metges, advocats, assessors, enginyers, etc.) d'ingressos mitjans, estan en la mateixa situació, o semblant,

en què es trobaven fa un segle els pagesos que estaven sent desplaçats pels tractors o els obrers fabrils esbandits per la sofisticació i l'automatització de la maquinària, amb la diferència que ara ja no hi ha, respectivament, «ciutats» o «oficines» on els desplaçats puguin refugiar-se i començar una nova vida productiva.

Això, si fa no fa, ja ho va imaginar profèticament Jeremy Rifkin el 1995 (a *The End of Work*). Deia Rifkin –i no era pas el primer a dir-ho– que els ordinadors, de programari cada vegada més sofisticat, portarien la nostra civilització a ser un món sense treball. Ara, una setzena d'anys més tard, podem assegurar que el procés ja ha començat. Martin Ford opinava el 2009 (a *The Lights in the Tunnel*) que les noves feines creades per la novíssima tecnologia (venedors de mòbils, programadors de servidors d'Internet o físics del CERN) no arribaran a una mínima fracció de la població..., i encara, perquè la història sembla demostrar que aquestes feines es volatilitzen fàcilment. Ford calcula que un 40% de la població activa americana podria ser substituïda ara mateix, en major o menor grau, per algun programa d'ordinador, i que probablement d'aquí deu anys ja no tindrà feina. I afegeix Ford que la deslocalització de les tasques informàtiques actuals (pròpies dels 1990, per cert) a l'Índia o on sigui només resol el problema provisionalment, fins que els sous d'allà siguin prou insuportables per fer la deslocalització inviable. *A Race Against the Machine*, Erik Brynjolfsson i Andrew McAfee, de l'MIT, donen la raó a Ford: els llocs de treball perduts des del començament de la crisi és molt poc probable que tornin.

A aquesta presumpta tendència social s'afegeix el fenomen, d'aparença merament econòmica, que vivim a Europa a tombant de 2012 i que consisteix, en paraules d'Obama, que la classe mitjana (europea) ja no és capaç de comprar els béns que els seus països produeixen. (La conclusió local immediata és aquesta: si només retallem no quedarà res, ni tan sols per comprar el que fabriquem; i viceversa: si ho volem comprar tot ens haurem d'endeutar... o bé exportar –a qui?– tant com puguem.) Aquest fenomen coincideix amb el descrit per Ford quan afirma que, a l'Occident en general, la tecnologia està creant riquesa a un ritme tal que està clarament fora de l'abast del potencial de consum de la població. L'article de *The Economist* del 4 de novembre de 2011 titulat *Luddite legacy*, és un excel·lent resum de tot això que comento. I fa només quatre

dies que Zygmunt Bauman ens deia a la Pedrera que no és que tinguem incertesa sobre el resultat del partit, és que han canviat les regles del joc mentre jugàvem, i encara no tenim ni idea de quines són les noves.

Perquè, dic jo, després d'haver passat deu mil anys de civilització basada en el concepte i la pràctica del treball, com transitar ara de cop cap una problemàtica civilització de lleure i benestar? Com educar-nos per al primer? com finançar el segon (sense treballar)? com redistribuir la riquesa que presumptament haurà creat el «treball» de les nostres màquines entre els seus usufructuaris humans? Ja es veu que som davant d'un problema insòlit i apassionant. Entrem en una cova sobre la qual ningú ens ha donat indicacions ni plànols.

És clar que, després d'aquesta tirallonga de seriositats inquietants, jo ara m'hauria de veure obligat, per ser correctament optimista i fer veure que els camins del futur són infinits, a esmentar per exemple els casos d'Amazon i d'eBay, que han creat ells sols 600.000 llocs de treball de gent que en viu (a base d'imaginar productes per a aquests dos entorns). I a recordar que els humans tenim, sobre les màquines, les capacitats d'imaginar, sentir, aprendre, crear, improvisar, intuir i actuar espontàniament. (De moment.) I a reblar-ho dient que a les màquines, per molt intel·ligents que siguin, els és difícilíssim de copiar-nos bé. (També de moment... si més no fins que arribi la temible singularitat i ens agafi confessats). I a citar per exemple Marina Gorbis, de l'Institute for the Future, de Palo Alto, que, com sol ser habitual en els comentaristes benèvolos (que, a diferència de Nicholas Carr amb l'Internet –v. Nodes 48–, no volen que ens espantem)<sup>26</sup>, diu coses com aquestes: «Potser sí que les màquines ens substituiran, però ho faran per 'amplificar-nos', per permetre'ns fer coses que mai no hauríem somiat poder fer» i també «La nova simbiosi home-màquina ens restituirà la dignitat del treball, i farà del món un lloc diferent». Amén

<sup>26</sup> L'autor es refereix a l'article de la plana 383. (n. de l'ed.)