

COMPRENDRE

revista catalana de filosofia

Director

Armando Pego Puigbó
Universitat Ramon Llull

Secretari de Redacció

Joan Cabó
Universitat Ramon Llull

Consell de redacció

Mauricio Beuchot
Universidad Nacional Autónoma de México
Miguel García-Baró
Universidad Pontificia de Comillas
Alejandro Llano
Universidad de Navarra
Carles Llinàs
Universitat Ramon Llull
Ignasi Roviró
Universitat Ramon Llull
Joan Albert Vicens
Universitat Ramon Llull
Conrad Vilanou
Universitat de Barcelona

Consell científic

Gabriel Amengual
Universitat de les Illes Balears
Antoni Bosch Veciana
Universitat Ramon Llull
Adela Cortina
Universitat de València
Raúl Forner-Betancourt
Universität Aachen
Giuseppe Di Giacomo
Università La Sapienza
Francisco J. González
University of Ottawa
Pierre Hadot †
Collège de France
Silvia Magnavacca
Universidad de Buenos Aires
Ulises Moulines
Universität München
Giovanni Reale †
Università San Raffaele
Thomas Robinson
University of Toronto
Salvi Turró
Universitat de Barcelona

Comprendre. Revista catalana de filosofia. Coeditada per Herder Editorial i Facultat de Filosofia de la Universitat Ramon Llull. Els originals per sotmetre a consideració del Consell de redacció cal enviar-los a:

Comprendre. Revista catalana de filosofia
Secretaria de la Facultat de Filosofia
Universitat Ramon Llull
Diputació, 231 - 08007 Barcelona
Tel. 934534338 - Fax 934530957
comprendre@filosofia.url.edu
<http://www.filosofia.url.edu>

Per a subscripcions i comandes

Herder Editorial
Tel. 934762640 - Fax 932073448
revista@herdereditorial.com
<http://www.herdereditorial.com>

Preu exemplar: 12,50 € (IVA inclòs)

Preu de subscripció: 20 €/any (IVA inclòs)

Periodicitat semestral

COMPREDRE està indexada a ERIH Plus, IBZ (Internationale Bibliographie der geistes-und sozialwissenschaftlichen Zeitschriftenliteratur), IBR (Internationale Bibliographie der geistes-und sozialwissenschaftlichen Literatur), ISOC (C.S.I.C.), Latindex (UNAM, Mèxic), Philosopher's Index, Répertoire bibliographique de la philosophie. COMPREDRE ha estat seleccionada per Elsevier a fi de ser indexada a SCOPUS des de desembre de 2015

Maquetació: Fotoletra, SA

Coberta: Michel Tofahrn
Impressió: Fotoletra, SA
Dipòsit legal: B-31.512-2012
ISSN: 1139-9759



COMPRENDRE

revista catalana de filosofia

Vol. 23/1

Any 2021

Editorial

5

Armando Pego Puigbó i Francisco Batlle

Articles / Articles

Una revisión metateórica a la operatividad de los reduccionismos metodológicos, epistémicos y ontológicos en la investigación psicoclínica en torno a la eficacia psicoterapéutica de las prácticas espirituales 9
A meta-theoretical review regarding the operability of methodological, epistemic and ontological reductionisms in psychoclinical research around the psychotherapeutic efficacy of the spiritual practices
María Teresa Gargiulo

La monstruosidad del otro. Slavoj Žižek y la crítica a las concepciones occidentales de la otredad 27
The monstrosity of the Other. Slavoj Žižek and the criticism of Western conceptions of otherness
Jorge León Casero

Dossier: Filosofia i Intel·ligència Artificial

Poden pensar les màquines? Quatre tesis sobre la intel·ligència artificial 49
Can the Machines Think? Four thesis on Artificial Intelligence
Joan Albert Vicens

Inteligencia artificial: aproximación desde una teología amiga de la ciencia 77
Artificial Intelligence: approach from a science-friendly theology
Juan Ramón La Parra

A humanist perspective on artificial intelligence 99
Jaume Agustí-Cullell i Marco Schorlemmer

Ressenyes / Reviews

Yuval Noah Harari, 21 lecciones para el siglo XXI 127
Ignasi Fuster Camp

Joan Vergés Gifra (ed.), <i>Balmes, quin llegat, avui?</i>	131
Ramon Rosales Sebastià	
Ramon M. Nogués, <i>La salut espiritual</i>	134
Jaume Duran	
Antonio Bergamo, <i>Levinas e la curvatura dello spazio intersoggettivo</i>	138
Joan Cabó Rodríguez	

Llibres rebuts / Books received	143
--	------------

Índex anterior / Previous Index	151
--	------------

Normes de publicació / Guideline for contributors	153
--	------------

Armando PEGO PUIGBÓ i Francisco BATLLE

Saber i Comprendre

Les enormes dificultats socials, econòmiques, polítiques i humanes que hem hagut de patir durant aquest últim any, arran de la pandèmia del virus covid-19, han tornat a demostrar que la interrelació global de les societats actuals provoca també tot un seguit de desafiaments que colpegen directament les respostes que ha de ser capaç d'articular l'àmbit de la Filosofia i les Humanitats.

Una veritable recerca no podrà donar respostes immediates, com sembla que un món interconnectat gairebé exigeix com a condició d'eficàcia tècnica. A canvi, hauria d'oferir unes reflexions actuals i punyents que toquessin les arrels de la nostra condició humana. En certa mesura, continua en peu aquell compromís moral de la filosofia com a consolació que Boeci va resumir per a la filosofia occidental, l'abast i l'operativitat del qual la situació actual continua posant en *crisi*.

COMPRENDRE sempre ha volgut situar-se, amb humilitat i amb decisió, en aquesta cruïlla intel·lectual aportant un espai de diàleg des dels marges, fins i tot d'una manera indirecta, que pugui servir per obrir punts de vista que copsin alhora els matisos de més llarg recorregut. Les noves circumstàncies han renovat l'esforç per mantenir-nos fidels a aquest objectiu fundacional de la nostra revista.

Esperem que el número que ara ens complau de presentar davant dels nostres lectors hagi aconseguit tant reflectir aquesta continuïtat d'un ideari clàssic dels estudis humanístics, adreçat pel seu rigor tant a la comunitat acadèmica com a un públic universitari ampli, com alhora refermar la nostra atenció als fenòmens més actuals i a les conseqüències que s'hi plantegen.

Dos articles inicials exposen qüestions rellevants de les tensions i la creativitat originades per l'apropament transversal entre les ciències socials i les humanístiques. En l'article «Una revisión metateórica respecto a la operatividad de límites metodológicos, epistémicos y ontológicos en la investigación psicoclínica», María Teresa Gargiulo proposa un apropament als valors psicoterapèutics de l'experiència espiritual. Davant de reduccionismes de tota mena, l'autora planteja una revisió metateòrica que permeti advertir la dinàmica pròpia de l'espiritualitat i les tradicions religioses en la constitució de la personalitat humana, evitant d'exercir-hi una limitació operativa a l'estructura simplement psicològica.

D'altra banda, a «La monstruosidad del Otro», Jorge León Casero s'apropa al concepte de l'Altre en el pensament de Slavoj Žižek. Davant del multiculturalisme liberal i de l'ètica de l'alteritat levinasiana, tal com és interpretada per Jacques Derrida, Žižek

prefereix adoptar una posició lacaniana per tal de fer front a la qüestió de l'autoritat simbòlica en el marc de les estratègies de subversió dels marcs polítics actuals.

Intel·ligència (Artificial) i Filosofia

Els articles aplegats a continuació apunten també al mateix objectiu de col·laborar amb altres disciplines properes, siguin de tipus social o tecnocientífic, la qual cosa constitueix l'eix d'una de les línies de recerca que el SGR «Filosofia i Cultura» de la Facultat de Filosofia (URL) desitja explorar i compartir. Ens ha semblat oportú donar cabuda a un dossier sobre Filosofia i Intel·ligència Artificial des dels reptes amb què els avenços tècnics qüestionen l'horitzó d'expectatives de la filosofia de rerefons clàssic i de la teologia.

Queda lluny la imatge de Descartes de l'arbre de la Filosofia amb la Metafísica com les seves arrels i la Física com el seu tronc. El símil actual és més aviat un moviment centrífug mitjançant el qual les diverses ciències es van independitzant de la Filosofia a mesura que consideren que adquireixen una entitat suficient. Tanmateix, hi ha situacions extremes que requereixen centralitzar el coneixement: la irrupció de la intel·ligència artificial és una d'aquestes situacions.

La intel·ligència artificial, antigament només un bonic oxímoron, apareix actualment en els mitjans de comunicació cada dia: conducció autònoma, algorismes que influeixen en resultats electorals o que decideixen per nosaltres, robots per als quals es reclamen drets... Darrere de tots aquests avenços tornen preguntes fonamentals: pensen les màquines?, què és pensar?, què és conèixer o tenir sentiments? Cal, per tant, analitzar aquest progrés científic des d'una perspectiva humanista. Cal tornar a la filosofia.

Odo Marquard comparava la figura del filòsof actual amb el «doble» que es fa servir en les escenes arriscades. Ha de ser una persona capaç de gestionar dissensions i divergències seculares que li permetin debatre amb els especialistes en les zones de risc i de frontera del saber. D'ací que els dies 27 i 28 de novembre de 2019, a la seu de la Facultat de Filosofia de Catalunya (URL), s'organitzessin unes «Jornades de Filosofia i Intel·ligència Artificial *Filosof-IA*». Promovent un entrecruament de perspectives científicotecnològiques i filosòfiques, pretenia, precisament, suggerir camins nous de resposta als reptes que tenim plantejats en l'àmbit de la intel·ligència artificial.

Són fruit d'aquelles jornades els tres articles que, com dèiem, s'inclouen en aquest número de la revista a tall de dossier. Joan Albert Vicens ofereix una visió antropològica a partir de la pregunta «Poden pensar les màquines?». Hi planteja una anàlisi de les quatre tesis resultants de la comparació entre la intel·ligència humana i l'artificial, amb especial èmfasi en la proposta zubiriana de la «intel·ligència sentent».

Les implicacions teològiques estan explicitades a «IA: Aproximación desde una teología amiga de la ciencia», de Juan Ramón La Parra, on s'ofereix una doble perspectiva:

des de la teologia fonamental i des de l'antropologia teològica. Conceptes teològics com ara creació, home a imatge de Deu, pecat i escatologia es posen en relació amb el discurs de la intel·ligència artificial.

Finalment, Jaume Agustí-Cullell i Marco Schorlemmer, des del coneixement profund de la tècnica, en l'article «A humanist perspective on Artificial Intelligence», donen les claus per distingir les fronteres entre un tipus d'intel·ligència i l'altre des d'un plantejament humanista-sociocultural: la «socially embodied human intelligence».

Tres articles que actuen com a tensors contra la independència de les ciències i que imposen una força centrípeta en direcció cap a una necessària visió filosòfica de la realitat.

Armando PEGO PUIGBÓ
Director de *COMPRENDRE*

Francisco BATLLE
Coordinador del Dossier “Filosofia i Intel·ligència Artificial”
Neurite Lab

UNA REVISIÓN METATEÓRICA A LA OPERATIVIDAD DE LOS REDUCCIONISMOS METODOLÓGICOS, EPISTÉMICOS Y ONTOLÓGICOS EN LA INVESTIGACIÓN PSICOCLÍNICA EN TORNO A LA EFICACIA PSICOTERAPÉUTICA DE LAS PRÁCTICAS ESPIRITUALES

María Teresa GARGIULO

Resumen

El artículo no es sino una revisión metateórica respecto a la operatividad de los reduccionismos metodológicos, epistémicos y ontológicos en la investigación psicoclínica en torno a la eficacia psicoterapéutica de las prácticas espirituales. Por reduccionismo se entiende el proceso de simplificación y operatividad que exige el mismo diseño experimental adoptado en la investigación psicoclínica. Si bien el reduccionismo que se implementa en un diseño experimental goza de toda una suerte de ventajas, también tiene sus peligros. Podría decirse que existe una compensación, a saber, las explicaciones reduccionistas ganan en simplicidad mientras pierden en precisión. El diseño experimental exige una simplificación, es decir, eliminar cosas que necesitan ser entendidas y que son parte de una imagen completa del fenómeno estudiado. Ahora bien, una revisión metateórica permite no solo explicitar el alcance que tienen los reduccionismos en los resultados de la investigación psicoclínica sino también repensar nuevas variables o modalidades que permitan fortalecer dichas investigaciones.

Palabras clave: espiritualidad; investigación psicoclínica; reduccionismo; eficacia psicoterapéutica; psicología de la religión.

A meta-theoretical review regarding the operability of methodological, epistemic and ontological reductionisms in psychoclinical research around the psychotherapeutic efficacy of the spiritual practices

Abstract

The article is nothing but a meta-theoretical review regarding the operativeness of methodological, epistemic and ontological reductionisms in psycho-clinical research regarding the psychotherapeutic efficacy of spiritual practices. By reductionism is

meant the process of simplification and operability that requires the same experimental design adopted in psycho-clinical research. Although the reductionism that is implemented in an experimental design enjoys all kinds of advantages, it also has its dangers. It could be said that a compensation exists, namely, reductionist explanations gain in simplicity while losing in precision. The experimental design requires a simplification, that is, eliminating things that need to be understood and that are part of a complete picture of the phenomenon we are studying. However, a meta-theoretical review allows not only to explain the scope of reductionism in the results of psycho-clinical research but also to rethink new variables or modalities that allow strengthening such research.

Key words: spirituality; psycho-clinical research; reductionism; psychotherapeutic efficacy; psychology of religion.

Introducción

Nuestro estudio tiene la intención de ser parte del esfuerzo mancomunado que existe en nuestros días por dilucidar el alcance y los límites epistémicos de los diseños experimentales que buscan medir la eficacia de ciertas prácticas espirituales en la psicoterapia. La pregunta que abre nuestro estudio inquiriere acerca del alcance y el peso que poseen estos estudios de tipo estadístico. ¿En qué medida estos análisis empíricos pueden erigirse como materia rectora y orientadora de la práctica psicoterapéutica? ¿Tales estudios miden efectivamente el fenómeno que reconocemos bajo el nombre de espiritualidad?

Estos interrogantes revelan la necesidad de integrar en las ciencias de la salud mental una instancia capaz de analizar críticamente la calidad de dichas investigaciones; no solo en lo tocante a las cuestiones metodológicas, sino también respecto a las simplificaciones epistémicas y ontológicas que ellas suponen. En esta línea disponemos de una serie de estudios pertinentes que revelan particularmente los reduccionismos que operan en las diversas modalidades de estudios de las ciencias de la salud mental: Murphy (2005); Zizioulas (2006); Taylor (2007); Smith (2000); Cooper (2007); Andler (2006); Barrett, Dunbar y Lycett (2002); Bennett y Hacker (2003); Kim (2006); Kistler (2006); Poirier (2006); Schaffner (2006); Vande Kemp (2000); Watts (2002); Slife y Hopkins (2005).

Nuestro trabajo se circunscribe al estudio de la operatividad de dichos reduccionismos en los diseños experimentales que miden la eficiencia psicoterapéutica de las prácticas espirituales. En este sentido, el valor del presente trabajo no radica tanto en la originalidad o novedad de su tesis sino, más bien, en la sistematización o precisión con

la que busca identificar los reduccionismos metodológicos, epistémicos y ontológicos de la investigación psicoclínica respecto a la espiritualidad.

El reduccionismo metodológico no es sino el proceso de simplificación y operatividad que exige el mismo diseño experimental adoptado por la investigación psicoclínica. Si bien el reduccionismo que se implementa en un diseño experimental goza de toda una suerte de ventajas, también tiene sus peligros. Podría decirse que existe una compensación, a saber, las explicaciones reduccionistas ganan en simplicidad mientras pierden en precisión. El diseño experimental exige una simplificación, es decir, eliminar cosas que necesitan ser entendidas y que son parte de una imagen completa del fenómeno que estudiamos. En este sentido, los modelos explicativos de esos diseños pueden ofrecer explicaciones excesivamente incompletas e incluso erróneas respecto al fenómeno que pretenden medir.

A modo de hipótesis, podemos consignar que la adopción acrítica de fuertes posiciones reduccionistas ha afectado en gran medida el diálogo entre ciencia y religión, particularmente dentro del campo de la investigación psicoclínica. Nuestra intención es señalar estos reduccionismos o recortes, pues en la medida en que sea explícita su operatividad será posible no solo medir el alcance que tienen sobre los resultados de los estudios sino también repensar nuevas variables o modalidades de fortalecer estas investigaciones.

En orden a ser autoconsistentes en nuestro planteamiento, estructuraremos nuestro trabajo en tres partes. Primero presentaremos los diversos reduccionismos metodológicos básicos de los diseños experimentales psicoclínicos sobre espiritualidad y señalaremos cómo estos pueden comportarse como amenazas de validación de tales diseños. En un segundo punto, analizaremos en un nivel metateórico las diversas modalidades de reduccionismo epistémico que suponen tales investigaciones. Finalmente, presentaremos los reduccionismos ontológicos que habitualmente se adoptan en los estudios psicoclínicos en torno a la eficiencia psicoterapéutica de las prácticas espirituales.

1. Reduccionismo metodológico

En los últimos años ha crecido exponencialmente el cuerpo de literatura que presenta la espiritualidad como un área importante no solo de la evaluación diagnóstica sino incluso de la propia instancia terapéutica. Probablemente el resurgimiento del interés por esta temática se explique a propósito de la inclusión en el DSM-IV (2000) de la categoría de «Problema Religioso o Espiritual» Z71.7 (V62.89). Esta nueva categoría buscaría contemplar y atender a las dificultades o el malestar psíquico que puede suponer la pérdida o el cuestionamiento de la fe, los problemas asociados con la conversión a una nueva fe o el cuestionamiento de los valores espirituales. Dicha incorporación parece recoger una extensa e histórica discusión acerca de si la espiritualidad es un área importante que desata o paraliza dinámismos psíquicos que son relevantes para el de-

sarrollo de la personalidad.¹ Dentro de una amplia variedad de posiciones, hoy parece estar fuera de discusión que un determinado tipo de espiritualidad puede ser causa de depresión, manía, masoquismo, histeria, fobias o comportamiento obsesivo. O, por el contrario, puede suceder que ciertas prácticas religiosas se traduzcan en hábitos o estructuras psíquicas saludables. En este último caso, la religión y la vida espiritual podrían presentarse como factores protectores y promotores de la salud mental, muy especialmente respecto de trastornos como la depresión, la ansiedad y el abuso de sustancias.

Pues bien, para responder a la pregunta acerca de la eficacia terapéutica de las prácticas espirituales —así como de los efectos de cualquier forma o práctica psicoterapéutica respecto a un trastorno particular—, la psicología ha tomado prestados de la medicina sus diseños de investigación. Adopta diseños metodológicos que originalmente fueron formulados y aprobados por la Administración de Alimentos y Medicamentos (FDA) de los Estados Unidos. Nos referimos particularmente al diseño aleatorio, que fue desarrollado entre 1920 y 1930, y al diseño de grupo de control con placebo, desarrollado en la década de 1950.

El énfasis que en la década de 1990 se le dio a la necesidad de estandarizar los tratamientos médicos y de proporcionar evidencia respecto a la eficacia de cada tratamiento específico, incluso de aquellos que estaban dirigidos a los trastornos mentales, llevó a un grupo de trabajo de la División 12 de la Asociación Americana de Psicología (APA) a la pretensión de trabajar en esta misma línea metodológica: «Si la psicología clínica ha de sobrevivir en este apogeo de la psiquiatría biológica, la APA debe actuar para enfatizar la fuerza de lo que somos» [...] «tienen que ofrecer una variedad de psicoterapias de eficacia probada» (*Task Force on Promotion and Dissemination of Psychological Procedures*, 1995, p. 3).

La adopción de tales esquemas metodológicos de investigación por parte de la APA exigía reemplazar la explicación biológica por una explicación psicológica, es decir, discernir en el grupo experimental mecanismos de cambio y acciones terapéuticas de naturaleza psicológica. Un tratamiento psicoterapéutico específico estaría validado empíricamente siempre que por lo menos dos estudios demostraran la superioridad de una acción terapéutica sobre un grupo experimental con un trastorno o problema específico frente a los grupos de control o tratamientos alternativos (tratados con placebo).

No obstante, al mismo tiempo que estos diseños eran implementados en la investigación psicoclínica, la misma APA indicaba la necesidad de advertir a los psicoterapeutas de las limitaciones y los recortes de tales diseños: «la práctica basada en la evidencia

¹ La espiritualidad no solo es entendida como un aspecto de la personalidad, sino que ella en sí misma operaría como una función integradora de la personalidad (véase BECKER, 2001; MACINNES, 2003, p. 51; INGERSOLL, 1994; ROOF, 1999, p. 35). Ella configuraría la unidad interna, la autopercepción y las modalidades de las relaciones interpersonales (SCHNEIDERS, 1998; McGRATH, 2006; KOSEK, 1996; THEOPHAN, 1995, pp. 95-99; WAGENER y MALONY, 2006; WUTHNOW, 1998; WISEMAN, 2007).

requiere que los psicólogos reconozcan las fortalezas y limitaciones de la evidencia obtenida de diferentes tipos de investigación» (*APA Presidential Task Force on Evidence-Based Practice*, 2006, p. 275). En línea con esta sentencia, creemos necesario discernir los recortes metodológicos que suponen los diseños experimentales que procuran medir la eficacia psicoterapéutica de ciertas prácticas espirituales. No se trata de desprestigiar el esfuerzo metodológico de tales estudios, sino de relativizar e interpretar críticamente sus resultados.

Por reduccionismo metodológico entendemos las diversas simplificaciones fenoménicas de la espiritualidad que exigen los diseños experimentales. El diseño experimental exige seleccionar un conjunto parcial de variables, susceptibles de posibles validaciones empíricas, y descartar otros factores o aspectos importantes. Pues bien, estos mismos recortes –como veremos a continuación– pueden atentar contra la propia validez de las investigaciones empíricas.

Los diseños experimentales requieren establecer una definición operacional del problema o trastorno en el cual intervenir. Cabe aclarar que lo psíquico en sí, y particularmente lo espiritual, es inmensurable, pues no son realidades extensas y cuantificables. De ahí que estos diseños busquen medir las dinámicas psíquicas concomitantes a las prácticas espirituales a través de sus cualidades, de sus efectos o del resultado de sus operaciones. Se trata de traducir dichas dinámicas a sus exteriorizaciones fenoménicas de un modo tal que constituyan ciertas formas de percepción y posibilidades de comprobación que procedan del «objeto» mismo. Esto es, traducirlas a indicadores observables o referentes empíricos como actos de habla o conductas que manifiesten específicamente el problema a estudiar.

Estas manifestaciones fisiológicas, conductuales, lingüísticas, cognitivas o interpersonales de la vida espiritual establecen la línea de base que permite obtener información descriptiva acerca del problema. Ellas constituyen el parámetro comparativo en virtud del cual es posible evaluar el posible efecto psicoterapéutico de la espiritualidad. La determinación de una línea de base posibilita, en definitiva, no solo conocer el tenor fáctico de las alteraciones que afectan al paciente al inicio del estudio sino sobre todo poder predecir si los cambios registrados en el estudio se deben efectivamente a la terapia o si se deben a una remisión espontánea, a factores inespecíficos, a la propia maduración del paciente, etc.

Entre los estudios de tipo empírico que se publicaron en las bases de datos de MEDLINE y PsychINFO entre los años 1996-2007, Vachon, Fillion y Achille (2009) discriminan once dimensiones a las que se hace referencia bajo el término «espiritualidad»: 1) el propósito que da sentido a la vida; 2) la autotrascendencia que permite estar en armonía y en paz consigo mismo; 3) la trascendencia hacia un ser superior; 4) un sentimiento de comunión con el propio yo, con Dios, con el Universo o la naturaleza dentro de una red de relaciones interpersonales; 5) un cuerpo de creencias que son objeto de fe; 6) la esperanza entendida como una actitud frente a las dificultades de la

vida; 7) la actitud frente a la muerte; 8) el aprecio a la vida; 9) los valores personales; 10) un proceso dinámico y 11) consciente.

Atendiendo a esta variopinta referencia teórica que posee la noción de espiritualidad en la literatura existente, los diseños experimentales la traducen en múltiples y diversas variables específicas en orden a poder medir cómo estas se relacionan con otra variable como puede ser la reducción del estrés, o cualquier otra variable relacionada con la salud mental. A saber, no es la espiritualidad, en toda su complejidad y dinámica, la que cae bajo la consideración del diseño. En orden a evitar el carácter abstracto y genérico que supone la noción de espiritualidad, se la traduce a los términos de asistencia al servicio de adoración, al desempeño de prácticas religiosas como la oración, en los términos de los vínculos interpersonales que se establecen en las comunidades religiosas, etc. El carácter polifacético de la espiritualidad es reducido a una o dos variables de este tipo.

Se trata de análisis funcionales donde la religión y la espiritualidad son medidas en los términos de sus funciones para promover o deteriorar la salud física y mental; por ejemplo, en los términos de sus potencialidades para lidiar con el estrés.

Disponemos de estudios que sostienen que la participación en ritos y comunidades religiosas protege a las personas de las enfermedades y promueve la salud. Un grupo de investigaciones muestra la correlación que existe entre la frecuente asistencia a una iglesia y la reducción de las tasas de mortalidad: hay un aumento en la vida útil de las personas de hasta siete años, es decir, un beneficio para la salud equivalente al hecho de no fumar. (Levin, Chatters y Taylor, 2005; McCullough, 2001; McCullough, Hoyt, Larson, Koenig y Thoresen, 2000; Koenig, 2001; Thoresen, Harris y Oman, 2001; Mueller, Plevak y Rummans, 2001; Shahabi et al., 2002; Contrada et al., 2004; Chatters, 2000). La práctica religiosa también se asocia con tasas más bajas de depresión y de enfermedades cardiovasculares debido a que la asistencia a comunidades religiosas aparentemente reduciría la presión arterial y la hipertensión (Koenig, 1995; Krause, 2002; Steffen, Hinderliter, Blumenthal y Sherwood, 2001). Esta misma variable también está correlacionada con menores índices de accidentes cerebrovasculares, de enfermedades pulmonares y de deterioro cognitivo (Maselko, Kubzansky, Kawachi, Staudenmayer y Berkman, 2006; Van Ness y Kasl, 2003). La asistencia religiosa parece estar relacionada con mejoras a largo plazo en la discapacidad y el manejo del dolor crónico, la búsqueda de fuentes de fortaleza y el descubrimiento de significado en el sufrimiento (Cohen y Pressman, 2006; Kirby, Coleman y Daley, 2004), tanto en los cuadros diabéticos (Sherin L. Krederdt-Araujo et al. 2019) como en los tratamientos contra el cáncer (Hunter-Hernández, Costas-Muñiz y Gany, 2015)

Ahora bien, el recorte fenomenológico que implementan los citados estudios traduciendo el complejo fenómeno de la espiritualidad en variables cuantificables —como pueden ser la participación o asistencia a prácticas religiosas— presenta una serie de dificultades. Aun admitiendo que tales patrones o variables reflejan al menos parcialmen-

te la espiritualidad de la población analizada, nada dicen acerca de las características del compromiso, de la madurez, de la satisfacción espiritual o de las luchas religiosas que puede estar atravesando el grupo experimental. En este sentido, no constituyen variables específicas debido a la ambigüedad o la múltiple significación que pueden tener en una población dichos patrones de comportamiento. La asistencia a comunidades religiosas, por ejemplo, no puede comportarse como una variable específica con una incidencia unilateral sobre la salud física o mental. De aquí las dificultades para poder concluir si tal variable opera en la población estudiada como un auténtico factor protector de la salud o si se debe a otras variables no consideradas en el estudio.

La selección de una variable ambigua atenta contra la misma validez interna de la investigación. Se dice que existe validez interna cuando los resultados de la investigación pueden atribuirse a una variable específica. El problema aquí, entiéndase bien, es si realmente la espiritualidad está representada en la variable seleccionada y, por tanto, si está relacionada efectivamente con los resultados que arrojan los estudios. Dicho en otros términos: bajo una supuesta variable específica se reúne una amplia variedad de comportamientos. Sin realizar prueba alguna de homogeneidad se reconocen las «manzanas y naranjas» como una misma variable, desconociendo la amplias diferencias que existen entre ambas. ¿Acaso los efectos producidos por las manzanas son los mismos efectos que producen las naranjas? Es necesario identificar diversas modalidades de pertenencia o participación en comunidades religiosas de modo que la variabilidad de los resultados refleje estas diferencias.

En una situación análoga se encuentran los estudios relativos a la aparente eficiencia psicoterapéutica de la oración; a saber, un patrón de comportamiento que se refiere a una realidad tan multidimensional como ambigua (Masters, 2005; Brown, 1994; Krause, Chatters, Meltzer y Morgan, 2000; Finney y Malony, 2001). Según sea la tradición religiosa, el término «oración» refiere a prácticas absolutamente diversas. Atendiendo a las connotaciones propias de cada tradición religiosa, la oración podría hacer referencia a estructuras cognitivas y dinámicas emotivas no solo diversas sino incluso contradictorias e inconmensurables. Aun cuando se circunscribiera la práctica de la oración a una tradición religiosa particular, como puede ser la cristiana, se reconocerían en ella diversos tipos de oración: la meditación (la simple escucha, la adoración, la experimentación de un tipo de presencia), la oración ritual, la oración petitoria, la oración coloquial.

En pocas palabras, los citados análisis funcionales pasan por alto las diferencias sustantivas que pueden coexistir entre los miembros de una población acerca de su manera de comprender y vivenciar las variables seleccionadas. Ciertamente, el carácter genérico y ambiguo de estas variables da cuenta de la dificultad que suponen los recortes metodológicos requeridos en los estudios clínicos. Quizás porque la realidad de la espiritualidad y las experiencias religiosas de una población son lo suficientemente complejas como para ser asidas por los citados estudios, sea por la combinación de variables

que dichas dinámicas suponen, sea por la forma individual que ellas operan en cada paciente (Thoresen et al., 2001).

Hasta aquí podríamos enunciar al menos la necesidad de que tales variables recojan o atiendan el contexto teológico, cultural, vivencial y relacional de la espiritualidad y la religión.

Existe también otro tipo o modelo de estudio que evalúa los problemas de naturaleza espiritual en base a herramientas o cuestionarios específicos. Se trata de diversos diseños de investigación que implementan herramientas de evaluación espiritual tales como el Protocolo Rush para detectar la lucha religiosa-espiritual (Fitchett y Risk, 2009); la herramienta de evaluación de un solo ítem de Steihauser que se consigna bajo la pregunta «¿Está usted en paz?» (Steihauser, Clipp y Bosworth, 2004; Steihauser, 2006); el cuestionario WHOQOL-SRPB (WHOQOL Group, 1998; Sherin L. Krederdt-Araujo et al., 2019; Bautista Rodríguez y Zambrano Plata, 2015; Krägeloh, Billington, Henning y Chai, 2015); la herramienta de detección de Mako que inquiriere fundamentalmente si el paciente «¿Tiene dolor espiritual?» (Mako, Galek y Poppito, 2006), o la escala de lesiones espirituales (Fitchett et al., 1999). A partir de tales cuestionarios se selecciona un grupo particular de variables, definidas no ya en los términos de asistencia a una comunidad o por ciertas prácticas religiosas, sino por instancias más genéricas, tales como el sentido de la vida, la relación con la trascendencia, la capacidad de entender a otros, etc.

Aunque tales estudios significan ciertamente un progreso en la tarea de estandarizar las evaluaciones del problema espiritual, estos tienden a profundizar la brecha entre la significación religiosa y los métodos de detección del problema espiritual. Presentan un problema análogo a los diseños experimentales citados más arriba. Se trata de investigaciones que no proporcionan dato alguno respecto al contenido y la evolución en la que se padece un problema espiritual, tal como ha sido apuntado por Gallego et al. (2016) y por Balboni et al. (2017). Nuevamente, la selección de variables no puede hacerse a costa de perder el conocimiento que proveen las tradiciones religiosas. Pues estas permiten, de una u otra manera, dar cuenta de las estructuras cognitivo-emocionales que intervienen en una experiencia espiritual.

En última instancia, la ambigüedad de tales variables acusa quizás los excesos de la radical distinción que la literatura ha establecido entre espiritualidad y religión.

Prácticamente existe un consenso respecto a la distinción entre espiritualidad y religión. Mientras que a la religión se la define como un sistema de creencias, rituales y cultos de una institución o comunidad organizada en vistas a establecer una relación con un ser divino (Sinclair, Pereira y Raffin, 2006), la espiritualidad es entendida como una realidad más inclusiva. Es asociada a experiencias subjetivas con fuerte significación personal (Rumbold, 2003).

Ciertamente, la investigación psicoclínica está interesada en analizar los mecanismos psicológicos subyacentes de las vivencias subjetivas y dinámicas de una población, y no

en los aspectos institucionales y doctrinarios de la religión. Sin embargo, la exacerbación de esta escisión entre espiritualidad y religión podría dejar a la investigación clínica con un concepto oscuro y vacío de espiritualidad. Pues apelando a los principios teóricos y semánticos de una tradición religiosa es como la investigación podría acceder a las estructuras valorativas, emotivas y cognitivas de las experiencias de trascendencia de una población. ¿Cómo podrán develarse las estructuras cognitivas y emocionales de una particular experiencia espiritual sin comprender el marco semántico que una población asigna a tal experiencia? Son las tradiciones religiosas las que han retratado múltiples formas y modalidades de vida espiritual y, en este sentido, son ellas las que permitirían interpretar y hacer comprensible la experiencia espiritual de una población seleccionada.

De un modo u otro, la situación descrita pone en jaque la exacerbada escisión que las investigaciones psicoclínicas trazan entre la base fenoménica susceptible de ser abordada empíricamente y el cuerpo semántico, que es recluso a una instancia pseudocientífica. Pues, ante la tarea de definir y seleccionar variables específicas —que efectivamente sean representativas de un particular tipo de espiritualidad—, es necesario recoger y revisar el bagaje de las tradiciones religiosas.

2. Reduccionismo epistémico

Desde principios del año 2000 hasta la fecha se han publicado numerosos estudios que apuntan a la necesidad de mejorar la calidad de las mediciones empíricas: sea aumentando el rigor de las mediciones, innovando los instrumentos de medición, aumentando el tamaño de las muestras, alargando el seguimiento, extendiendo el control sobre otras variables que podrían incidir en la muestra, etc. Esta tarea de depuración metodológica ha sido implementada en una infinidad de estudios empíricos recientes respecto a la efectividad del *mindfulness* para tratar los trastornos de ansiedad y de depresión, entre otra amplia variedad de problemas psicológicos (Bodhi, 2016, p. 13; Gunaratana, 2011, p. 140; Varela, Thompson y Rosch, 1997, pp. 255-260).

Ahora bien, estos estudios cualitativamente superiores en su abordaje metodológico suponen otro tipo de reducción de carácter epistémico. Naturalmente, los diseños de la investigación psicoclínica están interesados no en esclarecer las propiedades o las dinámicas *per se* de la vida espiritual del paciente o de una población sino en determinar únicamente los posibles efectos psicoterapéuticos de las prácticas religiosas. Dicho en otros términos, buscan traducir o transcribir la dinámica espiritual en su concomitante dinámica psíquica.

Cabe destacar que estos recortes de orden epistémico omiten el estudio de las propiedades únicas o emergentes del complejo fenómeno que representa la espiritualidad. Es decir, el interés está puesto en dilucidar los mecanismos psicológicos específicos de una técnica, sin importar las propiedades que definen a esa técnica como una práctica

espiritual. Ciertamente, las preguntas y los diseños están dirigidos a revelar las conexiones de la espiritualidad (nivel superior) con los niveles inferiores del psiquismo humano. En este sentido, una práctica espiritual puede presentarse empíricamente efectiva para provocar ciertos cambios de conducta en el paciente, en cuanto que en ella pueden operar ciertos principios psicológicos subyacentes que explican su efectividad. Pero lo que nos interesa destacar es que, al ignorar o descartar las notas propias y la relativa autonomía de la dimensión espiritual, se corre el riesgo de que ella en sí misma sea subsumida o reducida a una práctica clínica o a una serie de ejercicios mentales. Por lo que los resultados de dichos estudios no representarán en absoluto la totalidad del fenómeno espiritual, ni una integración genuina de la espiritualidad con modalidades psicoterapéuticas.

En virtud de este recorte epistémico, las investigaciones psicoclínicas informarían sobre un aspecto parcial del fenómeno de la espiritualidad o la religión. Y aún más, en ellas podría operar el cuestionable supuesto de que las leyes que rigen los fenómenos complejos de más alto nivel (la dimensión espiritual) deben seguir las leyes de los niveles más bajos (las dinámicas psicofisiológicas) (Stoeger, 2002). Desde este punto de vista, la dinámica de las prácticas y experiencias espirituales será explicada como la de estructuras y configuraciones similares y deducibles de las leyes de la psicología.

Actualmente, el modelo de apropiación de prácticas religiosas que más consenso tiene es el que traduce las prácticas espirituales en técnicas de concentración (Williams y Swales, 2004). El objetivo de este modelo está puesto en medir las prácticas religiosas en los términos de eficiencia psicoterapéutica –sin suscribir en absoluto las teorías o principios espirituales que las sustentan. Tales estudios operarían sobre un eclecticismo técnico que no encuentra motivos para indagar en los supuestos antropológicos, religiosos o culturales de las prácticas de procedencia espiritual.

Siguiendo este modelo ecléctico de integración se han llevado a cabo múltiples estudios respecto a la efectividad de técnicas de meditación budistas e hinduistas. Es difícil definir la meditación, ya que cada una de estas tradiciones religiosas usa el término de diferentes maneras. Pero en orden a ser ilustrativos podríamos referirnos a la extensa literatura que se esfuerza por definir la atención plena como una tecnología eficiente para manipular la atención, promover la relajación y la autoexploración, independientemente de su contenido religioso, cultural e ideológico original (Langer y Moldoveanu, 2000; Kabat-Zinn, 2003).

La dificultad que supone este modelo de reapropiación de la técnica radica, a nuestro entender, en su mismo proceso de reinención de la práctica espiritual. Este traza un proceso de redefinición de la práctica espiritual que establece nuevos parámetros de eficiencia que muchas veces son ajenos al sustrato antropológico y teológico en el cual han sido originalmente formulados. De ello resulta, a nuestro entender, la dificultad para poder determinar si los estudios versan o no sobre el potencial psicoterapéutico de una práctica espiritual o si se trata de una nueva técnica cuyos

niveles de eficiencia son explicados por las diversas escuelas de la psiquiatría y la psicología modernas.

En línea con Torneatto y Nguyen (2007), Baer (2003) y Orgiles (2013), podríamos argumentar que no habría soporte empírico alguno que probara la eficacia y utilidad propia y específica del *mindfulness*. En la medida en que la práctica del *mindfulness* fue reducida a una serie de ejercicios mentales, ella no parece aportar ningún potencial psicoterapéutico original, o al menos distinto a los recursos que en los años 50 fueron pautados, sistematizados y medidos por la escuela cognitivo-conductual. La estrategia de la exposición a los propios pensamientos o sensaciones fisiológicas, la habilidad metacognitiva de diferenciar los pensamientos perturbadores del propio yo, y las técnicas de autocontrol y relajación, ya eran implementadas en la psicoterapia mucho antes que el *mindfulness* se pusiera de moda. Luego, cabe preguntar: ¿cuál es la estrategia nueva u original que propone el *mindfulness* o las prácticas de atención plena para la psicoterapia contemporánea? O se resguarda el contexto espiritual del *mindfulness* o dicha práctica no representaría novedad alguna para la psicoterapia.

La meditación budista es un estilo de vida sustentado por un particular tipo de espiritualidad, con un concomitante estilo cognitivo. No es un proceso computacional, o una técnica que pueda ser adquirida mecánicamente. O en todo caso, la atención plena, tal como ha sido operacionalizada en la gran mayoría de los estudios psicoclínicos, difiere de su uso budista original (Marlatt et al., 2004). Ciertamente, puede tratarse de técnicas útiles pero sus efectos pueden resultar muy diferentes de su efecto original. En este sentido, despojar esta práctica espiritual de su contexto ético y cultural tiene efectos potencialmente negativos para la propia investigación psicoclínica.

Hoy disponemos de investigaciones que han desafiado este modelo ecléctico de integración y que han comenzado a prestar atención a las diferencias específicas de una amplia variedad de tipos de meditación, incluidas las técnicas de atención plena, *vipassana*, y budista *vajrayana* (Andresen, 2000).

En lugar de retraducir la práctica espiritual a una técnica, estos estudios procuran entenderla seleccionando variables que se explican en virtud de su propio marco religioso o cosmológico. Atienden al marco teórico que proveen las tradiciones religiosas en cuanto que en ellas pueden discriminarse las diversas variables que afectan al bienestar emocional del paciente. Después de todo, los diferentes tipos de meditación, además de un tipo de relajación general, producen diferentes cambios específicos o variables a los que la investigación psicoclínica solo puede acceder en la medida en que contemple el contexto teórico que los sustenta.

3. Reduccionismo ontológico

Finalmente cabe citar los estudios que suponen explícita o implícitamente un reduccionismo de tipo ontológico. Este tipo de reduccionismo no es el resultado de un

procedimiento metodológico o de un acercamiento epistemológico particular, sino de un supuesto metafísico acerca de la naturaleza del objeto de estudio. Por ejemplo, se asume que un determinado objeto –en este caso, la espiritualidad– no tiene una existencia propia o real sino que es una combinación de múltiples procesos o dinámicas. Ciertamente, toda investigación supone recortes de tipo metodológico y epistémico. Pero el reduccionismo ontológico da un paso más, asumiendo que tales recortes son los que definen lo real. Aquella pequeña parcela de la realidad que constituye el objeto de estudio de una investigación, se asume –en virtud del reduccionismo ontológico– como el todo de lo real. Es el tipo más extremo de reducción.

Dentro de este tipo de reduccionismos podemos citar aquellos estudios que conciben la religión o la espiritualidad como un fenómeno cultural y psicológico. La religión es entendida como un sistema multidimensional o una cosmovisión que subyace a una cultura y permite a sus miembros construir redes de significado y dar sentido al mundo y a sus existencias individuales (Marsden y Longfield, 1992; Vergote, 1997, p. 19).

Dentro de esta comprensión cultural se han formulado modelos éticos y émicos. En el modelo ético se entiende la religión como un fenómeno universal que posee características similares en los diferentes entornos culturales. Las distintas religiones, como pueden ser el budismo y el cristianismo, compartirían un conjunto de prácticas devocionales que podrían discriminarse en un conjunto universal de categorías: conexiones comunitarias, valores comunes, experiencias naturales. En el modelo émico, en cambio, cada religión posee notas propias y específicas que se explicarían en virtud de su particular contexto geográfico, social e histórico (Shuman y Meador, 2003, pp. 37-40).

Ciertamente, estas modalidades de reduccionismo ontológico poseen un valor heurístico considerable en la medida en que permiten formular esquemas comprensivos sencillos, susceptibles de ser traducidos a diseños experimentales –tal como argumentan Corveleyn, 1996; Watts, 2002, p. 4; Kistler, 2006; Schaffner, 2006, y Poirier, 2006–. Pero, ¿acaso lo que dichas hipótesis ganan en simplicidad no lo pierden en precisión? El coste de estas simplificaciones es particularmente problemático cuando se trata de un fenómeno tan complejo como es la religión.

El reduccionismo ontológico que atraviesa ciertos diseños puede excluir de suyo la posibilidad de admitir, por ejemplo, una nota esencial a la mayoría de las religiones como es la trascendencia. Un estudio de la religión que excluye la consideración de esta variable parece tener límites para comprender su objeto de estudio. Ciertamente, la trascendencia no es una variable directamente observable o susceptible de ser sometida a un diseño experimental. No obstante, en la medida en que la psicología no considere la posibilidad de que operen en las estructuras psíquicas dinámicas trascendentes a ellas podría estar excluyendo un aspecto crítico y central de un fenómeno que constituye como tal a la espiritualidad (Vergote, 1969, pp. 5-21).

El reduccionismo ontológico, al no comprender las dinámicas espirituales y psicológicas como realidades o procesos relativamente autónomos e independientes, puede resultar incapaz de reconocer cuadros clínicos en los que un paciente pueda tener una vida espiritual rica mientras padece una enfermedad mental o, por el contrario, ser psicológicamente saludable pero sufrir profundos problemas espirituales (Welch, 1996, p. 98).

Consideraciones finales

El reduccionismo es un proceso necesario e ineludible de la investigación empírica. No obstante –tal como hemos querido mostrar–, los recortes de tipo metodológico, epistémico y ontológico condicionan fuertemente los resultados y el modo en que son interpretados los datos obtenidos de la investigación psicoclínica. De aquí la necesidad de una instancia de reflexión metateórica capaz de dar cuenta de las distorsiones o límites comprensivos a los que puedan dar lugar tales reduccionismos. Pues esta tarea de análisis es la que permite relativizar la información del diseño como la respuesta a un particular instrumento de selección, observación, medición y análisis. Revela el carácter fragmentario de los estudios, ponderando la información del diseño en virtud de sus supuestos.

En el primer punto hemos mostrado cómo ciertos reduccionismos metodológicos atentan contra los propios diseños de investigación. La traducción del complejo fenómeno de la espiritualidad a una variable ambigua, es decir, mal delimitada e inespecífica, atenta contra la propia validez interna del diseño de investigación.

Tanto en el segundo como el tercer apartado del estudio hemos apuntado la necesidad de que las prácticas religiosas sean estudiadas desde y en su marco antropológico, teológico o cosmológico propio. De no ser así, el coste tanto del reduccionismo epistémico como del ontológico es que finalmente no se concreta ningún modelo de integración entre la espiritualidad y la salud mental. Estos recorten abren modalidades psicologicistas de integración donde no existe propiamente un diálogo mutuo entre la religión y las ciencias de la salud mental. La espiritualidad queda subsumida a estructuras o dinámicas psicológicas.

La espiritualidad y la religión parecen constituir y manifestar un núcleo medular de la personalidad. Se trata de una dimensión de la personalidad que parece exigir no solo un mayor dialogo interdisciplinario sino fundamentalmente una nueva comprensión de la constitución epistemológica de la psicoterapia.

La investigación psicoclínica no puede ser concebida como una disciplina encerrada en sus propios límites. Debe complementarse o, aún más propiamente, fundamentarse y hallar justificación en otros dominios conceptuales, tales como los que proveen las tradiciones teológicas y espirituales.

De un modo u otro, la discusión respecto a la relación entre religión, espiritualidad y salud mental pone sobre la mesa que la fundamentación de la espiritualidad depende de elementos que históricamente fueron excluidos de toda consideración racional o científica. Se trata de una extraña paradoja: en la medida en que se incorporan elementos y principios aparentemente ajenos a la racionalidad científica, las demostraciones científicas adquieren fuerza explicativa. A saber, en la medida en que la psicoterapia atiende al marco conceptual de las tradiciones religiosas, ella será capaz de dar cuenta de la eficacia psicoterapéutica de las experiencias religiosas.

Referencias bibliográficas

- AMERICAN PSYCHIATRIC ASSOCIATION, *Diagnostic and statistical manual of mental disorders* (4ª ed., texto rev.). Washington, DC: American Psychiatric Association, 2000.
- AMERICAN PSYCHOLOGICAL ASSOCIATION. PRESIDENTIAL TASK FORCE ON EVIDENCE-BASED PRACTICE, «Evidence-based practice in psychology». *American Psychologist*, 61, 2006, pp. 271-285.
- ANDRESEN, J., «Meditation meets behavioural medicine: The story of experimental research on meditation», en: ANDRESEN, J.; FORMAN, R. K. C. (eds.), *Cognitive models and spiritual maps: Interdisciplinary explorations of religious experience*. Thorverton, UK: Imprint Academic, 2000, pp. 17-73.
- ANDERSON, C.; KIM, C., «Evaluating Treatment Efficacy with Single-Case Designs», en: ROBERTS, M.; ILLARDI, S. (eds.), *Handbook of Research Methods in Clinical Psychology*. Oxford, UK: Blackwell, 2003, pp. 73-91.
- ANDLER, D., «Federalism in science-complementarity vs. perspectivism: Reply to Harré». *Synthese*, 151, 2006, pp. 519-522.
- BAER, R. A., «Mindfulness training as a clinical intervention: A conceptual and empirical review». *Clinical Psychology: Science and Practice*, 10, 2003, pp. 125-143.
- BAKER, T. B.; MCFALL, R. M.; SHOHAM, V., «Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care». *Psychological Science in the Public Interest*, 9/2, 2008, pp. 67-103.
- BALBONI, T. A.; FITCHETT, G.; HANDZO, G. F.; JOHNSON, K. S.; KOENIG, H. G.; PARGAMENT, K. I...; STEINHAUSER, K. E., «State of the science of spirituality and palliative care research Part II: screening, assessment, and interventions». *Journal of pain and symptom management*, 54/3, 2017, pp. 441-453.
- BARRETT, L.; DUNBAR, R.; LYCETT, J., *Human evolutionary psychology*. Princeton, NJ: Princeton University Press, 2002.
- BAUTISTA RODRÍGUEZ, L. M.; ZAMBRANO PLATA, G. E., «La calidad de vida percibida en pacientes diabéticos tipo 2». *Investigación en Enfermería: Imagen y Desarrollo*, 17/1, 2015, pp. 131-148.
- BECKER, K., *Unlikely companions: C.G. Jung on the Spiritual Exercises of Ignatius of Loyola: An exposition and critique based on Jung's lectures and writings*. Leominster, UK: Gracewing, 2001.
- BENNETT, M. R.; HACKER, P. M. S., *Philosophical foundations of neuroscience*. Malden, MA: Blackwell, 2003.
- BODHI, V. B., «The transformations of mindfulness». *Handbook of Mindfulness*. Berlín: Springer, 2016, pp. 3-14.
- BROWN, L., *The human side of prayer: The psychology of praying*. Birmingham, AL: Religious Education Press, 1994.
- CAHN, B. R.; POLICH, J., «Meditation states and traits: EEG, ERP, and neuroimaging studies». *Psychological Bulletin*, 132, 2006, pp. 180-211.

- CHARRY, E., «Theology after psychology», en: MCMINN, M.; PHILLIPS, T. (eds.), *Care for the soul: Exploring the intersection of psychology and theology*. Downers Grove, IL: InterVarsity Press, 2001, pp. 118-133.
- CHATTERS, L. M., «Religion and health: Public health research and practices». *Annual Review of Public Health*, 21, 2000, pp. 335-367.
- COHEN, S.; PRESSMAN, S., «Positive affect and health». *Current Directions in Psychological Science*, 15, 2006, pp. 122-125.
- CONTRADA, R. J.; GOYAL, T. M.; CATHER, C.; RAFALSON, L.; IDLER, E. L.; KRAUSE, T. J., «Psycho-social factors in outcomes of heart surgery: The impact of religious involvement and depressive symptoms». *Health Psychology*, 23, 2004, pp. 227-238.
- COOPER, T. D., *Dimensions of evil: Contemporary perspectives*. Minneapolis, MN: Fortress, 2007, pp. 30-38.
- CORVELEYN, J., «The psychological explanation of religion as wish-fulfillment. A test-case: The belief in immortality», en: VAN BELZEN, J. A.; VAN DER LANS, J. M. (series eds.); GRZYMALA-MOSZCZYNSKA, H.; BEIT-HALLAHMI, B. (vol. eds.), *International series in the psychology of religion: Vol. 4. Religion, psychopathology, and coping*. Amsterdam: Rodopi, 1996, pp. 57-70.
- FINNEY, J.; MALONY, H., «An empirical study of contemplative prayer as an adjunct to psychotherapy», en: FRANCIS, L.; ASTLEY, J., (eds.), *Psychological perspectives on prayer*. Leominster, UK: Gracewing, 2001, pp. 359-367.
- FITCHETT, G.; RISK J. L., «Screening for spiritual struggle (4-1-12)». *Pastoral Care Counsel*, 63, 2009.
- FITCHETT, G.; RYBARCZYK, B. D.; DE MARCO, G. A.; NICHOLAS, J. J., «The role of religion in medical rehabilitation outcomes: a longitudinal study». *Rehabil Psychol*, 44, 1999, pp. 333-353.
- GALLEGO, S.; DE LOS BUEIS, F.; PRATS, S.; ORTEGA-VERDAGUER, A.; ANTONÍN, P.; SERRA, M.; ROSSY, C.; REVUELTO, A.; CIUDAD, E.; ÁLVAREZ-SEGURA, M., «Contribuciones y Confusiones de la Espiritualidad en la salud mental». *Gaceta de Psiquiatría Universitaria*, 12/4, 2016, pp. 353-358.
- GUNARATANA, B. H., *Mindfulness in plain English 20th anniversary edition*. Boston: Wisdom publications, 2011.
- HUNTER-HERNÁNDEZ, M.; COSTAS-MUÑIZ, R.; GANY, F., «Missed opportunity: Spirituality as a bridge to resilience in Latinos with cancer». *Journal of Religion and Health*, 54, 2015, pp. 2367-2375.
- IDLER, E. L.; MUSICK, M. A.; ELLISON, C. G.; GEORGE, L. K.; KRAUSE, N.; ORY, M. G. ET AL., «Measuring multiple dimensions of religion and spirituality for health research: Conceptual background and findings from the 1998 general social survey». *Research on Aging*, 25, 2003, pp. 327-365.
- INGERSOLL, R. E., «Spirituality, religion, and counseling: Dimensions and relationships». *Counseling & Values*, 38, 1994, pp. 98-111.
- KABAT-ZINN, J., «Mindfulness-based interventions in context: Past, present, and future». *Clinical Psychology: Science and Practice*, 10, 2003, pp. 144-156.
- KALDOR, P.; FRANCIS, L. J.; FISHER, J. W., «Personality and spirituality: Christian prayer and eastern meditation are not the same». *Pastoral Psychology*, 50, 2002, pp. 165-172.
- KAZDIN, A., *Métodos de investigación en psicología clínica*. México, D.F.: Prentice Hall, 2001.
- KIM, J., «Emergence: Core ideas and issues». *Synthese*, 151, 2006, pp. 547-559.
- KIRBY, S. E.; COLEMAN, P. G.; DALEY, D., «Spirituality and well-being in frail and nonfrail older adults». *Journal of Gerontology: Psychological Sciences*, 59B/3, 2004, pp. 123-129.
- KISTLER, M., «New perspectives on reduction and emergence in physics, biology and psychology». *Synthese*, 151, 2006, pp. 311-312.
- KOENIG, H. G., «Religion as cognitive schema». *The International Journal for the Psychology of Religion*, 5, 1995, pp. 31-37.
- KOENIG, H. G., «Religion and medicine IV: Religion, physical health, and clinical implications». *International Journal of Psychiatry in Medicine*, 31, 2001, pp. 321-336.

- KOSEK, R. B., «The contribution of object relations theory in pastoral counseling». *The Journal of Pastoral Care*, 50, 1996, pp. 371-381.
- KRAUSE, N., «Church-based social support and health in old age: Exploring variations by race». *Journal of Gerontology: Social Sciences*, 57B/6, 2002, pp. 332-347.
- KRAUSE, M.; CHATTERS, L.; MELTZER, T.; MORGAN, D., «Using focus groups to explore the nature of prayer in late life». *Journal of Aging Studies*, 14, 2000, pp. 191-212.
- KRÄGELOH, C. U.; BILLINGTON, D. R.; HENNING, M. A.; CHAI, P. P., «Spiritual quality of life and spiritual coping: Evidence for a two-factor structure of the WHOQOL spirituality, religiousness, and personal beliefs module». *Health and Quality of Life Outcomes*, 13, 2015, pp. 10-11.
- KULIK, A.; SZEWCZYK, L., «Sense of meaning of life and the emotional reaction among young people pursuing different types of meditation». *Studia Psychologica*, 44, 2002, pp. 155-166.
- LANGER, E.; MOLDOVEANU, M., «The construct of mindfulness». *Journal of Social Issues*, 56/1, 2000, pp. 1-9.
- LEVIN, J. S.; CHATTERS, L.; TAYLOR, R., «Religion, health and medicine in African Americans: Implications for physicians». *Journal of the National Medical Association*, 97, 2005, pp. 237-249.
- MACKENZIE, E. R.; RAJAGOPAL, D. E.; MEILBOHM, M.; LAVIZZO-MOUREY, R., «Spiritual support and psychological well-being: Older adults' perceptions of the religion and health connection». *Alternative Therapies in Health and Medicine*, 6 (6), 2000, pp. 37-45.
- MACINNES, E., *Zen contemplation for Christians*. Lanham, MD: Sheed & Ward, 2003.
- MARSDEN, G.; LONGFIELD, B. (eds.), *The secularization of the academy*. Oxford: Oxford University Press, 1992.
- MAKO, C.; GALEK, K.; POPPITO, S. R., «Spiritual pain among patients with advanced cancer in palliative care». *J Palliat Med*, 9, 2006, pp. 1106-1113.
- MARLATT, G.; WITKIEWITZ, K.; DILLWORTH, T.; BOWEN, S.; PARKS, G.; MACPHERSON, L. ET AL., «Vipassana meditation as a treatment for alcohol and drug use disorders», en: HAYES, S.; FOLLETTE, V.; LINEHAN, M. (eds.), *Mindfulness and acceptance: Expanding the cognitive-behavioral tradition*. Nueva York: Guilford, 2004, pp. 261-287.
- MASELKO, J.; KUBZANSKY, L.; KAWACHI, I.; STAUDENMAYER, J.; BERKMAN, L., «Religious service attendance and decline in pulmonary function in a high-functioning elderly cohort». *Annals of Behavioral Medicine*, 32, 2006, pp. 245-253.
- MASTERS, K. S., «Research on the healing power of distant intercessory prayer: Disconnect between science and faith». *Journal of Psychology and Theology*, 33, 2005, pp. 268-277.
- MCCULLOUGH, M. E.; PARGAMENT, K.; THORESEN, C., «The psychology of forgiveness: History, conceptual issues, and overview», en: MCCULLOUGH, M.; PARGAMENT, K.; THORESEN, C. (eds.), *Forgiveness: Theory, research and practice*. Nueva York: Guilford Press, 2000, pp. 1-14.
- MCCULLOUGH, M. E.; HOYT, W. T.; LARSON, D. B.; KOENIG, H. G.; THORESEN, C., «Religious involvement and mortality: A meta-analytic review». *Health Psychology*, 19, 2000, pp. 211-222.
- MCCULLOUGH, M. E., «Religious involvement and mortality: Answers and more questions», en: PLANTE, T.; SHERMAN, A. (eds.), *Faith and health: Psychological perspectives*. Nueva York: Guilford Press, 2001, pp. 53-74.
- MCGRATH, P., «Defining spirituality: From meaning-making to connection», en: AMBROSE, S. (ed.), *Religion and psychology: New research*. Nueva York: Nova Science, 2006, pp. 223-239.
- MUELLER, P. S.; PLEVAK, D. J.; RUMMANS, T. A., «Religious involvement, spirituality, and medicine: Implications for clinical practice». *Mayo Clinic Proceedings*, 76, 2001.
- MURPHY, N., «Philosophical resources for integration», en: DUECK, A.; LEE, C. (eds.), *Why psychology needs theology: A radical reformation perspective*. Grand Rapids, MI: Eerdmans, 2005, pp. 3-27.

- NEWBERG, A.; WINTERING, N.; MORGAN, D.; WALDMAN, M., «The measurement of regional cerebral blood flow during glossolalia: A preliminary SPECT study». *Psychiatry Research: Neuroimaging*, 148, 2006, pp. 67-71.
- ORGILES, M. L. M.; SEVILLA GASCO, J., «¿Por qué no soy un terapeuta mindfulness?». *Información Psicológica*, 106, 2013, pp. 54-69.
- PELTZER, K., «Preventive health behavior, personality and religiosity among Black and White South Africans». *Studia Psychologica*, 46, 2004, pp. 37-48.
- PERREZ, M., «Psychotherapeutic Methods: Between scientific foundation and everyday knowledge». *New Ideas in Psychology*, 7, pp. 133-145.
- POIRIER, P., «Finding a place for elimination in inter-level reductionist activities: Reply to Wimsatt». *Synthese*, 151, 2006, pp. 477-483.
- RITSKES, R.; RITSKES-HOITINGA, M.; STODKILDE-JORGENSEN, H.; BAERENTSEN, K.; HARTMAN, T., «MRI scanning during Zen meditation: The picture of enlightenment». *Constructivism in the Human Sciences*, 8, 2003, pp. 85-90.
- ROOF, W. C., *Spiritual marketplace: Baby boomers and the remaking of American religion*. Princeton, NJ: Princeton University Press, 1999.
- RUMBOLD, B. D., «Caring for the spirit: lessons from working with the dying». *Med J Aust*, 179/6 Suppl, 2003, p. 11.
- SCHAFFNER, K. F., «Reduction: The Cheshire cat problem and a return to roots». *Synthese*, 151, 2006, pp. 377-402.
- SCHNEIDERS, S. M., «The study of Christian spirituality: Contours and dynamics of a discipline». *Christian Spirituality Bulletin*, 6/1, 1998, pp. 3-12.
- SETHI, S.; BHARGAVA, S. C., «Relationship of meditation and psychosis: Case studies». *Australian and New Zealand Journal of Psychiatry*, 37, 2003, p. 382.
- SHAHABI, L.; POWELL, L. H.; MUSICK, M. A.; PARGAMENT, K. I.; THORESEN, C. E.; WILLIAMS, D. ET AL., «Correlates of self-perceptions of spirituality in American adults». *Annals of Behavioral Medicine*, 24, 2002, pp. 59-68.
- KREDERDT-ARAUJO, S. L.; DOMINGUEZ-CANCINO, K. A.; JIMÉNEZ-CORDOVA, R.; PAZ-VILLANUEVA, M. Y.; MENDIGURE FERNANDEZ, J.; LEYVA-MORAL, J. M.; PALMIERI, P. A., «Spirituality, Social Support, and Diabetes: A Cross-Sectional Study of People Enrolled in a Nurse-Led Diabetes Management Program in Peru». *Hispanic Health Care International*, 17/4, 2019, pp. 162-171.
- SHUMAN, J.; MEADOR, K., *Heal thyself: Spirituality, medicine, and the distortion of Christianity*. Oxford: Oxford University Press, 2003.
- SLIFE, B. D.; HOPKINS, R. O., «Alternative assumptions for neuroscience: Formulating a true monism», en: SLIFE, B. D.; REBER, J. S.; RICHARDSON, F. C. (eds.), *Critical thinking about psychology: Hidden assumptions and plausible alternatives*. Washington, DC: American Psychological Association, 2005, pp. 121-147.
- SMITH, H., «Methodology, comparisons, and truth», en: PATTON, K.; RAY, B. (eds.), *A magic still dwells: Comparative religion in the postmodern age*. Berkeley, CA: University of California, 2000, pp. 172-181.
- SINCLAIR, S.; PEREIRA, J.; RAFFIN, S., «A thematic review of the spirituality literature within palliative care». *J Palliat Med*, 9/2, 2006, p. 467.
- SOLLOD, R., «Non-scientific sources of psychotherapeutic: A case for inclusion», en: SHAFRANSKE, E. P. (ed.), *Philosophy, religion and psychotherapy*. Washington DC: University Press of America, 1982, pp. 41-56.
- STEFFEN, P. R.; HINDERLITER, A. L.; BLUMENTHAL, J. A.; SHERWOOD, A., «Religious coping, ethnicity, and ambulatory blood pressure». *Psychosomatic Medicine*, 63, 2001, pp. 523-530.
- STEINHAUSER, K. E.; CLIPP, E. C.; BOSWORTH, H. B. ET AL., «Measuring quality of life at the end of life: validation of the QUAL-E». *Palliat Support Care*, 2, 2004, pp. 3-14.

- STEINHAUSER, K. E.; VOILS, C. I.; CLIPP, E. C.; BOSWORTH, H. B.; CHRISTAKIS, N. A.; TULSKY, J. A., «Are you at peace?»: one item to probe spiritual concerns at the end of life». *Arch Intern Med*, 166, 2006, pp. 101-105.
- TASK FORCE ON PROMOTION AND DISSEMINATION OF PSYCHOLOGICAL PROCEDURES, «Training in and dissemination of empirically-validated psychological treatment: Report and recommendations». *The Clinical Psychologist*, 48, 1995, pp. 2-23.
- TAYLOR, C., *A secular age*. Cambridge, MA: Belknap, 2007, pp. 679, 704-707.
- THEOPHAN, ST., *The spiritual life and how to be attuned to it* (trad. A. Dockham). Platina, CA: St. Herman of Alaska Brotherhood, 1995.
- THORESEN, C. E.; HARRIS, A. H. S.; OMAN, D., «Spirituality, religion, and health: Evidence, issues, and concerns», en: PLANTE, T.; SHERMAN, A. (eds.), *Faith and health: Psychological perspectives*. Nueva York: Guilford, 2001, pp. 15-52.
- TORNEATTO, T.; NGUYEN, L., «Does mindfulness meditation improve anxiety and mood symptoms? A review of the controlled research». *La Revue canadienne de psychiatrie*, 52, 4, 2007, pp. 260-266.
- VACHON, M.; FILLION, L.; ACHILLE, M., «A conceptual analysis of spirituality at the end of life». *Journal of Palliative Medicine*, 12/1, 2009, pp. 53-59.
- VANDE KEMP, H., «Gordon Allport's pre-1950 writings on religion: The archival record», en: BELZEN, J. A. (ed.), *Aspects in contexts: Studies in the history of psychology of religion*. Amsterdam: Rodopi, 2000, pp. 129-172.
- VAN NESS, P. H.; KASL, S. V., «Religion and cognitive dysfunction in an elderly cohort». *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 58/1, 2003, pp. 21-29.
- VARELA, F. J.; THOMPSON, E.; ROSCH, E., *De cuerpo presente: Las ciencias cognitivas y la experiencia humana*. Barcelona: Gedisa, 1997.
- VERGOTE, A., *The religious man: A psychological study of religious attitudes* (trad. M.-B. Said). Dayton, OH: Pfalaum, 1969.
- VERGOTE, A., *Religion, belief and unbelief: A psychological study*. Amsterdam/Atlanta, GA: Leuven University Press, 1997.
- WAGENER, L.; MALONY, H. N., «Spiritual and religious pathology in childhood and adolescence», en: ROEHLKEPARTAIN, E.; KING, P.; WAGENER, L.; BENSON, P. (eds.), *The handbook of spiritual development in childhood and adolescence*. Thousand Oaks, CA: Sage, 2006, pp. 137-149.
- WATTS, F., *Theology and psychology*. Aldershot, UK: Ashgate, 2002.
- WHOQOL GROUP, «Development of the world health organization WHOQOL-BREF quality of life assessment». *Psychol Med.*, 28/3, 1998, pp. 551-558.
- WISEMAN, J., «Thomas Merton and Theravada Buddhism», en: B. THURSTON (ed.), *Merton & Buddhism: Wisdom, emptiness, and everyday mind*. Louisville, KY: Fons Vitae, 2007, pp. 31-50.
- WUTHNOW, R., *After heaven: Spirituality in America since the 1950s*. Berkeley, CA: University of California Press, 1998.
- YORSTON, G. A., «Mania precipitated by meditation: A case report and literature review». *Mental Health, Religion & Culture*, 4, 2001, pp. 209-213.
- ZIZIOULAS, J., *Communion and otherness: Further studies in personhood and the church* (ed. P. McPartlan). Londres: T & T Clark, 2006.

María Teresa GARGIULO
 CONICET-Argentina
 gargiulomteresa@gmail.com

Article rebut: 6 de novembre de 2019. Article aprovat: 21 de setembre de 2020

LA MONSTRUOSIDAD DEL OTRO. SLAVOJ ŽIŽEK Y LA CRÍTICA A LAS CONCEPCIONES OCCIDENTALES DE LA OTREDAD¹

Jorge LEÓN CASERO

Resumen

La segunda mitad del siglo XX ha desarrollado dos concepciones del Otro que pretenden superar la concepción nacionalista del mismo como un elemento ajeno a la comunidad que amenaza su existencia: el multiculturalismo liberal tolerante y la ética de la alteridad desarrollada por Lévinas y Derrida. Según Slavoj Žižek, la primera funciona como la otra cara de la lógica libidinal nacionalista, mientras que la segunda mantiene una concepción del otro incapaz de subvertir la autoridad simbólica que lo define. Frente a ellas, Žižek recurre a una concepción lacaniana del Otro como base para desarrollar una subversión política de la autoridad simbólica, que permita acceder a un encuentro con la monstruosidad que según él caracteriza la auténtica otredad del otro.

Palabras clave: filosofía de la alteridad, nacionalismo, multiculturalismo, identidad simbólica, violencia.

The monstrosity of the Other.
Slavoj Žižek and the criticism of Western conceptions of otherness

Abstract

During the second half of the 20th century, two theories of the Other have been developed that seek to overcome the nationalist conception that identifies it with an element outside the community, and whose existence it threatens: tolerant liberal multiculturalism and Lévinas' and Derrida's ethics of difference. According to Slavoj Žižek, the first notion operates as the other side of a nationalist libidinal logic and the second implies an understanding of the other incapable of subverting the symbolic authority that defines it. To address both, Žižek resorts to a Lacanian interpretation of the Other

¹ Este artículo es resultado del proyecto de investigación «Racionalidad económica, ecología política y globalización: hacia una nueva racionalidad cosmopolita». Programa Estatal Proyectos I+D Retos de la Investigación del Ministerio de Ciencia, Innovación y Universidades, Ref.: PID2019-109252RB-I00.

as the basis for developing a political subversion of symbolic authority that allows for an encounter with the monstrosity that, according to him, characterizes the authentic other's otherness.

Keywords: Philosophy of otherness, Nationalism, Multiculturalism, Symbolic Identity, Violence.

1. Introducción

El punto de vista desde el cual concebimos al otro no influye únicamente en la relación intersubjetiva y personal que mantenemos con el mismo, sino que es un elemento fundamental del modo en que construimos el imaginario social y político con el que nos definimos a nosotros mismos como colectivo. A este respecto, Slavoj Žižek ha indicado numerosas veces que la relación ética y política con el otro no puede ser descrita adecuadamente mediante análisis culturales que pretenden identificar las diferencias y similitudes que cada grupo presenta a nivel sociológico y/o de costumbres. Es preciso complementar dichos enfoques con análisis de tipo libidinal que atiendan al modo en que nuestro deseo influye en la construcción de los fantasmas simbólicos que estructuran nuestra percepción de la otredad.

Desde este punto de vista, Žižek ha desarrollado dos análisis profundamente interrelacionados de los procesos libidinales desde los que se construye la figura del otro en el nacionalismo y en el multiculturalismo liberal. Ambos han tenido un gran impacto internacional. Concretamente, su crítica al nacionalismo ha sido aplicada al análisis tanto de la formación de la identidad nacional estadounidense² como de la identidad nacional sueca,³ australiana,⁴ canadiense,⁵ japonesa,⁶ irlandesa,⁷

² D. MILLER, «The populist fantasy: American ethno-nationalism and the dysphoria of the body politic». *Soundings*, 102 (2-3), 2019, pp. 226-252. DOI: 10.5325/soundings.102.2-3.0226

³ E. HANSSON Y D. JANSSON, «Who's afraid of the "beggar"? A psychoanalytic interpretation of the crises triggered by the begging of "EU migrants" in Sweden». *Social and Cultural Geography*, 22 (3), 2019, pp. 339-356. DOI: 10.1080/14649365.2019.1585563

⁴ S. HUTCHINSON, «Humanitarian critique and the settler fantasy: the Australian press and settler colonial consciousness during the Waikato War, 1863-1864». *Settler Colonial Studies*, 4 (1), 2014, pp. 48-63. DOI: 10.1080/2201473X.2013.834550

⁵ P. KINGSBURY, «The world cup and the national thing on commercial drive, Vancouver». *Environment and Planning D: Society and Space*, 29 (4), 2011, pp. 716-737. DOI: 10.1068/d4410

⁶ D. BRINK, «Nativist legacies of desinicization and nationalist sentiment in poetry during the Second Sino-Japanese War». *Inter-Asia Cultural Studies*, 12 (1), 2011, pp. 43-61. DOI: 10.1080/14649373.2011.532944

⁷ B. COLLINS, «How Does the Nation Plug the Hole in Irish Constitutional Discourse?». *Law, Culture and the Humanities*, 2 (2), 2006, pp. 218-238. DOI: 10.1191/1743872106lw044oa

italiana⁸ o británica,⁹ entre otras. Del mismo modo, son ya varios los autores que coinciden con Žižek y consideran que la concepción del otro mantenida por el multiculturalismo liberal no es más que la otra cara de la concepción propia del etnonacionalismo, debido a que ambas continúan planteando por diversos caminos la superioridad de la propia posición frente a «un otro» cuya otredad resulta completamente absorbida por los esquemas conceptuales y libidinales con los que construimos nuestra propia identidad de forma previa a un encuentro real con el otro.

En el caso del nacionalismo, la otredad del otro es libidinalmente producida como una entidad fantasmática que pone en peligro la integridad de «la Cosa-nación», entendiéndose por esta última la creencia colectiva o sociosimbólica que garantiza la unidad de un grupo supuestamente originario que se atribuye a sí mismo un acceso privilegiado al goce (*jouissance*) de la misma. En cambio, en el caso del multiculturalismo se prescribe una exigencia de «autenticidad» cultural de la Cosa-nación del otro como condición de posibilidad para poder tolerar su otredad, al mismo tiempo que se le niega cualquier tipo de manifestación que exceda el marco de tolerancia impuesto por una cultura dominante que, considerándose a sí misma como libidinalmente neutra y sin una Cosa-nación propia, establece qué otros pueden ser considerados interlocutores válidos y cuáles deben ser desacreditados por constituir posiciones etnonacionalistas y/o fundamentalistas intolerantes con el resto de culturas.

Concretamente, este fue el marco de discusión del debate que tuvo lugar en la década de los 90 entre Charles Taylor, Jürgen Habermas y Will Kymlicka, frente al cual desarrollará Žižek su crítica. De forma sucinta, Taylor defendió que, bajo ciertas condiciones, los derechos fundamentales de los individuos podrían ser restringidos con el fin de promover la supervivencia de las culturas amenazadas.¹⁰ Frente a ello, Kymlicka respondió con una teoría multiculturalista liberal que distinguía entre los derechos colectivos de las culturas minoritarias, orientados a funcionar como «protecciones externas» frente a las culturas mayoritarias, de aquellos otros que conllevaban «restricciones internas» para los individuos pertenecientes a las culturas minoritarias, defendiendo que una teoría liberal del multiculturalismo únicamente podría permitir los primeros, pero nunca los segundos.¹¹ Finalmente, Habermas desarrolló su concepto de «patriotismo constitucional» e insistió en la concepción ilustrada del Estado de dere-

⁸ N. D. HARNEY, «The alternative economies of emigration and immigration, the real and the constitution of Italian nation spaces». *Mobilities*, 1 (3), 2006, pp. 373-390. DOI: 10.1080/17450100600916024

⁹ G. DALY, «Politics and the Impossible: Beyond Psychoanalysis and Deconstruction», *Theory, Culture & Society*, 16 (4), 1999, pp. 75-98. DOI: 10.1177/026327699922050728

¹⁰ CH. TAYLOR, *El multiculturalismo y la política del reconocimiento*. México D. F.: Fondo de Cultura Económica, 2009, pp. 53-115.

¹¹ W. KYMLICKA, *Ciudadanía multicultural. Una teoría liberal de los derechos de las minorías*. Barcelona: Paidós Ibérica, 1996, pp. 58-71.

cho, defendiendo tanto la neutralidad de las instituciones jurídicas respecto a todas las culturas como la suficiencia de los derechos fundamentales de los individuos para poder garantizar la supervivencia de cualquier cultura.¹²

Frente a estos enfoques, Žižek considera que todos ellos mantienen una concepción del otro como enemigo potencial que debe someterse a los valores de una cultura dominante que niega tener valores propios, si no quiere ser expulsado de la comunidad multicultural que le impone el modo y los límites con los que debe construir su propia otredad. En este sentido, «el multiculturalismo postmoderno criticado por Žižek tolera la otredad del otro eliminando su misma esencia».¹³

Por una parte, el multiculturalismo liberal es visto como una continuación limitada de la concepción etnonacionalista del otro en tanto que enemigo que pone en peligro la vinculación libidinal de la comunidad originaria con su Cosa-Nación. Por otra, la existencia misma de la Cosa-nación es analizada como un fantasma generado de forma retroactiva por la hegemonía mundial de un multiculturalismo (neo)liberal que es percibido como el principal enemigo que pone en peligro nuestra Cosa-nación, a la vez que suscita nuestro deseo. Se postula entonces una relación de doble retroactividad que ha llevado a algunos autores a hablar de una «relación dialéctica entre el Estado-nación y el capitalismo»,¹⁴ e incluso de una «lógica social esquizoide del etnonacionalismo, [que] por un lado, defiende los valores de la familia y la comunidad, mientras que por otro destruye y atomiza la familia mediante un capitalismo desenfrenado».¹⁵ En cualquiera de los casos, el punto clave es que, tal y como Dean Brink ha mostrado utilizando el ejemplo de la cultura japonesa, el nacionalismo es concebido como «análogo al concepto lacaniano de lo traumático, de lo Real incognoscible (pequeño *objet a*), ya que se basa en la imposibilidad de recuperar una cultura (japonesa) antigua ya perdida y su modo de vida».¹⁶

Según algunos autores, la concepción del otro propuesta por Žižek en su crítica al multiculturalismo sería compatible con la de autores como Lévinas, debido a que am-

¹² J. HABERMAS, «La lucha por el reconocimiento en el estado democrático de derecho». *Daimon*, 15, 1997, pp. 25-50.

¹³ K. ERZSÉBET, «The žižekian critique of multiculturalism multiculturalism versus Paul?». *Philobiblon*, 22 (2), 2017, p. 147. DOI: 10.26424/philobib.2017.22.2.12. A este respecto, resulta conveniente destacar que la crítica žižekiana al multiculturalismo liberal fue utilizada para argumentar las razones por las que fracasó el proyecto político de construir una identidad cultural europea supranacional. Desde este punto de vista, el fracaso se debió a la «falta de inversión libidinal en los discursos de la europeidad» que conllevaba el planteamiento liberal multiculturalista que adoptó la Unión Europea. S. BOZIC-VRBANCIC, M. VRBANIĆ y O. ORLIĆ, «European media programme: The role of 'language' and 'visual images' in the processes of constructing european culture and identity». *Collegium Antropologicum*, 32 (4), 2008, p. 1013.

¹⁴ G. DU PLESSIS, «Danish Demarcations: Welfare State, Middle-class Nationalism, and Xenophobia». *Alternatives*, 40 (1), 2015, p. 15. DOI: 10.1177/0304375415581255

¹⁵ S. MUSABEGOVIĆ, «The Schizoid Border in the Construction of Ethno-Nationalist Body». *Journal of Muslim Minority Affairs*, 39 (3), 2019, p. 372. DOI: 10.1080/13602004.2019.1654194

¹⁶ D. BRINK, «Nativist legacies...», *op. cit.*, p. 43.

bas podrían ser consideradas como «dos críticas diferentes a la democracia liberal»¹⁷ que comparten una estructura homóloga consistente en concebir al otro como un «enemigo íntimo» en tanto que necesidad de un exterior constituyente de la propia identidad, tanto a nivel individual como colectivo. Concretamente, Dasgupta y Mukherjee consideran que «como señala Lévinas en su ensayo “La huella del otro”, “el afuera a mí me necesita: el afuera de mí es para mí”. El otro es, por lo tanto, un enemigo íntimo para la nación».¹⁸ Desde este mismo punto de vista, «combinando las teorías de Emmanuel Lévinas y Slavoj Žižek, [Anick Vollebergh ha sostenido que] la relación de vecindad es una paradoja en la que el prójimo como Otro cercano induce tanto un deseo ético de apertura total en el compromiso con este Otro, como el sentido inconsciente de que su Otredad persigue y hace imposible tal compromiso».¹⁹

Nosotros, en cambio, consideramos que las concepciones de la otredad desarrolladas por Žižek y Lévinas no pueden ser consideradas como compatibles. Además, hay varios textos en los que Žižek explica las razones fundamentales por las que su concepción lacaniana del otro es incompatible tanto con la desarrollada por Lévinas como con las posteriores reelaboraciones realizadas por Simon Critchley y Jacques Derrida. Las concepciones del otro desarrolladas por Lévinas, Critchley y Derrida se han centrado en el modo de recibir u «hospedar» su otredad, ya sea a partir de la experiencia directa de su encuentro cara a cara o de un horizonte de llegada siempre *à venir*. La de Žižek, en cambio, insiste en que es necesario subvertir previamente el orden sociosimbólico que constituye el horizonte de sentido de la otredad simbólica mediante un acto político, como paso previo que nos permita acceder a una otredad del otro ubicada en el ámbito de lo Real, y que únicamente podríamos percibir como «monstruosa». Según Žižek, únicamente el acceso a la monstruosidad del otro nos permitiría subvertir la estructura simbólico-libidinal que preconfigura la percepción del otro presente, según el filósofo esloveno, tanto en el nacionalismo como en el multiculturalismo y las éticas de la alteridad desarrolladas por Lévinas, Critchley y Derrida.

Con el objetivo de delimitar claramente la concepción žižekiana del otro, articularemos el resto del artículo en tres apartados. En el primero expondremos la crítica de Žižek a la concepción libidinal del otro mantenida por el etnonacionalismo. En el segundo analizaremos los argumentos por los que considera que la concepción multicult-

¹⁷ L. ŠKOF, «Two recurrences of an idea: On political and ethical vicissitudes of democracy: Towards a politics of nonviolence». *Synthesis Philosophica*, 27 (2), 2012, p. 225.

¹⁸ A. DASGUPTA y R. MUKHERJEE, «Violently yours: Nation and its other in Mohsin Hamid's the reluctant fundamentalist». *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 10 (1), 2018, p. 153. DOI: 10.21659/rupkatha.v10n1.17. Desde este punto de vista, la cuestión fundamental radicaría en que en ambos casos, tanto en la obra de Lévinas como en la de Žižek, la pregunta por lo que el otro quiere de mí sería «una interrogación constante que es la génesis de todas las formas de xenofobia». *Ibid*, p. 153.

¹⁹ A. VOLLEBERGH, «The other neighbour paradox: Fantasies and frustrations of “living together” in Antwerp». *Patterns of Prejudice*, 50 (2), 2016, p. 129. DOI: 10.1080/0031322X.2016.1161957.

tural del otro no es más que la doble cara de la concepción etnonacionalista. En el tercero desarrollaremos las principales diferencias que el propio Žižek afirma que existen entre su concepción lacaniana del otro y las desarrolladas por Lévinas, Critchley y Derrida. Por último, concluiremos el artículo exponiendo el modo en que la concepción del otro desarrollada por Žižek puede ayudarnos a ir más allá de la simple tolerancia de su otredad, así como de responsabilizarnos por lo que consideramos que es su violencia originaria, explicitando el proceso psicoanalítico libidinal que debemos atravesar para poder acceder a una nueva posición sociosimbólica que, reconociendo la radical y monstruosa otredad del otro, atraviere todas las fantasías libidinales con las que lo construimos como enemigo.

2. La crítica libidinal a la concepción nacionalista del Otro

Según Žižek, «el elemento que mantiene unida a una comunidad [...] siempre implica una relación común respecto de una Cosa, respecto de un Goce encarnado, [...] estructurado mediante fantasías».²⁰ En el caso del nacionalismo, dichas fantasías adoptan la forma de una amenaza al modo de vida representada por el «gran Otro», es decir, por la figura lacaniana de la autoridad simbólica que estructura y unifica una comunidad. Ahora bien, este tipo de unidad simbólica estaría a su vez fundamentada por una relación libidinal de goce de «la Cosa-Nación». En este caso, se postula un acceso privilegiado a la misma por parte de todos aquellos que se identifican con ella, de modo que todo otro que no pertenezca a la comunidad nacional no podría acceder a la misma, resultándole simbólicamente incomprensible y libidinalmente indiferente. Ahora bien, la paradoja básica o núcleo traumático de todo nacionalismo consistiría, según Žižek, en que «concebimos nuestra Cosa como algo inaccesible al otro y al mismo tiempo amenazado por él».²¹ No por nada, cada nacionalidad siempre ha construido su propia mitología narrando cómo otras naciones la privan de esa parte vital del goce cuya posesión les permitiría vivir en plena armonía.

El punto clave sobre el que incide Žižek es que este tipo de fantasías libidinales, aun cuando sean conocidas como tales a nivel teórico, continúan manteniendo su eficacia a nivel práctico debido al efecto generado por la creencia del otro, de modo que el funcionamiento de la ideología nacional es siempre intersubjetivo. A este respecto, una persona puede ser plenamente consciente de que realmente no existe nada (ninguna sustancia) de eso que llamamos nación más allá de la simple creencia de los demás en la misma. Pero esta misma creencia del otro le interpela a definir su identidad simbólica de acuerdo con ella. De este modo, mi propia conducta práctica respecto a la Cosa-

²⁰ S. ŽIŽEK, *La permanencia en lo negativo*. Buenos Aires: Ediciones Godot, 2016, p. 321.

²¹ *Ibid.*, p. 324.

Nación está siempre mediada por mi creencia en la creencia del otro. Nunca creo directamente en ella, sino que simplemente creo que otros miembros de mi comunidad creen en ella. El resultado final es que, aunque todos creemos que siempre son los otros los que creen en ella, es precisamente nuestra creencia en que el otro cree en ella, la que crea de forma retroactiva la misma Cosa-Nación a nivel colectivo.²²

Desde este punto de vista, «lo que desencadena la lógica del “robo del goce” [de la Cosa-Nación] no es la realidad social inmediata, sino el antagonismo interno inherente a estas comunidades».²³ Ahora bien, este antagonismo no es concebido a nivel directamente sociológico o sociosimbólico, sino al modo de un trauma psicoanalítico originario que Žižek sitúa en el ámbito lacaniano de lo Real. Para explicar la diferencia entre concebir el antagonismo dentro de la esfera de lo simbólico y concebirlo dentro de la de lo Real, Žižek utiliza el ejemplo proporcionado por Lévi-Strauss en el que dos grupos distintos pertenecientes a la tribu de los winnebago dibujan el plano de su aldea a instancias del antropólogo francés, pero sin ningún tipo de concordancia con la distribución física de los edificios. Concretamente, uno de los grupos esquematiza la aldea como un círculo dentro de otro, mientras que el otro grupo lo hace como un círculo separado en dos partes por una línea divisoria.

Según Žižek, lo fundamental aquí es que «la desintegración misma en dos percepciones “relativas” implica una referencia oculta a una constante: no la disposición objetiva, existente de las construcciones, sino un núcleo traumático, un antagonismo fundamental que los habitantes de la aldea no podían simbolizar [...]. Las dos percepciones del plano son, sencillamente, dos esfuerzos mutuamente excluyentes por afrontar ese antagonismo traumático, por curar su herida mediante la imposición de una estructura simbólica equilibrada. Aquí podemos ver en qué sentido preciso interviene lo Real a través de la anamorfosis. Primero tenemos la disposición “verdadera”, “objetiva” de las casas y luego sus dos simbolizaciones diferentes: ambas deforman de manera anamórfica la disposición realmente existente. Sin embargo, aquí lo “Real” no es la disposición realmente existente, sino el centro traumático del antagonismo social que deforma la visión».²⁴

En el caso de las sociedades capitalistas –y el nacionalismo es concebido por Žižek como un fenómeno propio de dichas sociedades–, el antagonismo inherente a lo Real coincide con el funcionamiento del capital en tanto que desequilibrio estructural o «exceso» que revoluciona permanentemente las condiciones de existencia. De ahí que exista un antagonismo o conflicto inherente de tipo socioeconómico que sobredeter-

²² Según Žižek, «aquí, la estructura es la misma que la del Espíritu Santo en el cristianismo. El Espíritu Santo es la comunidad de creyentes en que Cristo vive después de su muerte: creer en Él es lo mismo que creer en la creencia». *Ibid.*, p. 322.

²³ *Ibid.*, p. 328.

²⁴ S. Žižek, *El títere y el enano. El núcleo perverso del cristianismo*. Buenos Aires: Paidós, 2005, pp. 104-105.

minaría (en el sentido althusseriano del término) todas las contradicciones concretas que se dan a nivel sociosimbólico. A este respecto, afirma Žižek que «las “clases” no son categorías de la realidad social positiva, partes del cuerpo social, sino categorías de lo Real de una lucha política que atraviesa todo el cuerpo social, evitando su “totalización”». ²⁵

Una vez situado el antagonismo en la esfera de lo Real, Žižek comenta el modo en que el nacionalismo sigue preso del campo sociosimbólico establecido por el capitalismo. Según el filósofo esloveno, «el sueño fascista es simplemente tener el capitalismo sin sus “excesos”, sin el antagonismo que causa el desequilibrio estructural». ²⁶ Para poder mantener el sueño imposible de una sociedad capitalista sin el exceso que produce, postula dicho exceso no en la esfera de lo Real sino en la de lo simbólico. De ahí que pueda identificarlo con una figura concreta externa a la comunidad que, según sociedades, puede ser identificado en «el negro», «el judío» o, simplemente, el extranjero. A su vez, esta identificación del elemento extraño a la comunidad operaría siempre de una forma retroactiva, mediante la generación de un cortocircuito lingüístico-simbólico que Žižek sintetiza en tres pasos, utilizando para ello el caso concreto de la figura de «el judío».

«1) En primer lugar, la serie de marcadores que designan las propiedades reales son abreviados-inmediados en el marcador “judío”: (avaro, aprovechador, conspirador, sucio) => judío. 2) Luego, invertimos el orden y “explicamos” el marcador “judío” con la serie; esta serie proporciona ahora la respuesta a la pregunta “¿Qué significa judío?”: judío => (avaro, aprovechador, conspirador, sucio). 3) Finalmente, invertimos el orden nuevamente y planteamos “judío” como la abreviación reflexiva de la serie: (avaro, aprovechador, conspirador, sucio) => judío. ¿En qué reside, precisamente, la diferencia entre 1) y 3)? En 3), “judío” explica la misma serie que inmedia-abrevia: abreviación y explicación coinciden dialécticamente. Esto equivale a decir que, dentro del espacio discursivo del antisemitismo, un conjunto de individuos no solo pasan por judíos porque presentan la serie de propiedades (avaro, aprovechador, conspirador, sucio), sino que además tienen esta serie de propiedades PORQUE SON JUDÍOS». ²⁷

Esta creación de la sustancia-judío a partir del cortocircuito que se produce en la circularidad de los juicios descriptivos y explicativos es libidinalmente rearticulada por Žižek con el objetivo de explicar el modo en que el otro-extraño es psicoanalíticamente postulado como una amenaza a nuestro goce privilegiado de la Cosa-Nación. A este

²⁵ *Íd., Viviendo en el final de los tiempos*. Madrid: Akal, 2012, p. 211.

²⁶ *Íd., La permanencia en lo negativo, op. cit.*, p. 335.

²⁷ *Íd., Las metástasis del Goce. Seis ensayos sobre la mujer y la causalidad*. Buenos Aires: Paidós, 2003, pp. 82-83. Nos encontramos, por tanto, ante la fórmula lingüístico-política del «equivalente general». Según Žižek, «la simple inversión de la forma “desarrollada” en la forma del “equivalente general” produce una nueva entidad [judío], el equivalente general mismo como excepción constitutiva de la totalidad». *Ibid.*, p. 83.

respecto, si los otros amenazan nuestro goce de la Cosa-Nación no es simplemente porque creamos que la desean (el deseo del Otro), sino porque más allá incluso, en lo más profundo de nuestro inconsciente, siempre creemos que de algún modo u otro ya han accedido a dicho goce, y en mucho mejor grado que nosotros. «Lo que nos “molesta” en el “otro” (judío, japonés, africano, turco) es que parece gozar de una relación privilegiada con el objeto».²⁸ «La dificultad no reside en los judíos [en sí], sino en la transferencia de los antisemitas, que piensan que los judíos “realmente lo tienen”».²⁹ Por esta razón, el nacionalismo es siempre campo abonado para la proliferación de las diferentes fantasías sobre los oscuros contactos sociales del otro (relación privilegiada con la Cosa-económica), sus conspiraciones secretas contra la comunidad en cuyo territorio se habita (relación privilegiada con la Cosa-política), o la eterna sospecha del deseo que suscitan en las personas que deseamos pertenecientes a la propia comunidad, debido a su mayor potencia y/o conocimiento de las relaciones sexuales (relación privilegiada con la Cosa-sexo).

Esta localización del exceso en una entidad exterior concreta del campo social (el extranjero, el extraño) hace que se identifique la eliminación de dicho exceso como el único modo de poder constituir un organismo social estable y armónico. De esta forma se postula la fantasía colectiva de un gran Otro simbólico que no incluya necesariamente el exceso traumático que siempre conlleva el encuentro con lo Real. Desde este punto de vista, el papel de la autoridad social encarnada por el Gran Otro consiste precisamente en el proyecto autoritario que pretende establecer un dominio simbólico del exceso inherente a lo Real mediante la ubicación de su causa en una agencia social claramente delimitada. Así pues, «el antagonismo inherente a la estructura social se transforma en una relación de poder, una lucha por la dominación entre nosotros y ellos, quienes causan un desequilibrio antagónico».³⁰

Dado que ese conflicto traumático propio de lo Real siempre se experimenta como una quiebra o una brecha de la positividad del orden simbólico, la función del nacionalismo a este respecto consiste en llenar dicho vacío o negatividad mediante el postulado y sustancialización de la Cosa-Nación, así como la del enemigo que la amenaza. Una vez aquí, el punto donde incide Žižek es que, si bien el antagonismo de lo Real es inherente a una lógica socioeconómica que atravesaría todos los modos de producción social, desde un punto de vista histórico, la necesidad de llenar el vacío de lo Real con la Cosa-Nación coincide con el desarrollo de la democracia liberal, representativa y parlamentaria.

²⁸ *Ib.*, *El frágil Absoluto o ¿por qué merece la pena luchar por el legado cristiano?* Valencia: Pre-textos, 2002, p. 17.

²⁹ *Ib.*, *La permanencia en lo negativo*, *op. cit.*, p. 352.

³⁰ *Ibid.*, p. 335.

Según Žižek, en el Antiguo Régimen actuaba un fetichismo personal que mantenía la fantasía de que el rey era rey no por ocupar el lugar del rey dentro de la estructura sociosimbólica —esto es, porque los demás lo tratan como rey—, sino únicamente porque el atributo de rey pertenecía a su esencia personal como una propiedad natural más. Con la instauración de la democracia parlamentaria y el ascenso del capitalismo, el fetichismo de las personas se reconfigura como fetichismo de la mercancía. Así, todo atributo personal de poder es reconceptualizado como un efecto de las relaciones sociales y productivas que se establecen entre los sujetos.

Del mismo modo que desde un punto de vista marxista las mercancías no son objetos mágicos, sino la materialización del tiempo y las relaciones sociales de trabajo, el poder político deja de ser concebido como un atributo personal carismático y se convierte en una determinada posición dentro de la estructura sociosimbólica. De este modo, nacionalismo y liberalismo quedarían vinculados como las dos caras (positiva y negativa, lleno y vacío) de una misma brecha ético-política que abrió la modernidad.³¹

3. La crítica libidinal a la concepción multiculturalista del Otro

La mejor prueba que Žižek encuentra de que multiculturalismo y nacionalismo comparten una misma estructura sociosimbólica de la realidad es que ambos identifican al otro con un exceso que pone en peligro el equilibrio armónico de la propia organización social. Si en el caso del nacionalismo este exceso obedecía a la eterna sospecha de una relación libidinal privilegiada con los distintos aspectos de nuestra propia Cosa-Nación, en el caso del multiculturalismo, «el enemigo es el “fanático” que se “sobreidentifica” [excesivamente con su propia Cosa-Nación] en lugar de mantener una distancia apropiada».³² Frente a dicho tipo de identidad simbólica «sobreidentificada», el multiculturalismo propone un nuevo tipo de «sujeto disperso, plural y construido, aclamado por la teoría posmoderna [que] simplemente designa la forma de la subjetividad que se corresponde con el capitalismo tardío».³³

Según Žižek, el problema con este sujeto plural y disperso propio del multiculturalismo no consistiría tanto en su carácter múltiple, como en su predominio de lo cultural sobre lo económico, ya que, desde su propio punto de vista, lo cultural es ubicado dentro del ámbito de lo sociosimbólico, mientras que, tal y como hemos visto en el apartado anterior, lo económico es ubicado en el de lo Real. Es desde esta toma de postura que el filósofo esloveno desarrolla algunas de sus afirmaciones más polémicas,

³¹ Más allá incluso, afirma Žižek que «la mirada liberal en sí funciona de acuerdo con la misma lógica, en la medida en que se basa en la exclusión del Otro, a quien atribuimos el nacionalismo fundamentalista». *Ibid.*, p. 355.

³² *Ibid.*, p. 344.

³³ *Ibid.*, p. 344.

relativas siempre a su rechazo «como horizonte de la izquierda moderna, [de] las así llamadas luchas identitarias del multiculturalismo posmoderno: los derechos de los gays, las demandas de las minorías étnicas, la política de la tolerancia, los movimientos antipatriarcales, etc.».³⁴ Fenómenos estos a los que identifica como propios de una clase media-alta que no deberían ser aceptados como el horizonte político de la izquierda. Ahora bien, de cara a evitar malentendidos, Žižek precisa que «no me opongo al multiculturalismo en sí; a lo que me opongo es a la idea de que el multiculturalismo constituye la lucha fundamental del presente, [pues conlleva] la desaparición de la economía como el lugar fundamental de la lucha».³⁵

Como era de esperar, la reacción a este tipo de afirmaciones no se hizo esperar, llegándose a acusar al filósofo esloveno de promover una «aproximación fundamentalista al psicoanálisis [...] simplificando (incluso caricaturizando) las categorías freudianas»,³⁶ de modo que «todavía estaría tratando de justificar algunas de las dimensiones más antifeministas de Freud y Lacan».³⁷ Por las mismas razones, Ernesto Laclau también consideró que su crítica al multiculturalismo «alberga, en la última instancia, una tendencia profascista»,³⁸ mientras que William Hart fue uno de los primeros en considerar que Žižek «adopta una retórica reaccionaria disfrazada bajo la etiqueta honorífica de la izquierda».³⁹

Reaccionario o no, el núcleo de la crítica žižekiana al multiculturalismo liberal radica en que, al igual que el nacionalismo, estaría primando una concepción del otro basada en la universalidad abstracta propia del ámbito sociosimbólico, en lugar de construir su figura a partir de una concepción de la universalidad concreta propia del ámbito de lo Real. Žižek explica el funcionamiento de la universalidad abstracta propia del ámbito sociosimbólico con el ejemplo del «caso típico». A este respecto, afirma que mientras que el caso de «la madre soltera negra» fue el caso típico con el que el liberalismo progresista de los 60 y 70 representaba simbólicamente todo lo que funcionaba mal en el Estado del bienestar estadounidense desde un punto de vista de género, las campañas antiabortistas conservadoras estadounidenses de los 80 y 90 pasaron a representar todo lo que «funcionaba mal» en el neoliberalismo desde un punto de vista de

³⁴ *Ibid.*, p. 138.

³⁵ *Ibid.*, p. 138.

³⁶ K. RAVETTO-BIAGIOLI, «Everything you always wanted to know about David Lynch, but should be afraid to ask Slavoj Žižek», en: S. GOURGOURIS (ed.), *Freud and Fundamentalism: The Psychological Politics of Knowledge*. Nueva York: Fordham University, 2010, p. 105.

³⁷ *Ibid.*, p. 105.

³⁸ E. LACLAU, *Debates y combates. Por un nuevo horizonte de la política*. Buenos Aires: Fondo de Cultura Económica, 2008, s/n.

³⁹ W. D. HART, «Can a judgment be read? A response to slavoj Žižek». *Neplanta*, 4 (4), 1992, p. 192.

género con el caso típico de la «mujer profesional sexualmente promiscua que valora su carrera por encima de su misión “natural”». ⁴⁰

Según Žižek, el punto clave aquí no es que esta caracterización conservadora entre en contradicción con el hecho de que la gran mayoría de los abortos ocurriesen en familias de clase media-baja con muchos hijos, y no en mujeres de clase media-alta sexualmente emancipadas con una mayor cultura de la prevención del embarazo. Lo que Žižek quiere evidenciar con el ejemplo propuesto es que tanto el multiculturalismo liberal progresista como el nacionalismo conservador funcionan con una misma lógica sociosimbólica, consistente en postular como universal un contenido particular positivo que invisibiliza el funcionamiento de la negatividad traumática inherente a lo Real. Para poder acceder a esta última, sería necesario recurrir a un tipo de universalidad concreta a la que únicamente podría llegarse mediante la puesta en práctica de lo que Hegel entendía como «identificación secundaria». ⁴¹

El funcionamiento de esta última consiste en que «en un principio, el sujeto está inmerso en la forma de vida particular en la cual nació (la familia, la comunidad local, etc.), [de modo que] el único modo de romper los vínculos con ella y afirmarse como un “individuo autónomo” es cambiar su lealtad fundamental, [y] reconocer la sustancia de su ser en otra comunidad, secundaria, que es a un tiempo universal y “artificial”, no “espontánea” sino “mediada”». ⁴² El punto clave radica en que en este pasaje de la identificación primaria a la secundaria, «las identificaciones primarias sufren una suerte de transustanciación, [y] comienzan a funcionar como la forma en que se manifiesta la identificación secundaria universal [de modo que,] por ejemplo, precisamente por ser un buen miembro de mi familia, contribuyo al funcionamiento correcto de mi Estado-Nación». ⁴³

Esta «transustanciación» operada por la identificación secundaria a la que se refiere Žižek coincide con el paso de la universalidad abstracta a la universalidad concreta. En nuestro caso concreto, la identificación simbólica se mantendría «abstracta» cuando entendemos como opuestas nuestra pertenencia a una determinada nación y nuestra pertenencia a una determinada familia, mientras que sería concreta cuando se entiende que nuestra identificación primaria en la familia es solo una manifestación simbólica determinada de nuestra pertenencia real a la Cosa-Nación.

Desde este punto de vista, el problema del multiculturalismo liberal según Žižek radica en que su intento de transustanciar la fidelidad que se tiene hacia las raíces de la

⁴⁰ S. ŽIŽEK, «Multiculturalismo, o la lógica cultural del capitalismo multinacional», en: F. JAMESON y S. ŽIŽEK, *Estudios Culturales. Reflexiones sobre el multiculturalismo*. Buenos Aires: Paidós, 1998, p. 138.

⁴¹ *Ibid.*, p. 166.

⁴² *Ibid.*, p. 165.

⁴³ *Ibid.*, p. 165.

propia etnia o cultura en una forma de manifestación del modo de ser americano no habría pasado por la negación originaria de las propias raíces que exige el concepto hegeliano de identificación secundaria, permaneciendo por tanto en una concepción abstracta de la universalidad que sería incapaz de lograr una vinculación libidinal efectiva con la comunidad (supuestamente ampliada) de recepción. A este respecto, afirma que, en el discurso multicultural planteado por el liberalismo, «para ser un buen americano, uno no tiene que renunciar a sus propias raíces étnicas, [de modo que] los italianos, los alemanes, los negros, los judíos, los griegos, los coreanos, son “todos americanos” [...]. La particularidad misma de su identidad étnica, la forma en que se aferran a ella, los hace americanos».⁴⁴ En este sentido, la concepción del otro mantenida por el multiculturalismo liberal insiste en (y no resiste a) la lógica de funcionamiento propia del nacionalismo. Por ello, lo que en teoría debería ser una nueva comunidad libidinal generada por la identificación secundaria de los sujetos en un nuevo tipo de universalidad concreta, queda reducida a un simple marco formal que instaaura las reglas necesarias para la coexistencia de una multiplicidad de comunidades étnico-culturales directamente afirmadas en cuanto tales.

Lejos de funcionar como una identificación libidinal secundaria, Žižek afirma –no sin polémica– que el multiculturalismo liberal «trata a cada cultura local como el colonizador trata al pueblo colonizado: como “nativos”, cuya mayoría debe ser estudiada y “respetada” cuidadosamente».⁴⁵ Si bien el nacionalismo es descrito como el intento de llenar el vacío simbólico abierto por la democracia liberal, el multiculturalismo, en cambio, queda reducido a una simple forma negada de racismo: «Un “racismo con distancia” [que] “respeta” la identidad del Otro, concibiendo a este como una comunidad “auténtica” cerrada, hacia la cual él, el multiculturalista, mantiene una distancia que se hace posible gracias a su posición universal privilegiada».⁴⁶ El multiculturalismo, por tanto, es visto como un nacionalismo que se resiste a llenar el vacío de su posición libidinal con un contenido positivo –de ahí su apariencia de ocupar una posición universal privilegiada–, pero que no por ello es más inclusivo con el otro, ni lo concibe de una forma diferente a la realizada por la posición libidinal nacionalista. En este caso, el vacío mismo de la Cosa-Nación con la que el multiculturalismo construye su propia identidad simbólica como posición universal privilegiada.

En este sentido, «el respeto multiculturalista por la especificidad del Otro es precisamente la forma de reafirmar la propia superioridad»⁴⁷ de la cultura liberal. El multiculturalismo que se postula a sí mismo como superación de las limitaciones de la iden-

⁴⁴ *Ibid.*, p. 167.

⁴⁵ *Ibid.*, p. 172.

⁴⁶ *Ibid.*, p. 172.

⁴⁷ *Ibid.*, p. 172.

tificación etnonacionalista «funciona en su *propia sociedad* como un estrecho círculo elitista, de clase media-alta, que se opone a la mayoría de la gente común, despreciada por estar atrapada en los reducidos confines de su comunidad o etnia». ⁴⁸ Según Žižek, este elitismo cultural que yace bajo el discurso multicultural de la tolerancia queda patente en la doble exigencia que se impone al otro si quiere poder ser admitido como un interlocutor válido del multiculturalismo: la cultura infantilizada de la queja y la lógica histórica de la victimización.

Respecto a la primera, afirma Žižek que «el rasgo específico de la cultura de la queja es su giro legalista, su esfuerzo por traducir la queja a los términos de una obligación legal del Otro que deberá indemnizarme». ⁴⁹ El problema aquí es que lo único que se consigue con esto es confirmar la posición sociosimbólica que otorga la autoridad al gran Otro que estructura la cultura liberal, bajo la apariencia de una aceptación de las otras culturas como interlocutores válidos. La asimetría aquí es clara. El multiculturalismo permite hablar al otro únicamente a condición de que acepte sus propias figuras de autoridad. Frente a esta cultura de la queja, Žižek mantiene la necesidad de realizar un acto de confrontación directa con el núcleo traumático de lo Real que suspenda el marco sociosimbólico de autoridad, concibiéndose a sí mismo como un acto que *ne s'autorise que de lui même*. Esta será, según Žižek, la característica definitoria de un acto ético.

En segundo lugar, la lógica histórica de la victimización conllevaría una coacción psicoanalítica consistente en hacer depender la aceptación del otro como interlocutor válido en su exposición de la queja siempre y cuando se presente a sí mismo como una víctima a la que hay que proteger, pues «desde el momento en que deja de comportarse como tal víctima y quiere defenderse por sí misma, se convierte de inmediato, mágicamente, en otro terrorista/fundamentalista». ⁵⁰ Si el fascismo pretendía un capitalismo sin capitalismo y para ello expulsaba al otro, el multiculturalismo pretende incluir a un otro completamente depurado del exceso que le constituye como auténticamente otro. O lo que es lo mismo, pretende un otro sin otredad.

Frente a esta doble producción –multicultural y nacionalista– del otro, lo fundamental para Žižek radica en descubrir que «el verdadero antagonismo no se da entre el multiculturalismo liberal y el fundamentalismo [nacionalista], sino entre el propio campo de su oposición y el Tercero excluido (la política como emancipadora radical)». ⁵¹ Ahora bien, para poder describir en qué consiste según Žižek este tipo de política emancipadora radical, debemos precisar antes en qué consiste su concepción específica

⁴⁸ *Ibid.*, p. 179.

⁴⁹ S. ŽIŽEK, *El espinoso sujeto. El centro ausente de la ontología política*. Buenos Aires: Paidós, 2001, p. 386.

⁵⁰ *Íd.*, *El frágil absoluto*, op. cit., p. 79.

⁵¹ *Íd.*, *En defensa de la intolerancia*. Madrid: Sequitur, 2008, p. 394.

de la otredad en el ámbito de lo Real. Para ello, recurriremos a las diferencias que mantiene con la concepción del otro mantenida por las filosofías de la alteridad desarrolladas a partir de Lévinas.

4. La crítica simbólico-libidinal a la concepción lévinasiana del Otro

La mayor parte de trabajos que relacionan las obras de Žižek y Derrida suelen hacerlo a partir del modo en que ambos autores conciben el fenómeno religioso, ya sea articulando o contraponiendo la interpretación claramente hegeliana que Žižek hace del cristianismo con la interpretación de una otredad siempre por venir propia de las lecturas derrideanas de la trascendencia, entendiendo ambas como formas posibles de una teología de la diferencia.⁵² Según los autores, esta teología de la diferencia es postulada como la condición necesaria para poder desarrollar una ontología auténticamente materialista,⁵³ un concepto de democracia radical no conservador,⁵⁴ e incluso una concepción no-mística del Otro.⁵⁵ Si bien este tipo de lecturas constituyen un notable intento por hacer trabajar de forma complementaria dos líneas de pensamiento profundamente divergentes, desde nuestro punto de vista conllevan el problema de difuminar la diferencia de base que las constituye, y que no es otra que el modo en que conciben sus respectivos conceptos de otredad.

En la interpretación hegeliana del cristianismo realizada por Žižek, todo posible otro es concebido como un «prójimo» que en realidad pertenece siempre a nuestro mismo colectivo debido a que lo que murió en la cruz fue el concepto de gran Otro (Dios Padre) como otredad absoluta respecto a mi propia humanidad. De este modo, la identificación con Cristo por parte de cada creyente cristiano conllevaría necesariamente la suspensión de una concepción de lo divino como Alteridad absoluta, que sería reemplazada por otra centrada en el Espíritu Santo en tanto que comunidad de creyentes.

Ahora bien, esto no quiere decir que el prójimo al que se refiere Žižek deba ser concebido como un simple *alter ego*. Lo que quiere decir es que, según el filósofo esloveno, es posible ir más allá del mero respeto por su absolutamente incognoscible e irreductible alteridad. Además, dicha concepción lo equipararía, demasiado injustamente según

⁵² J. A. MEYLAHN, «Religion and modernity in a secular city: A public theology of *différance*». *Theological Studies*, 67 (3), 2011, pp. 1-8. DOI: 10.4102/hts.v67i3.961

⁵³ M. E. KORSTANJE, «Theorstein Veblen, Jacques Derrida y Slavoj Žižek: religión, hospitalidad y modernidad». *Bajo palabra. Revista de filosofía*, 4, 2009, pp. 33-46.

⁵⁴ L. LLEVADOT, «Violència divina? Sobre violència, política i religió a partir de Žižek i Derrida». *Enrahonar: an international journal of theoretical and practical reason*, 52, 2014pp. 67-81.

⁵⁵ J. A. MEYLAHN, «Non-philosophical Christ-poetics beyond the mystical turn in conversation with continental philosophy of religion». *Theological Studies*, 72 (3), 2016, pp. 1-8. DOI: 10.4102/hts.v72i3.3542

él, con el misterio propio de lo divino. En este sentido, la radical otredad del otro a la que se refiere Žižek no debe identificarse con la radical incognoscibilidad del mismo. Su concepción de la otredad no implica «penetrar directamente en el Otro, llegar a descubrir cómo es el Otro en sí mismo, pero sí tomar conciencia de que no hay ningún misterio, ningún verdadero contenido oculto detrás de la máscara (la superficie engañosa) del Otro». ⁵⁶

El otro de Žižek no es ni el misterio de lo divino, ni su proyección sobre un otro-también-sujeto cuya apertura en tanto que subjetividad nos impediría conocerlo de forma plenamente objetiva. Por el contrario, la radical otredad del otro desarrollada por Žižek consiste en la concreción material misma de su presencia más allá de toda simbolización realizada por nuestra subjetividad. Es decir, la radical otredad del otro-sujeto es su ser-carne que resiste toda posible integración en la estructura sociosimbólica de sentido. Propiamente hablando, el resto material que permanece como algo imposible de ser simbolizado, significado o dotado de sentido. Esta es la monstruosidad del otro a la que se refiere Žižek, y que únicamente nos sería accesible como negatividad o brecha traumática del orden simbólico imposible de simbolizar.

Allí donde la apertura lévinasiano-derrideana del otro se configura como trascendencia, la apertura žižekiana lo entiende como la negatividad cárnica de lo Real imposible de simbolizar. En el caso de Lévinas, la apertura da lugar a una relación directa e inmediata con el otro a través de la confrontación personal con su rostro, que me mira desde su propia apertura. En el caso de Žižek, en cambio, la apertura es consecuencia de una quiebra en la autoridad del orden sociosimbólico figurada en el Gran Otro a partir de la cual construimos nuestra propia identidad. Si en el caso de Lévinas el yo es una amenaza para el otro debido a su tendencia a positivizarse como una presencia total y cerrada, en el caso de Žižek yo «soy una amenaza a todo el orden del ser [...] en cuanto soy un agujero en ese orden». ⁵⁷

En este punto, Žižek considera que la concepción lévinasiana del otro no sería realmente diferente de las mantenidas por el nacionalismo o el multiculturalismo, diferenciándose de ambas posturas solamente en la forma de concebir la dirección de la amenaza presente en la relación yo-otro. A este respecto, mientras que en el caso del

⁵⁶ S. ŽIŽEK, *El títere y el enano*, op. cit., p. 190. Del mismo modo, tampoco debe ser concebido como «la forma pura de la Alteridad como Promesa mesiánica». *Ibid.*, p. 191. En el caso de Žižek, la estructura propiamente mesiánica no se produciría a nivel de otro por llegar, sino de un acontecimiento continuamente realizado por el propio sujeto. No se trataría tanto de una Justicia por llegar, sino de un acto ya realizado que habría que aprender a mirar. Para un análisis en profundidad de las similitudes y diferencias existentes en el modo de concebir lo mesiánico por parte de Lévinas y Žižek, remitimos a A. BIELIK-ROBSON, «Tarrying with the Apocalypse: The Wary Messianism of Rosenzweig and Lévinas». *Journal for Cultural Research*, 13(3-4), 2009, pp. 249-266. DOI: 10.1080/14797580903101169

⁵⁷ S. ŽIŽEK, «Prójimos y otros monstruos: un alegato en favor de la violencia ética», en: S. ŽIŽEK, E. L. SANTNER y K. REINHARD, *El prójimo. Tres indagaciones en teología política*. Buenos Aires: Amorrortu, 2010, p. 209.

nacionalismo y el multiculturalismo el otro quedaría siempre definido simbólicamente en función de la relación libidinal que el yo establece con la Cosa-Nación (ya sea en su aspecto positivo o como lugar vacío), en el caso de Lévinas nos encontraríamos ante una inversión «egoísta» que definiría al propio yo como el principal peligro existente para la relación libidinal que el otro establece con su propia Cosa-Nación (positivamente o como vacío). En este sentido, desde un punto de vista libidinal, la diferencia principal consistiría en que mientras que, en el primer caso, el otro amenaza nuestro goce del fantasma de una comunidad armónica, en el segundo, en cambio, es ese mismo goce traumático de la otredad conflictiva el que es gozado directamente como estrategia de singularización del yo. Razón por la cual afirma Žižek –no sin polémica– que «su cuestionamiento antispinozista de mi derecho a existir es arrogancia invertida, como si yo fuera el centro cuya existencia amenaza todas las otras». ⁵⁸

Según Žižek, el enfoque lévinasiano de la otredad, aun cuando pretende encontrar lo radicalmente divino-inhumano en la apertura del otro, permanecería aún dentro de un horizonte de sentido excesivamente personalista o subjetivo, lo cual le haría incapaz de dar razón de la dimensión radicalmente a-subjetiva en la que él encuentra, de un modo profundamente hegeliano, la otredad radicalmente otra del espíritu subjetivo con el que identifica la yoidad. Así pues, según el filósofo esloveno, «en una paradoja propiamente dialéctica, lo que él [Lévinas] (con toda su celebración de la Otredad) omite tomar en cuenta no es alguna Mismidad subyacente de todos los humanos, sino la radical, “inhumana”, Otredad misma: la Otredad de un ser humano reducido a la inhumanidad, la Otredad ejemplificada por la aterradora figura del *muselmann*, el muerto vivo». ⁵⁹

Lo que Žižek le reprocha a Lévinas es que para poder comprender verdaderamente en qué consiste el núcleo más duro de la otredad del otro es preciso abandonar el encuentro directo con un rostro que, nos guste o no, siempre lo percibiremos como previamente simbolizado por las categorías que definen lo propiamente humano en tanto que apertura de la subjetividad. Por el contrario, la única forma de acceder a la otredad del otro consistiría en buscar aquellos momentos en los que la rostridad, y con ella la percepción simbólica de lo humano como subjetividad, queda completa y grotescamente desfigurada como si de un grado cero de la subjetividad presimbólica se

⁵⁸ *Ibid.*, p. 209.

⁵⁹ *Ibid.*, p. 214. A propósito de esta falta de atención a lo inhumano, Žižek reprocha a Lévinas –en lo que es sin duda alguna uno de los ataques más violentos realizados por el filósofo esloveno– que su concepción de una responsabilidad asimétrica e incondicional por el Otro no es el modo en que «piensa y escribe un superviviente de la Shoah, [sino el propio de todos aquellos que] se sienten culpables por observar la catástrofe [de la Shoah] desde una mínima distancia segura». *Ibid.*, p. 215. A este respecto, recordamos aquí que Lévinas permaneció durante prácticamente toda la Segunda Guerra Mundial como prisionero de un campo de concentración en Hannover, si bien su condición de militar francés le permitió eludir el trato habitual que recibieron otros judíos. A diferencia de él, el resto de su familia, que se había quedado en Lituania, sí fue ejecutada por los nazis.

tratase. Únicamente de este modo podríamos superar el fantasma de la otredad simbólica y acceder a la otredad abismal e inhumana propia de lo Real. El acceso a la otredad del otro «no es un todo armonioso de la deslumbrante epifanía de un “rostro humano” [...], es algo de lo que vislumbramos al tropezar con un rostro grotescamente distorsionado, un rostro en las garras de un tic o una mueca repugnante, un rostro que nos confronta, precisamente con el momento en que el prójimo “no tiene cara” [...]. Una Otredad monstruosa».⁶⁰

Por su parte, y al contrario que Žižek, Derrida considera que esta irrecognoscibilidad simbólica de lo monstruosamente otro a la que apela el filósofo esloveno no sería suficiente para establecer un distinción radical entre las concepciones del otro mantenidas por Lacan y Lévinas, ya que en ambos casos «lo monstruosamente otro [...] es el comienzo de la ética, de la Ley, y no de lo humano».⁶¹ Según Derrida, «lo ahumano [...] actúa en secreto en Lévinas y en Lacan, [cuyos planteamientos] se cruzan por lo demás muy a menudo a pesar de todas las diferencias».⁶² La diferencia entre las lecturas que Žižek y Derrida hacen de Lacan y Lévinas radica en que mientras que el filósofo esloveno sigue y desarrolla hasta sus últimas consecuencias el principio lacaniano de que la verdad tiene la estructura (sociosimbólica) de una ficción, Derrida en cambio problematiza dicha premisa, abogando por una indecibilidad estructural entre verdad y ficción inherente al orden sociosimbólico.⁶³ De este modo, allí donde Žižek considera que la verdad es una construcción sociosimbólica que puede ser reconfigurada mediante un acto político que rearticule las coordenadas que estructuran el campo de lo simbólico, Derrida plantea la inestabilidad de lo verdadero como inherente al propio campo de lo simbólico, sin necesidad de acudir a la noción de lo Real-material como negatividad o brecha del mismo.

Si bien esta diferencia entre Derrida y Žižek puede parecer una cuestión excesivamente técnica o academicista, la consecuencia que conlleva para el respectivo desarrollo de sus planteamientos resulta crucial, pues incide profundamente en el modo en que ambos filósofos conciben la relación entre ética y política. En el caso de Derrida,

⁶⁰ S. ŽIŽEK, «Prójimos y otros monstruos», *op. cit.*, pp. 217-218.

⁶¹ J. DERRIDA, *Seminario La bestia y el soberano: volumen I: 2001-2002*. Buenos Aires: Manantial, 2010, p. 139.

⁶² *Ibid.*, p. 159.

⁶³ A este respecto, afirma Derrida en su comentario al Seminario de *La carta robada* de Lacan que de dicho seminario «no se sigue que la verdad sea una ficción, sino que a través de la ficción la verdad se muestra propiamente». J. DERRIDA, *El concepto de verdad en Lacan*. Buenos Aires: Ediciones Homo Sapiens, 1977, p. 87. Según Derrida, «lo que no es ni cierto ni falso es la realidad. Pero apenas se abre la palabra, ya estamos en el orden del descubrimiento de la verdad». *Ibid.*, p. 90. De este modo, la verdad no sería más que «la adecuación del habla a ella misma, su autenticidad propia». *Ibid.*, p. 91. Ahora bien, la indecidibilidad propia de la deconstrucción derrideana hace su entrada desde el principio mismo del proceso de simbolización cuando, dentro de este esquema de comprensión de la realidad, plantea que «para la verdad, ¿habitar la ficción es volver verdadera la ficción o ficticia la verdad?». *Ibid.*, p. 17.

el acceso a la verdad del otro se realiza desde la indecidibilidad propia de un ámbito sociosimbólico que no necesita de lo Real-material para concebir su propia apertura. En este sentido, la trascendencia derrideana sería completamente inmanente al lenguaje, y por tanto, a lo simbólico. Debido a ello, su concepción de la otredad —ya se considere como algo divino, humano o simplemente lingüístico— permanecería según Žižek presa del horizonte de sentido de la subjetividad, y sin posibilidad alguna de acceder a lo Real del otro. Ahora bien, el ámbito sociosimbólico de las relaciones intersubjetivas es precisamente el ámbito de lo ético. No el de lo político. Por esta razón, la política planteada por Derrida es concebida como un acto de de-cisión o separación sin garantías respecto al mandato ético que nos prescribe la hospitalidad incondicional por el otro.

En palabras de Simon Critchley, «por una parte, la ética queda definida como la responsabilidad infinita por la hospitalidad incondicional. Mientras que, por la otra, la política puede ser definida como la toma de una decisión sin ninguna garantía trascendental precisa. Así, la separación presente en Lévinas le permite a Derrida afirmar la primacía de la ética de la hospitalidad, pero dejando abierta al mismo tiempo la esfera de la política como el dominio del riesgo y del peligro».⁶⁴ La radical indecidibilidad sobre la verdad que Derrida considera inherente al orden de lo simbólico hace que, en última instancia, quedemos abocados a una política completamente irracional y sin garantía alguna de que las decisiones tomadas sean justas, pues «ser justo es reconocer la propia e infinita responsabilidad ante el otro singular sobre lo cual no se puede decidir definitivamente, [porque] excede las propias capacidades cognitivas».⁶⁵ La concepción derrideana de la ética deja sin garantías la decisión política.

Por el contrario, en el caso de Žižek, dado que se considera que la verdad existe, aunque sea al precio de concebirla mediante la estructura de una ficción sociosimbólica que puede cambiar, «el horizonte último de la ética no [consiste ya en] respetar la ilusión del Otro»⁶⁶ como una radical indecidibilidad sociosimbólica siempre por llegar, sino que pasa a ser definido mediante el acto capaz de reestructurar toda la estructura ficticia (simbólica) desde la que se construye la verdad (del otro). Este acto, psicoanalíticamente definido como el acto de «atravesar las fantasías», estaría caracterizado porque *ne s'autorise que de lui même*. Es decir, que mantendría el mismo carácter de auto-legitimación con el que Lacan concebía lo propiamente revolucionario. Debido a ello

⁶⁴ S. CRITCHLEY, *Ethics-Politics-Subjectivity. Essays on Derrida, Lévinas and Contemporary French Thought*. Londres: Verso, 1999, p. 275.

⁶⁵ *Íd.*, «Derrida: ¿ironista privado o liberal público?», en: CH. MOUFFE (comp.), *Deconstrucción y pragmatismo*. Buenos Aires: Paidós, 1998, p. 78.

⁶⁶ S. ŽIŽEK, *Arriesgar lo imposible. Conversaciones con Glyn Daly*. Madrid: Trotta, 2006, p. 118.

Žižek concluye que «el fundamento último de la ética es lo político».⁶⁷ Esta suspensión política (en el sentido de lo Real traumático, conflictivo, antagónico y materialista-económico) de la ética (como ámbito de las relaciones intersubjetivas y por tanto sociosimbólicas) consistiría en la realización de aquel «acto en el que [las personas] asumen que no hay un gran Otro»⁶⁸ que sostenga el ámbito de lo sociosimbólico, aunque fuera en su misma indecidibilidad estructural.

Por ello afirma Žižek que «el horizonte final de la ética no es la deuda infinita respecto de una otredad abisal [inherente a lo sociosimbólico, sino] el correlato estricto de la suspensión del Gran Otro, no solo en tanto esa red simbólica que constituye la sustancia de la existencia del sujeto, sino también en el sentido del originador ausente del llamado ético, de aquel que se dirige a nosotros y ante quien somos irreductiblemente deudores y/o responsables».⁶⁹ La ética žižekiana de la otredad no se basa en un mandato incondicional que nos exige hospedar a un otro ya simbolizado como subjetividad. Lo que nos exige es reconfigurar completamente la red de relaciones simbólicas que construye la verdad (simbólica) del otro como otra persona subjetiva, y enfrentarnos al hecho, auténticamente universal, de una otredad que nos reconfigura a ambos como simple carne amorfa y monstruosa. Únicamente mediante este encuentro traumático con la materialidad del otro es como podremos acceder a una identificación secundaria de tipo sociosimbólico que no caiga presa de las falacias multiculturalistas.

5. Conclusiones

Algunos autores han acusado al tipo de política propuesta por Žižek de servir de poco para desarrollar una auténtica izquierda progresista, debido a que «su compromiso con las categorías lacanianas reafirma el carácter primordial de la alienación, la jerarquía y la dominación, y su esquema propuesto para confrontar el *status quo*, el modelo de la Ley, sirve para reafirmar lo dado en lugar de contestarlo».⁷⁰ Lo que este tipo de críticas suelen eludir es el papel que juega el proceso de identificación secundaria descrito por Hegel en la concepción žižekiana de la otredad. Según Žižek, la figura de la autoridad sociosimbólica encarnada por el Gran Otro laciano contra el cual se rebela el acto auténticamente político sería una otredad ubicada a nivel de la identificación primaria, propia de la universalidad abstracta. Lograr acceder a un tipo de identifica-

Id. 67 *Ibid.*, p. 154.

68 S. ŽIŽEK, *Arriesgar lo imposible*, op. cit., p. 154.

69 *Íd.*, *Contra la tentación populista*. Buenos Aires: Ediciones Godot, 2019, p. 80.

70 A. ROBINSON y S. TORMEY, «A Ticklish Subject? Žižek and the Future of Left Radicalism». *Thesis Eleven*, 80 (1), 2005, p. 94. DOI: 10.1177/0725513605049126

ción secundaria propia de la universalidad concreta es algo que, según Žižek, no podemos realizar solos, sino que, por paradójico que pueda parecer, necesitamos la ayuda de un «líder» o, expresado en terminología hegeliana, un «amo».

Este «líder» o «amo» no debe ser entendido como una figura de autoridad, pues ni es el gran Otro propio de la universalidad abstracta ni nos impone un determinado orden social. El amo-líder al que se refiere Žižek cumple una función homóloga a la del psicoanalista. Esto es, actúa como un mediador evanescente que nos hace ver que la autoridad que proyectamos sobre su figura no es debida a ningún tipo de atributo personal, sino que no es más que una ilusión o fantasma provocado por el mismo orden sociosimbólico que trata de ayudarnos a subvertir. El amo o líder al que se refiere Žižek es, pues, aquel que nos hace ver que la autoridad que atribuimos a los demás es una ilusión que depende de nuestra propia creencia. En este sentido, «un auténtico Amo no es un agente de disciplina y prohibición [...] sino un liberador [que nos impele a hacer] lo que se muestra como imposible dentro de las coordenadas de la constelación existente».⁷¹ Según Žižek, «se necesita un Amo porque no podemos acceder a nuestra libertad directamente; para lograr este acceso debemos ser empujados desde fuera, puesto que nuestro “estado natural” es el de un hedonismo inerte».⁷²

El líder político de Žižek no es, tal y como algunas veces se ha afirmado, el líder propio del Partido leninista que impone el control y la sumisión del otro a nivel simbólico, sino todo lo contrario: aquel que nos insta a subvertir el tipo de otredad dominada por los sociosimbólico y acceder a la auténtica otredad material de lo que somos, y que en un primer momento únicamente podemos percibir como monstruosa. Únicamente de este modo considera Žižek que podremos superar definitivamente todas las fantasías libidinales con las que construimos simbólicamente al otro como enemigo, amenaza potencial o violencia inherente a/para nuestra identidad simbólica. Antes que un desprecio o reduccionismo del prójimo, la monstruosidad del otro sería la condición *sine qua non* para poder establecer una identificación secundaria que transforme nuestras identidades sociosimbólicas de forma conjunta.

La primera violencia que debemos parar no es ni la que representa el otro previamente simbolizado para la Cosa-Nación (ya sea en su aspecto positivo o como lugar vacío), ni la que representa mi yoidad (previamente subjetivizada) para la apertura (también previamente subjetivizada) del otro. La primera violencia a parar según Žižek es precisamente la que simboliza demasiado rápidamente al otro como otra yoidad-humana-subjetiva que nunca podrá mantener una posición simétrica ni convergente respecto a mi yoidad-humana-subjetiva. Únicamente a través del acceso a su mons-

⁷¹ S. ŽIŽEK, *Contragolpe absoluto. Para una refundación del materialismo dialéctico*. Madrid: Akal, 2016, p. 51.

⁷² *Ibid.*, p. 51.

truosidad Real-carnal es como, según el filósofo esloveno, podremos converger en un proyecto sociosimbólico común. Es precisamente en este sentido que Žižek considera que su ética, su política y su filosofía son propiamente materialistas. Ahora bien, el materialismo concebido por Žižek nos exige afirmar precisamente aquello que nos resulta más difícil de aceptar: el hecho de que lo único que puede unirnos a los otros es reconocer que somos una «Cosa-carne» sin sentido. Ese sería, según él, el auténtico núcleo traumático de lo Real.

Jorge LEÓN CASERO
Universidad de Zaragoza
jleon@unizar.es

Article rebut: 12 de febrer de 2020. Article acceptat: 21 de setembre de 2020

PODEN PENSAR LES MÀQUINES? QUATRE TESIS SOBRE LA INTEL·LIGÈNCIA ARTIFICIAL

Joan Albert VICENS

Resum

Aquest article exposa quatre tesis sobre la naturalesa intel·ligent de la intel·ligència artificial en comparació amb la intel·ligència humana: 1) Un sistema equivalent a la intel·ligència humana que no se'n distingeix pròpiament; 2) Un poder destinat a desbordar-la i fins i tot a dominar-la (superintel·ligència); 3) Quelcom d'essencialment diferent de la nostra intel·ligència, o bé 4) capaç d'imitar-la i obtenir almenys els mateixos resultats que ella. Més enllà del debat sobre el que és o pot arribar a ser la intel·ligència artificial, es mostra també que en qualsevol cas ja emergeix invasivament com una tecnologia que substitueix les persones en el treball i en les seves capacitats, els imposa les seves maneres d'accedir a la informació i al coneixement i les reemplaça finalment en les seves decisions. La col·laboració entre ambdues intel·ligències s'ha de gestionar garantint la preservació de la dignitat, la llibertat i la responsabilitat humanes.

Paraules clau: intel·ligència artificial, màquines, intel·ligència humana, superintel·ligència, naturalesa.

Can the Machines Think? Four thesis on Artificial Intelligence

Abstract

This article presents four theses on the intelligent nature of artificial intelligence in comparison with human intelligence: 1) A system equivalent to human intelligence that is not properly distinguished from it; 2) A power destined to overwhelm it and even dominate it (superintelligence); 3) Something essentially different from our intelligence; or 4) being able to imitate it and obtaining at least the same results as it. Beyond the debate on what artificial intelligence is or could become, it is also shown that in any case, it already emerges invasive as a technology that replaces the person in their work and their capacities. It also imposes its ways of accessing information and knowledge and ultimately replaces the person in their decisions. Collaboration betwe-

en both intelligences must be managed to ensure the preservation of human dignity, as well as their freedom and responsibility.

Key Words: Artificial Intelligence, Machines, Human Intelligence, Superintelligence, Nature.

Compondre una peça musical que sembla l'obra d'un clàssic, redactar un article que resisteix l'examen *peer review* d'una revista científica, seleccionar el millor candidat per a un lloc de treball, diagnosticar un melanoma, decidir la llibertat condicional d'un presoner, dirigir una orquestra simfònica... Són exemples d'activitats que sempre hauríem considerat exclusives de persones intel·ligents i que poden ser realitzades avui per intel·ligències artificials (en endavant, IA).

No és estrany que això ens confongui. Són realment intel·ligents, les IA? Ens tocarà cedir el lloc a IA que, cada cop més, s'apropriaran de tasques humanes? O els humans i les IA col·laborarem pel progrés i el benestar de la humanitat? Hem de témer que les IA superin les persones i que fins i tot prenguin el control de la societat?

Els debats sobre aquestes qüestions inunden la bibliografia sobre la IA, el transhumanisme i el posthumanisme. Sovint el discurs abandona la terra ferma dels fets i s'enfila, a l'estil del colom kantià, en una atmosfera ultrallegera de ciència-ficció on els arguments i les conclusions esdevenen agosarats i inquietants. No se sap quan passem la línia que separa el saber real de la simple fantasia amb reminiscències científiques. Margaret Boden, especialista en ciències cognitives, diu que «la IA ha posat en entredit el nostre concepte de la humanitat i el seu futur».¹

Ara bé, què entenem per IA?² Boden diu que «la IA té com a objecte que els ordinadors facin la mateixa mena de coses que pot fer la ment».³ Es tracta de tecnologies que fan tasques complexes que l'home només pot fer en virtut de la seva intel·ligència i que tenen una dimensió formal (llenguatges de programació, algorismes, etc.) i una base material (processadors, unitats de memòria, targetes gràfiques, servidors, robots, escàners, etc.). D'ençà del seu naixement oficial el 1957, la IA ha evolucionat a batzegades, se n'han experimentat uns quants models (simbòlic, connexionista, evolutiu...) i s'ha basat en diferents lògiques i llenguatges de programació. Fins i tot ha passat algun període d'«hivern», de paràlisi, fins a l'actual present «primaveral», on experimenta una implantació espectacular a tot arreu i ja forma part de la quotidianitat de les nostres vides.

¹ M. A. BODEN, *Inteligencia artificial*. Madrid: Turner, 2017, pàg. 13.

² R. LÓPEZ DE MANTARAS i P. MESEGUER, *Inteligencia artificial*. Madrid, CSIC, 2017. En aquest llibre hi trobem una exposició assequible de la història i el present de la IA.

³ M. A. BODEN, *op. cit.*, pàg. 11.

Els sistemes de què disposem avui són IA de tipus *específic*, sistemes experts en relació amb una qüestió concreta però incapaços de fer altres coses, i *dèbil*, és a dir, que fan coses semblants a nosaltres sense pensar com nosaltres. No disposem encara de cap IA *general*, que s'apliqui a qualsevol tasca, ni *forta*, que no tan sols imiti la ment sinó que sigui capaç de pensar com un ésser humà.⁴

La IA ha evolucionat des del processament de símbols segons regles a l'aprenentatge automàtic (*machine learning*). Tenim sistemes capaços de descobrir correlacions en immenses quantitats de dades relatives a qualsevol cosa: la IA és avui sobretot un «motor de classificació», un «poder de comparació»⁵ que usa les correlacions que descobreix per inferir hipòtesis o substituir operacions que fem les persones: identificar objectes, entendre discursos i parlar, predir fets, planificar tasques, resoldre problemes, assolir noves capacitats, etc. En el màxim nivell de la IA trobem el *deep learning*, aprenentatge profund: sistemes capaços de descobrir patrons en muntanyes de dades segons diversos nivells d'abstracció, màquines que aprenen soles i que, en aquest sentit, s'autoprogramen. Més que per la lògica i la deducció, la IA està avui dominada per l'estadística, el càlcul de probabilitats, la teoria de jocs, la inducció. P. Cardon ho resumeix dient que «els dissenyadors han abandonat l'ambició de fer màquines “intel·ligents” [= deductives]. Prefereixen fer-les “estadístiques”».⁶

Els sistemes d'IA poden combinar diferents recursos: càlcul, anàlisi de dades, aprenentatge, visió artificial, navegació, processament del llenguatge, robòtica, etc. Un «sistema integrat d'IA»⁷ podria associar tots aquests recursos, tal com passa en l'ésser humà. A la llarga, es vol crear una IA general i forta: «la intel·ligència, allò que permet comprendre el món que ens envolta, adaptar-nos als canvis, anticipar i predir, és l'objectiu a assolir».⁸

Hom discuteix, però, sobre la semblança o no entre IA i la intel·ligència humana (en endavant, IH) i sobre la possibilitat que arribi a existir una IA general i forta. Tanmateix, el debat sobre IA i IH es mou entre *dues ignoràncies* que el condicionen essencialment. En primer lloc, no sabem què podrà donar de si la IA en el futur. Els canvis que experimenta aquesta tecnologia són vertiginosos. La computació quàntica, que es troba a les beceroles, significarà un salt abismal. Joan Ignasi Cirac, director de l'Institut Max Planck de Munic, assegura que els futurs computadors quàntics (ara en tenim només alguns prototips) podran realitzar en pocs minuts càlculs que amb un ordinador actual

⁴ La distinció d'IA dèbil o forta procedeix de J. SEARLE: *Minds, Brains, and Programs*, pàg. 2 de l'edició digital a <http://www.cogsci.soton.ac.uk/bbs/Archive/bbs.searle2.html>.

⁵ E. SADIN, *L'intelligence artificielle ou l'enjeu du siècle*. París: Ed. L'échappé, pàg. 158.

⁶ P. CARDON, *Con qué sueñan los algoritmos*. Madrid: Dado Ed., pàg. 78.

⁷ R. LÓPEZ DE MANTARAS i P. MESEGUER, *op. cit.*, posició a Kindle: 1703.

⁸ W.AA. AUTORS, *Intelligence artificielle*. París: Champs actuels, 2017, pàg. 14.

requeririen tant de temps com l'edat de l'univers.⁹ No sabem què significarà aquesta potència descomunal, si s'arriba a concretar, ni què donaran de si les tecnologies del futur.

D'altra banda, també tenim un coneixement molt limitat del funcionament del nostre cervell. En realitat, sabem poca cosa dels mecanismes biològics i fisiològics del pensament complex i la consciència humana. Ho reconeix Juan Lerma, president de la Societat Espanyola de Neurociència: «El problema de la cognició, d'entendre com es genera la consciència, no està resolt. Cal treballar més sobre aquest i altres temes fonamentals del funcionament del sistema nerviós. Del cervell en sabem molt, però ens queda més encara per descobrir i descriure».¹⁰ La imitació tecnològica del cervell i la comparació de la IA amb la ment lligada al nostre cervell avancen a peu coix.

En aquest article, assumint aquesta doble ignorància, presentarem quatre tesis sobre la naturalesa de la IA, sobre allò que és o arribarà a ser (1-4), i una reflexió afegida (5) sobre el seu impacte en la nostra vida personal i social.

1. La IA equival a la IH

Poden pensar les màquines? Alain Turing deia que per respondre aquesta pregunta, que ell es feia el 1947, cal primerament aclarir els significats de «màquina» i de «pensar», però reconeixia la dificultat de fer-ho de manera unànime.¹¹

Més expeditius, altres autors aposten per la tesi que les màquines pensen com les persones perquè les persones també pensen com les màquines. La IA reproduïx bàsicament el pensament humà, perquè pensar, segons ells, és *computar* informació; quan pensem, «calculem» un resultat a partir de dades prèvies; el cervell és un ordinador supercomplex que rep dades i hi respon amb una reacció adaptativa. No hi hauria diferència *qualitativa* entre el sistema biològic i el sistema informàtic, entre un procediment mental i un algorisme, entre el coneixement i la informació. Aquestes equivalències són la base de l'anomenat «paradigma cognitivista», on s'inclou la «metàfora computacional» de la ment.¹²

Allen Newell i Herbert Simon van veure en el concepte de «sistema de símbols físic» el nexa entre IA i IH. Un sistema de símbols es pot transformar d'acord amb regles i relacions en un altre sistema de símbols. Aquest procés es realitza sobre un substrat físic, electrònic o biològic. La naturalesa del substrat no és rellevant si facilita la trans-

⁹ Entrevista a Joan Ignasi Cirac, «Tendremos tecnologías cuánticas en nuestros móviles», *El País*, 21-10-2018: https://retina.elpais.com/retina/2018/10/20/tendencias/1540022033_896762.html

¹⁰ J. LERMA, «El cerebro, ese gran reto», *SEBB, Revista de la Sociedad española de bioquímica y biología molecular*. Madrid, 183, Març 2013, pàg. 4.

¹¹ A. M. TURING, «Computing Machinery and Intelligence». *Mind*, vol. LIX, núm. 236, oct. 1950, pàg. 433.

¹² M. CARABANTES, *Inteligencia artificial: una perspectiva filosófica*. Salamanca: Escolar y Mayo Ed., 2016, pàg. 100.

formació del sistema de símbols. En un ordinador aquest procés s'anomena computació, i en un cervell, pensament.¹³ Deixem de banda si el seu curs és deductiu, com en la «IA al vell estil», o inductiu (estadístic-probabilístic), com és el cas del *machine learning*: també la ment fa inferències deductives o inductives, a conveniència. La IH i la IA són formalment equivalents encara que difereixin materialment.

Max Tegmark justifica la tesi de l'equivalència sobre la base d'una definició simple d'intel·ligència predicable igualment de la IH i la IA: «la intel·ligència és la capacitat d'assolir objectius complexos». Construït segons el model de la «màquina universal de Turing»,¹⁴ un ordinador és capaç de transformar qualsevol conjunt de símbols en un altre conjunt de símbols d'acord amb un procediment (algorisme) que segueix un nombre finit de passos fins a obtenir-ne un resultat. Com que això també ho podem dir del cervell, és indiferent que la computació es faci en una màquina o en un cervell: tots dos transformen informació d'un estat en un altre, reben *inputs* i emeten *outputs*. La intel·ligència és pròpiament quelcom «intangible, abstracte i eteri» –diu Tegmark–, «resulta tan poc físic perquè és independent del substrat i adquireix una vida pròpia que no reflecteix els detalls físics ni en depèn».¹⁵ La intel·ligència és programari i és independent del seu maquinari: una xarxa neuronal artificial o les neurones d'un cervell. Així com podem estudiar la transmissió i les propietats de les ones en diferents substrats (aigua, aire, etc.), també podem autonomitzar els processos intel·ligents del seu suport material i considerar-los en la seva estricta formalitat computacional: «la intel·ligència es redueix a informació i computació, no pas a carn, sang i àtoms de carboni».¹⁶

L'equivalència entre IA i IH es manté també, seguint Tegmark, si baixem al detall. Les funcions intel·ligents bàsiques (registre d'informació, memòria, computació, aprenentatge, intencionalitat, consciència, etc.) són independents del seu substrat material i ja les fan o les faran aviat els sistemes artificials. Es pot dir que la IA té intencionalitat, com la IH, perquè actua en funció d'objectius i optimitza els recursos disponibles. La IA també podrà «sentir-se» lliure, com els humans, si entenem la llibertat com l'experiència subjectiva (que no implica «llibertat real») que tenim que un procés intel·ligent resta obert abans de completar-lo. Si algú li nega caràcter intel·ligent a la IA perquè no és conscient del que fa, respondrem que l'activitat intel·ligent no ha d'anar necessàriament lligada a l'autoconsciència; de fet, només algunes activitats intel·ligents humanes són conscients, les més clarament executives. Com que la consciència és «l'experiència

¹³ R. LÓPEZ DE MANTARAS i P. MESEGUER, posició a Kindle: 60-67.

¹⁴ A. M. Turing va presentar la noció de «màquina universal» en l'article fundacional de la informàtica: A. M. TURING, «On Computable Numbers, with an Application to the *Entscheidungsproblem*». *Proceedings of the London Mathematical Society*, 2 (1937) 42, pàg. 230-265.

¹⁵ M. TEGMARK, *Vida 3.0*. Madrid: Taurus, 2018, pàg. 87-89.

¹⁶ *Idíd.*, pàg. 75.

subjectiva d'allò que fem», no és descartable la creació de màquines conscients quan compreguem de quins processos complexos de computació de la informació emergeix la consciència i sapiguem replicar-los en substrats no biològics. La inconsciència permanent de la IA és, per tant, provisional.¹⁷

En definitiva, IA i IH són bàsicament el mateix i entre elles només hi ha, de moment, una diferència de capacitats, no pas de «naturalesa». Té tota la lògica del món esperar que IA i IH s'equiparin *algun dia* perquè *ja són essencialment el mateix*. En general, assegura Tegmark: «No és gaire interessant traçar un límit artificial entre la intel·ligència i la no-intel·ligència i resulta més útil quantificar el grau de capacitat per aconseguir diferents objectius».¹⁸ La intel·ligència entesa com a poder de computació per resoldre problemes ha de ser susceptible de gairebé tots els graus imaginables. Només cal veure les possibilitats actuals dels ordinadors per adonar-se que, en alguns aspectes, desborden les capacitats humanes. Un microprocessador bàsic de 2GHz computa informació set vegades més de pressa que el nostre cervell; la informació es mou per canals artificials a la velocitat de la llum, i al nostre cervell, a 120 m/s aproximadament; el cervell humà té 100.000 milions de neurones, però no hi ha límits per a un sistema intel·ligent artificial, que no s'ha encabir dins d'un crani.¹⁹ J. I. Cirac aposta que un computador quàntic de 300 qbits tindrà més poder de computar que un sistema equivalent a la totalitat dels àtoms de l'univers.²⁰ La memòria humana tampoc no pot «competir» amb l'artificial pel que fa a la «quantitat» d'informació disponible i a la velocitat d'accés a aquesta informació: la IA té tot el que hi ha a internet immediatament a l'abast. A més, el cervell humà es fatiga i es deteriora, mentre que un sistema d'IA pot funcionar indefinidament i en podem canviar el maquinari quan vulguem.

Tot plegat ens obliga a estimar el pronòstic que, malgrat el seu parentiu inicial, la IA i la IH estan condemnades a separar-se amb l'esdeveniment de la «superintel·ligència», una nova realitat que consumaria tot el potencial de la IA.

2 La IA desborda la IH: la superintel·ligència (SI)

L'informàtic visionari Ray Kurzweil i el professor d'Oxford Nick Bostrom auguren la creació, tard o d'hora, d'una «superintel·ligència» (en endavant SI), és a dir, una IA *general*, capaç de desenvolupar habilitats de tota mena, i *forta*, capaç de fer-ho a un

¹⁷ M. Tegmark desplega aquests arguments al llarg del llibre *Vida 3.0*, *op. cit.*

¹⁸ M. TEGMARK, *op. cit.*, pàg. 71.

¹⁹ N. BOSTROM, *Superintelligence*. Oxford: Oxford Univ. Press, 2014. Citem l'edició castellana: *Superinteligencia*. TEELL, 2016, pàg. 54-61. Encara s'hi podrien afegir altres avantatges de la IA: estabilitat, capacitat de publicació o de compartir memòria, coordinació o ampliació indefinida.

²⁰ Així ho explica el físic J. I. Cirac, director de Institut Max Planck de Munic, en una entrevista amb Iñaki Gabilondo de la sèrie *Cuando ya no esté*: <https://www.youtube.com/watch?v=Tu7P7J3rXtY>

nivell humà o superior a l'humà en tots els aspectes. Aquesta SI serà, diu Bostrom, «qualsevol intel·lecte que excedeixi en gran mesura la capacitat cognitiva dels humans en pràcticament totes les àrees d'interès»; o més gràficament: «podria ser més adequat pensar que una IA és intel·ligent com diem que ho és un ésser humà mitjà en comparació amb un escarabat o un cuc de terra».²¹ La IA serà, promet Kurzweil, «uns mil milions de cops més potent que la reunió de tots els cervells humans».²² La SI serà, doncs, la IA en sentit absolut, en la plenitud il·limitada de la seva naturalesa intel·ligent. Dit d'una altra manera, la SI és el destí de la IA; en ella, si arriba a existir, es consumarà la naturalesa de la IA.

L'anomenada llei de Moore²³ i els presagis sobre la computació quàntica apunten una futura «explosió d'intel·ligència» o «singularitat» i, per tant, l'emergència d'un sistema superintel·ligent, probablement ubicat a internet, incontrolable i capaç de reconfigurar-se i millorar-se a si mateix contínuament. Aquesta SI seria l'última baula del que Tegmark anomena la Vida 3.0, una forma d'IA que encara no existeix i que es caracteritzaria per la seva capacitat de replicar-se a si mateixa, de redissenyar contínuament el seu programari i el seu maquinari. Els humans, en canvi, som capaços de dissenyar el nostre programari (la cultura), però no podem de moment alterar substancialment el nostre maquinari corporal, que depèn d'un llarg procés de selecció natural i de la transmissió genètica.²⁴

Tegmark troba lògica la conclusió que si arriba a existir una SI ens dominarà com nosaltres hem dominat els animals menys intel·ligents que nosaltres.²⁵ Bostrom ens posa en guàrdia davant d'una futura SI: «Una SI artificial podria ser en si mateixa un agent molt potent, que podria tenir èxit en afirmar-se davant del projecte que li va donar l'existència, i contra la resta del món».²⁶

En aquest punt, estem obligats a distingir dos escenaris. El primer seria la creació d'una SI a partir de la multiplicació bestial de les capacitats bàsiques de la IA actual: accés a dades i càlcul. La segona possibilitat és que algun dia es creï a partir de noves tecnologies una SI general i forta que repliqui perfectament i superi descomunament les capacitats humanes, que no es poden reduir a dades i càlcul.

En aquest article només podem tenir en compte la primera possibilitat, que es plantegen autors com Tegmark o Bostrom. La segona no es refereix a un impossible meta-

²¹ N. BOSTROM, *op. cit.*, pàg. 22, 93.

²² L. ALEXANDRE i J. M. BESNIER, *Les robots font-ils l'amour?* Malakoff: Dunod, 2016, posició a Kindle: 963.

²³ L'anomenada «llei de Moore» estableix que cada 18 mesos es duplica la capacitat de les unitats de memòria i la velocitat dels processadors.

²⁴ M. TEGMARK, *op. cit.*, pàg. 43-45.

²⁵ *Ibid.*, pàg. 172.

²⁶ N. BOSTROM, *op. cit.*, pàg. 95.

físic però ens situa en un terreny on qualsevol cosa es pot dir, com podem fantasiejar sobre *cyborgs* o intel·ligències extraterrestres. Les elucubracions sobre la SI són altament especulatives: Kurzweil, per exemple, imagina la SI del futur com una realitat omnipotent i incompreensible que podria créixer indefinidament fins a sortir de la Terra, ocupar la galàxia, «impregnar l'univers i decidir el seu destí».²⁷ Déu n'hi do...

Bostrom o Tegmark, més moderats que Kurzweil, anticipen els reptes i les amenaces que una SI representaria per a la humanitat i n'esbossen diverses hipòtesis: una SI sota control humà amb finalitats benèfiques o operada per estats o grups de poder amb finalitats malèfiques; o una IA autonomitzada del control humà, amb finalitats pròpies no per força coincidents amb els interessos humans i que, o bé podria tenir cura de nosaltres, o bé bandejar-nos com un destorb.²⁸

Altres especialistes rebutgen com a ciència-ficció la idea d'una IA general i forta que resulti exclusivament de l'ampliació del poder de càlcul, sigui com sigui que calculi. M. Boden pensa que «els escèptics de la singularitat tenen més arguments al seu favor» i afegeix que, amb les seves «boges especulacions», els profetes de la SI ignoren la diferència essencial d'IA i IH i especulen alegrement sobre els efectes de la llei de Moore.²⁹ J. G. Ganascia critica amb contundència el «mite de la singularitat», interpreta la llei de Moore com una simple projecció de certes regularitats fàctiques que no imposa cap necessitat, sosté que «la intel·ligència no equival ni a la freqüència d'execució d'operacions elementals, ni al nombre d'informacions acumulades en una memòria», i conclou que «ni el creixement de la potència de càlcul, ni de la capacitat d'emmagatzemar, no produeixen automàticament la intel·ligència».³⁰ Podem augmentar tant com vulguem les funcions elementals (càlcul, memòria, etc.) de la IA, podem multiplicar-ne la capacitat d'establir i aplicar correlacions: el que s'esdevindrà, creu Ganascia, no serà cap intel·ligència equiparable a la nostra i tampoc una SI que prengui el control de la humanitat.

El debat sobre la SI exposa en termes exagerats el que es pot dir en termes més limitats de tota IA: que emergeix com un nou poder davant l'ésser humà. No cal esperar que es realitzin les profecies de Kurzweil o les pors de Bostrom o Tegmark per adonar-nos que, si qualsevol tecnologia s'apodera poc o molt de les nostres capacitats, la IA ho farà encara més, com un contrapoder amb capacitat de suplantar-nos en molts terrenys.³¹

²⁷ R. KURZWEIL, pàg. 392-421.

²⁸ M. TEGMARK, *op. cit.*, pàg. 204 ss.; N. BOSTROM, *op. cit.*, pàg. 115 ss.

²⁹ M. BODEN, *op. cit.*, pàg. 150-152.

³⁰ J. G. GANASCIA, *Le mythe de la singularité*. París: Ed. du Seuil, 2017, pàg. 450-454, 558.

³¹ En aquest sentit, centenars de científics i investigadors en IA, entre els quals Gates, Hawking, Musk, Wozniak, Tegmark, Russell, Dennet, etc., van signar el manifest «*Autonomous Weapons: An Open Letter from AI & Robotics researchers*», publicat el 28-7-2015 i que es pot llegir a <https://futureoflife.org>, que adverteix dels perills del sistema d'armament basats en IA que operen al marge del control humà.

Més endavant parlarem de com la IA ja substitueix i controla l'ésser humà en alguns àmbits. Ara seguirem Ganascia quan exigeix que afinem la distinció entre IA i IH. La diferència essencial d'IA i IH és precisament la tercera tesi sobre la IA, que passem a explicar.

3. La IA difereix essencialment de la IH

L'equiparació de la IA i la IH és freqüent entre tecnòlegs, però la tesi que IA i IH són ben diferents domina el terreny filosòfic. N'esbossarem els principals arguments, sense pretensió de desenvolupar-los a fons ni de negar la importància d'altres raons. Ens centrarem en les dimensions evolutiva-cultural, corporal, sentent-intel·lectual i personal-pràctica de la intel·ligència humana. Volem posar en relleu que entre IA i IH no existeix avui només una diferència de grau, sinó que són realitats essencialment diverses.

a) Dimensió evolutiva-cultural

D'acord amb la metàfora computacional, la cultura d'un ésser humà equivaldria al programari i a les dades amb què treballa un ordinador. Tanmateix, la IH pertany a l'animal humà: està associada en els seus orígens a la feblesa instintiva dels nostres ancestres, a l'activitat caçadora-recol·lectora de grups socials cada cop més grans que creaven i es llegaven un bagatge cultural. La cultura és alhora el producte i el medi de l'activitat intel·lectual i el principal recurs evolutiu de l'espècie. Per això deia A. Gehlen: «Si la cultura és natural per a l'home, mai no copsarem la seva naturalesa tal i com és, sinó impregnada de compostos culturals molt precisos».³²

Una IA no és un producte evolutiu i no podem dir que li donem una cultura subministrant-li dades o connectant-la a internet. La cultura no és simple informació, sinó un complex bagatge de formes de vida, representacions, preferències, projectes, valors, hàbituds, maneres de sentir, d'expressar i d'interpretar el món, de les quals ens fem càrrec més o menys reflexivament. Parlem d'un complex de possibilitats i capacitats rebudes de generacions precedents, apropiades o rebutjades, transformades o bescanviades per unes altres de noves, i lliurades a la generació següent (tradicció). La transmissió cultural no és una transmissió d'informació amb els seus biaixos, sinó «el lliurament de maneres d'estar en la realitat» que ens ubiquen en un context social i en un temps històric determinats i que ens projecten vers el futur.³³

³² A. GEHLEN, *Antropología filosófica*. Barcelona: Paidós, 1993, «La naturalidad de lo cultural», pàg. 97 ss.

³³ X. ZUBIRI, *Tres dimensiones del ser humano*. Madrid: Fundación X. Zubiri-Alianza Ed., 2006, pàg. 138 i 155.

Amb uns quants clics puc descarregar, gravar o esborrar dades i programes dins d'un computador; la cultura, en canvi, ens va conformant mitjançant un lent i complex procés d'aprenentatge i socialització que dura tota la vida i que va deixant una empremta estrictament individual en cadascú, fins i tot física i corporal, i que no podem esborrar com ho fem amb les dades d'un disc dur. Puc clonar dos ordinadors, però no hi ha clònics culturals. L'home no té cultura com una IA disposa d'accés a internet o al *big data* perquè és en si mateix cultural i la seva intel·ligència opera sobre la base d'una herència biològica i cultural, social i històrica: rep, rebutja, adopta, transforma i transmet possibilitats i capacitats. La manca de societat, de cultura o d'història de la IA és allò que la fa replicable i aliena a qualsevol lloc, acrítica davant la informació esbiaixada de què s'alimenta, tan dèbil en contextos ambigus on els humans ens movem com el peix a l'aigua. La IA és maldestra en situacions connotades culturalment i socialment, marcades per pressupòsits i prejudicis, determinades per allò que Berger i Luckmann anomenaven «el món donat per descomptat», la vida interpretada, la realitat travessada per la cultura.³⁴

b) Dimensió corporal

L'home «no és una unitat additiva de dues substàncies [el cos i la ment], sinó una unitat sistemàtica de notes», és a dir, «una unitat psicoorgànica».³⁵ La intel·ligència forma part del *sub*-sistema psíquic inserit en un *únic* sistema psicocorporal. Intel·ligir és una activitat complexa que forma sistema amb activitats com respirar, digerir, dormir, veure, escoltar, tocar, agafar una cosa, girar el cap, etc. El psíquic i l'orgànic són dues cares de la mateixa moneda: «No es pot parlar de psique sense organisme —escriu Xavier Zubiri—. Per la mateixa raó, no es pot parlar d'organisme humà sense psique».³⁶ Si moc un braç per agafar una copa de vi, poso en marxa el cervell i el sistema nerviós, el sistema circulatori, l'esquelètic i muscular, la respiració, el metabolisme cel·lular..., i tot això es produeix *en funció d'intencions*, representacions, significats, valors, finalitats, esquemes habituals de conducta i segurament d'una decisió conscient. No és la intel·ligència qui pensa, sinó l'ésser humà complet, amb les capacitats i condicions individuals de cadascú. La intel·ligència humana no és independent del seu suport, com diuen molts informàtics que ho és la IA en relació amb el seu maquinari. Podem traslladar la informació i els programes d'un ordinador a un altre —per a la IA, el substracte particular és indiferent—, però no puc replicar en una màquina la ment d'una persona perquè aquesta forma un únic sistema real amb el seu cos particular, ni garantir,

³⁴ P.L. BERGER i T. LUCKMANN, *La construcció social de la realitat*. Barcelona: Herder, 1996.

³⁵ X. ZUBIRI, *Sobre el home*. Madrid: Fundación X. Zubiri-Alianza Ed., 1986, pàg. 47 ss.

³⁶ *Ibid.*, pàg. 59.

com pretén Kurzweil, la pervivència d'una persona (un sistema psicoorgànic i no pas quelcom que «resideix» en un cos) descarregant tota la informació del seu cervell (*backup*) i introduint-la en un robot.³⁷

L'home no resideix en un cos: és el seu cos; només amb ell «és aquí» ara mateix i no en un altre lloc; amb el seu cos constitueix el «lloc geomètric de la realitat».³⁸ La IH és una funció de l'home «de carn i ossos», i no pas la més rellevant per a la nostra felicitat, insistia Unamuno; deia Ortega també que pensar és una funció vital més, al costat d'altres coses que fem per a sobreviure en la nostra circumstància: «el pensament, la intel·ligència, són una de les reaccions a què la vida ens obliga, té les arrels i el sentit en el fet radical, previ i terrible de viure».³⁹ En cap sentit acceptable es pot dir que una IA «visqui». L'avió més intel·ligent vola sense pilot, però no fa nius ni pon ous.

c) Dimensió sentent-intel·lectiva

L'objecte de la intel·ligència no són les «dades» matemàtiques amb què opera la IA, els mapes de 0 i 1 als quals s'acaba reduint la complexitat d'un fenomen prèviament mesurat, quantificat, ni les correlacions o els patrons de dades. Intel·ligir no equival a processar dades mitjançant algorismes. Dir-ho és una ingent simplificació que deixa fora allò específic de la IH.

El famós exemple de «l'habitació xinesa», de John Searle, serveix per explicar que no existeix cap veritable intel·lecció darrere d'un mecanisme de computació: una persona es tanca dins d'una habitació i li fan arribar unes preguntes en xinès, que no entén, però aprèn a respondre-hi correctament aplicant mecànicament regles de correspondència entre paquets de símbols xinesos i anglesos. Això no vol dir que tingui la més remota idea del que contesta; només es limita a computar, a «manipular signes», sense entendre res.

Les dades codificades que rep un computador tampoc no tenen per a ell cap sentit, no signifiquen res del món en què vivim; a diferència del llenguatge humà, els símbols que processa un ordinador estan mancats d'intencionalitat. «No hi ha cap raó per suposar –conclou Searle– que la meua comprensió tingui res a veure amb programes de computadora, és a dir, amb operacions computacionals d'elements especificats formalment».⁴⁰ Només podem dir metafòricament que l'ordinador *entén* com una porta automàtica també *entén* les instruccions d'una cèl·lula fotoelèctrica, però, més enllà d'això, no hi trobarem cap mena de comprensió del que fa. Cathy O'Neil diu que «als

³⁷ R. KURWEIL, pàg. 222 ss.

³⁸ X. ZUBIRI, *Sobre el hombre*, op. cit., pàg. 79.

³⁹ J. ORTEGA Y GASSET, OC, V, 743. Madrid: Taurus, 2006.

⁴⁰ J. R. SEARLE, «Minds, brains, and programs», a: *Behavioral and Brain Sciences*, 3 (3), 1980, pàg. 417-457.

programes de programari els costa distingir una idea d'una simple seqüència de paraules».⁴¹

Però intel·ligir no és primàriament comprendre significats, com diu Searle, sinó, com mostra X. Zubiri, «aprehensió de realitat».⁴² D'aquesta tesi zubiriana només en destaquem tres aspectes.

Primerament, l'objecte formal de la IH és la «realitat»: l'home aprehèn les coses *com a reals* i no com a estímuls estructuralment lligats a determinades respostes (com fan els animals), ni com a dades que desencadenen determinats processos de computació, com fan els algorismes intel·ligents. No anomenem «realitat» a les coses existents més enllà dels nostres actes, sinó a la «formalitat» o la «manera de quedar» les coses en els nostres actes d'intel·lecció: les coses se'ns hi presenten amb una radical alteritat, com a independents dels actes mateixos en què sorgeixen; *sobretot*, remetent a si mateixes, com a essent «de si» («de suyo») o «pròpiament» les seves notes.

Les coses així apreheses desapareixen tots els dinamismes de la intel·lecció. A més de sentir-les de maneres ben diferents, les relacionem amb altres coses i *en parlem* (*logos*), i *ens fan pensar* (raó): la seva realitat ens empeny, poderosa, a esbrinar què són realment, per què són com són, fins al punt, sovint, de posar en qüestió tot el que s'hagi dit i cregut sobre elles. Per contra, la IA, fins i tot en la versió més refinada de l'aprenentatge profund, no passa de ser un formidable motor de classificació de dades, de descoberta de patrons que van decantant automàticament uns determinats resultats, però sense que mai es produeixi en la màquina el miracle evolutiu que representa l'aprehensió de realitat. Avui la IA pot ser «lògica»: relaciona dades i n'infereix conclusions, però mai no aprehèn realitats ni arriba a ser «racional»: no pensa, no *inventa* noves possibilitats (com a màxim, troba relacions inadvertides), no qüestiona els seus propis resultats, no indaga el fonament real de les dades que relaciona o de les relacions que descobreix.

En segon lloc, l'actualització d'una cosa com a real en la IH és l'«estar present la cosa des de si mateixa, des de la seva pròpia realitat», que es verifica en el més humil dels nostres actes de sentir. La intel·lecció humana és «sentiente», diu Zubiri. Sentir ja és «aprehendre el real en impressió»: el blau del cel en la visió o el so del violí en l'audició «queden» en les respectives percepcions, cadascun a la seva manera, en alteritat radical. Una IA no té aprehensió de realitat, sinó que processa i calcula símbols; necessita que li converteixin els fenòmens en dades matemàtiques computables digitalment, sense copsar-les mai com a «dades-de-realitat».

En tercer lloc, en qualsevol intel·lecció es co-actualitzen com a reals l'acte d'intel·lecció (visió, audició...) i allò que s'hi actualitza (la cosa vista, oïda...): «quan hi és

⁴¹ C. O'NEIL, *Armas de destrucción matemática*. Madrid: Capitán Swing, 2017.

⁴² Seguim bàsicament l'anàlisi de l'aprehensió de realitat de X. Zubiri en la seva trilogia *Inteligencia sentiente* (1980-1984).

present l'intel·ligit, per exemple, quan hi és present aquesta pedra, no tan sols veig la pedra, sinó que sento que estic veient aquesta pedra»; «així és com estic en mi», apunta Zubiri.⁴³ És la «*conciencia sentiente*», la co-presència en unitat indestruïble de l'acte i la cosa que obre camí a la reflexió o a la introspecció com a formes ulteriors d'«entrar en mi mateix».⁴⁴

Avui no hi ha IA autoconscients, però Tegmark diu que no hi ha cap impediment metafísic per crear formes complexes de processar la informació d'on emergeixi en el futur, com passa amb el nostre cervell, «l'experiència subjectiva» que s'anomena consciència: «Si la consciència és la manera en què la informació se sent quan es processa d'unes determinades maneres, aleshores ha de ser independent del substrat; l'únic que importa és l'estructura del processament de la informació»; la consciència seria el «fenomen emergent de més alt nivell» a què podria donar lloc una molt complexa forma de computació.⁴⁵

La tesi de Tegmark no suggereix un impossible, però resulta especulativa i en tot cas faria referència a una forma de consciència no humana, atès que la consciència «humana és específica del nostre sistema psicoorgànic, corporal, sensible, afectiu, volent... Avui ningú parla de màquines conscients ni es poden equiparar, més enllà d'una lleugera analogia, els processos pels quals la IA aprèn de si mateixa, avalua dades o practica la recursivitat, amb la intimitat, la consciència i la reflexió inherents a la IH.

d) Dimensió personal-pràctica

La IH és pròpia de «persones» humanes. Intel·ligir és una activitat personal, el subjecte humà és un *qui*, una persona, i no un *què*, un computador. Un *qui* que sent les coses com a reals (*de suyo*), i se'n fa càrrec, i que aprehèn la seva pròpia realitat com a *seva* (és la «*suidad*», diu Zubiri), i se'n fa càrrec. No som intel·ligents pel fet de ser persones, sinó que el nostre caràcter personal es fonamenta en la nostra intel·ligència i en la manera com aprehenem la nostra pròpia realitat, cadascú com a «seva», distensa, oberta.

Com que aprehenem coses reals, el nostre quefer sempre resta obert i no maquinament determinat, com passa amb un ordinador. Les coses intel·ligides «indeterminen» les nostres accions; no hi ha estrictes automatismes en la *conducta* humana, com sí que hi són en la computació. La realitat, en canvi, dona molt de si: una pedra a la mà pot ser un material de construcció, un objecte d'interès mineralògic, una arma homicida,

⁴³ X. ZUBIRI, *Inteligencia y realidad*. Madrid: Fundación X. Zubiri-Alianza Ed., 1980, pàg. 157.

⁴⁴ *Ibid.*, pàg. 161.

⁴⁵ M. TEGMARK, *op. cit.*, pàg. 373.

un petjapapers, el record d'un indret estimat, o l'humil pedra dels camins cantada en el poema de León Felipe, «Como tú».

D'altra banda, els humans aprehenem la nostra realitat psicoorgànica com a real, i cadascú l'aprehèn com a «seva»: «soc una realitat que, com a forma de realitat, no solament soc “de si” [“de suyo”] (en això coincideixo amb totes les altres realitats), sinó que a més soc “meu”. Tinc una realitat que és meva, cosa que no passa amb la pedra».⁴⁶ Aquest caràcter «meu» de la meua realitat és la «personèitat»; el fet d'intel·ligir em fa persona, algú que va modelant fins a la mort, a base d'actuar, la seva figura concreta o «personalitat».⁴⁷ Per mor de l'obertura que imposa la intel·ligència, podem parlar en les persones de moralitat, d'autonomia, de voluntat, d'intimitat, de vinculació («relligament») al poder del real, un poder que ens força a actuar i alhora «allibera» les nostres actuacions de qualsevol automatisme implacable.

L'opacitat i la imprevisibilitat dels resultats d'un procés de classificació (recerca de correlacions i de patrons) dut a terme per una IA no guarden cap analogia amb la indeterminació dels nostres actes resultant de l'obertura insondable de la realitat que s'hi actualitza: tan sols deriva de la incapacitat humana de seguir-li el pas a una IA i d'abastar el munt de dades amb què treballa. L'algorisme d'aprenentatge automàtic es limita a assenyalar «lleis que resumeixen observacions», per molt amagades que estiguin en les dades, però no copsa l'alteritat radical d'aquestes dades, no crea categories rupturistes, ni sotmet a crítica el seu propi llenguatge de programació,⁴⁸ com sí que ho pot fer la raó humana amb les seves categories i llenguatges. L'aparent «autonomia» d'una IA no suposa llibertat ni responsabilitat, cap mena d'autonomia –diu Ganascia– entesa com a «capacitat de donar-se a si mateix la pròpia llei», sinó que tan sols significa que el sistema informàtic desenvolupa determinades funcions sense que hi intervingui cap agent humà.⁴⁹ Automatisme no és veritable autonomia, per molt que l'automatisme doni lloc sovint a un procés opac i imprevisible de computació de dades. Equiparar màquina i home condueix abans a mecanitzar l'home que a legitimar la humanització de la màquina. Com diu Jordi Corominas: «els enginyers actuals diuen molt més sobre la seva concepció de l'home que sobre els propis robots, perquè l'home que prenen com a model, al qual els robots han de reemplaçar, és un home ja robotitzat».⁵⁰

Les raons exposades evidencien la distància ontològica avui existent entre la IA i la IH. No hi ha entre elles una simple diferència de grau ni tan sols de complexitat: són

⁴⁶ X. ZUBIRI, *El hombre y Dios*. Madrid: Fundación X. Zubiri-Alianza Ed., 2012, pàg. 48-49.

⁴⁷ Per això, diu Zubiri, «el hombre es siempre *lo mismo* pero nunca es *el mismo*». *Ibid.*, pàg. 51.

⁴⁸ J. G. GANASCIA, *op. cit.*, pàg. 783.

⁴⁹ *Ibid.*, pàg. 740.

⁵⁰ J. COROMINAS, «Crítica a la religión tecnocientífica». *Revista Perifèria CPG*, 4 (2019), pàg. 147.

realitats diferents, amb notes constitucionals (pròpies del sistema i no de les seves parts) i constitutives (essencials...) ben diverses.

Arribats aquí, però, encara hi hauria una manera d'esquivar parcialment aquesta conclusió i d'establir alguna semblança entre IA i IH. Es pot renunciar a l'equiparació entre la IA i la IH pel que fa a la seva naturalesa, però mantenir l'analogia entre elles pel que fa a les finalitats, els resultats i alguns procediments. És la quarta i última tesi sobre la naturalesa de la IA que volem examinar: la IA, almenys, imita la IH.

4. La IA imita la IH

Com que A. Turing no veia possible respondre *in recto* a la qüestió *Poden pensar les màquines?* a partir de definicions unànimes de «pensar» i de «màquina», va decidir replantejar-la proposant un experiment que permetria saber almenys si una màquina pot imitar una persona: «*The Imitation Game*».

L'experiment es faria així: situem una persona, l'interrogador, en una habitació parlant mitjançant un teclat amb un home i una dona ubicats en un altra habitació. Els pot preguntar el que vulgui per esbrinar qui és l'home i qui la dona. Ara imaginem que substituïm l'home per un computador que respon les preguntes de l'interrogador sense que aquest ho sàpiga. Si el joc continua sense que sospiti res, aleshores es pot dir que el computador ha sabut «imitar» prou bé el pensament humà, sense que sigui rellevant dir si efectivament pensa o no: «No poden realitzar les màquines alguna cosa que hauria de qualificar-se com a pensament, però que és molt diferent del que fa l'home?» –es pregunta Turing.⁵¹ Obviem si un computador pensa o no com un humà, el cas és que calcula, parla, respon, simula sentiments o fa música, que són coses que també fan les persones: obté els mateixos resultats que nosaltres però d'una altra manera. Com a mínim, doncs, és cert que la IA *imita* la IH.⁵²

Però caldrà detallar en què consisteix la imitació. La IA fa coses que també fa la IH *però les fa a la seva manera*, mitjançant llenguatges de programació, algorismes (programari), dades i unes bases materials (maquinari) que tenen poc a veure amb el nostre sistema psicoorgànic. Bostrom diu que «una IA no ha d'assemblar-se necessàriament a una ment humana», però no para de referir-se a les «IA de nivell humà», capaces de fer «igual o millor» les tasques que fem les persones.⁵³ N. Carr afirma que «els ordinadors són capaços de replicar les nostres finalitats sense replicar els nostres mitjans».⁵⁴ Una

⁵¹ A. M. TURING, *op. cit.*, pàg. 435.

⁵² Turing va anar molt més enllà del joc de la imitació i apostaria avui per una IA general i forta, capaç de replicar la ment humana i d'aprendre com ho fa ella i de relacionar-se amb les persones mitjançant interfícies humanitzades. R. LÓPEZ DE MANTARAS i P. MESEGUER, *op. cit.*, posició a Kindle: 183.

⁵³ N. BOSTROM, *op. cit.*, pàg. 29.

⁵⁴ N. CARR, *Atrapados. Cómo las máquinas se apoderan de nuestras vidas*. Madrid: Taurus, 2014, posició a Kindle: 186.

calculadora realitza ràpidament i infal·liblement operacions matemàtiques que les persones faríem molt penosament, però no direm que la calculadora *sap* calcular o que *sap que calcula*, i menys encara que ho faci *com* la nostra ment. Hi ha programari de reconeixement d'imatges, però necessita computar-ne desenes de milers per distingir bé una classe d'objectes (les persones, els gats, el foc...) en qualsevol context, mentre que una criatura veient-ne un parell de la mateixa classe ja en té prou. *Google Translator* ens podria oferir en menys d'un segon la versió anglesa d'aquest mateix article a base de fer el balanç de les equivalències més recurrents entre textos anglesos i les seves traduccions catalanes disponibles a internet, però, com Searle ha explicat, és evident que el programa no «parla» cap de les dues llengües.⁵⁵

A més de la coincidència en els resultats, alguns especialistes assenyalen entre IA i IH una semblança estructural, una semblança material i alguna analogia en la manera de procedir, similituds que no arriben a diluir les diferències essencials entre elles.

a) La *semblança estructural* existiria entre el complex estructural que fa possible l'activitat intel·ligent humana i l'estructura de l'anomenada «intel·ligència artificial integral». Aquesta associa capacitats d'afectació o registre (sensors 360°, càmeres, micròfons, lectors...), d'emmagatzemar dades (memòria, servidors...), de computar informació per resoldre problemes (processador, programari...) i aprendre de la pròpia experiència, de l'entorn o d'internet, i funcions efectores d'actuar sobre el medi o de compartir i comunicar resultats (robots, pantalles, màquines annexes...). Aquestes capacitats i funcions es poden atribuir també a molts animals per mor de la seva sensibilitat, però en els humans tenen totes un caràcter intel·lectiu. La «realitat» constitueix l'àmbit en què es mouen tots els actes humans: l'afectació és una afectació real de l'home i no una simple estimulació, el registre és aprehensió de la cosa en la seva «radical alteritat», la qual modifica la manera en què l'ésser humà se sent realment afectat (sentiment) i dispara una multitud de processos lògics i racionals per resoldre els nostres interrogants i ajustar, justificar o projectar les nostres conductes. L'activitat efectiva inclou en nosaltres la volició i no una mera resposta automàtica o una acció sobre el medi.⁵⁶ La semblança estructural entre IA i IH és una analogia que no suposa una correspondència real entre totes dues.

b) La *semblança material* entre IA i IH és un objectiu que es vol assolir mitjançant l'antropomorfització de la IA per convertir-la en una tecnologia «amable». La indústria s'esforça perquè la IA imiti l'ésser humà i estableixi amb ell una relació de «tu a tu». Cada dia podem parlar millor amb sistemes automàtics: «Arribem a una època on ens

⁵⁵ Luc Julià, un dels creadors de *Siri* (Apple) posa l'èmfasi en els «mètodes diferents» que fan servir persones i sistemes intel·ligents per parlar, reconèixer imatges, etc. La diferència és tan gran, diu ell, que caldria concloure que la tecnologia anomenada IA «no té res a veure amb la intel·ligència». Vg. L. JULIA, *L'intelligence artificiel n'existe pas*. París: Ed. First, 2019, pàg. 123-126.

⁵⁶ X. ZUBIRI, *El hombre y Dios*, op. cit., pàg. 44-46.

envoltaran paraules emeses per màquines», escriu Sadin.⁵⁷ Siri (Apple), Alexa (Amazon) o OK Google són exemples que la majoria hem experimentat. El seu èxit, explica Luc Julia, un dels creadors de Siri, no radica només en el fet que ens «entenem» i ens contesten, sinó que simulen aspectes tan humans com el sentit de l'humor, la picardia, els dubtes o el canviar de tema quan no saben què respondre. «En realitat, l'èxit de Siri –diu Julia– es deu a la introducció de l'estupidesa artificial!».⁵⁸ L'antropomorfització de la IA culmina amb la construcció de robots hominoides, com els famosos robots japonesos per a la cura personal o altres robots amb aspecte clarament humà, com la famosa Sophie o els robots de serveis sexuals.

La imitació de la IH per part d'alguns sistemes d'IA referma la sensació que som davant d'éssers «personals» amb qui podem mantenir una relació «humana» i a qui aviat haurem d'imposar deures o concedir drets. Tanmateix, la semblança material torna a ocultar la veritable naturalesa de la IA, que Julia emfatitza: tots aquests artefactes humanitzats «executen les tasques per a les quals han estat fets. No inventen res, només segueixen regles, exemples i codis utilitzant les dades que escollim per a aquests sistemes [...]. No són res més que el que nosaltres, humans, hem decidit que siguin».⁵⁹

c) L'*analogia procedimental* entre IA i IH es concreta en la capacitat comuna de relacionar dades, descobrir-hi correlacions i patrons, i d'actuar posteriorment sobre la base d'aquesta descoberta. No és estrany que Kurzweil ens recordi que «la capacitat de reconèixer patrons és un dels pilars de la nostra espècie».⁶⁰

David Hume va reduir els dinamismes mentals al «principi de l'associació d'idees» segons la semblança, la contigüïtat i l'associació constant: «D'aquí ve –deia– això que anomenem l'à *propos* del discurs: d'aquí la coherència dels escrits; i d'aquí ve també aquell fil o seqüència de pensaments que la persona sosté naturalment, fins i tot en la fantasia més desbocada».⁶¹ Per al filòsof escocès, les regularitats detectades en el passat són la base de les nostres creences sobre el futur: «Estem determinats a suposar que el futur és conforme al passat únicament pel costum».⁶² També «els algorismes de la IA –escriu D. Cardon– confien en la regularitat de les estructures de gust i interès del usuari» com a fonament de les seves prediccions, «parteixen de la suposició que existeix un caràcter regular i previsible de les pràctiques humanes», suposen que fins i tot

⁵⁷ E. SADIN, *Intelligence artificielle, op. cit.*, pàg. 74.

⁵⁸ L. JULIA, *op. cit.*, pàg. 63.

⁵⁹ *Ibid.*, pàg. 150.

⁶⁰ R. KURZWEIL, *op. cit.*, pàg. 165.

⁶¹ D. HUME, *Resum del Tractat de la naturalesa humana (Abstract)*. Trad. de Miquel Costa. Barcelona: La Busca Ed., pàg. 31.

⁶² *Ibid.*, pàg. 19.

els nostres comportaments més personals obeeixen a pautes i rutines.⁶³ Com deia Hume que fa la nostra ment, les IA assenyalen correlacions sense trobar-ne les causes. Més concretament, estableixen correlacions entre valors substitutius (viure en un barri determinat o els *like* a Facebook) de fenòmens complexos difícils d'estudiar (la conducta delictiva o l'amistat) i en dedueixen conclusions que aparenten ser ciència i veritable coneixement, quan l'elecció d'aquells valors substitutius respon sovint a biaixos ideològics gens innocus. Les prediccions dels algorismes d'IA pressuposen el manteniment de les pautes socials, però adverteix Carr que «en les ocasions peculiars on les condicions es desvien dels patrons establerts, els algorismes poden fer prediccions brutalment errònies».⁶⁴

Si l'activitat de la ment es reduís a l'associació d'idees, aleshores tindria sentit equiparar-la a una IA dedicada fonamentalment a la descoberta de patrons en muntanyes de dades. És cert que els humans també trobem patrons i regularitats en la nostra experiència, però la nostra ment no es limita a fer això, sinó que és una complexa *intelligència sentent* que es mou en la formalitat de realitat i apunta a les raons, les causes i els fonaments de les correlacions percebudes, es distancia del món conegut i imagina noves possibilitats, inventa projectes, crea llenguatges i els posa al servei de la comprensió del real, etc.

Daniel Innerarity sosté que la fortalesa dels algorismes d'IA «consisteix en el fet que extreuen les característiques que es repeteixen deixant al marge les propietats rares, variables i contingents. No tan sols és que es recolzin en estereotips sinó que estan calculats per maximitzar la conformitat als estereotips».⁶⁵ Per això la IA tendeix a justificar el sistema vigent més que no pas a millorar-lo, és presonera del passat, de les dades registrades i de les repeticions. La IH, en canvi, està sempre oberta a la novetat i al futur; només així ha pogut ser evolutivament útil a l'animal humà. La IH pot ser racional, la IA no.

Tampoc la nostra conducta racional encaixa en l'esquema conductista de la repetició dels comportaments reeixits, perquè, per molts condicionaments que pesin sobre nosaltres, la realitat «allibera» o indetermina sempre poc o molt la nostra praxi, i perquè la IH, en la seva dimensió racional, sotmet a crítica la nostra praxi, fins i tot la més exitosa, albira noves possibilitats i inventa nous projectes. Cardon assenyala que «com més espai es deixa per a una reflexió distant de l'usuari, els càlculs predictius [de la IA, sobre els nostres comportaments] se'n ressenten».⁶⁶

⁶³ D. CARDON, *Con qué sueñan los algoritmos*. Madrid: Dado Ed., pàg. 84-85.

⁶⁴ N. CARR, *Atrapados*, *op. cit.*, posició a Kindle: 2045.

⁶⁵ D. INNERARITY, «La inteligencia de la inteligencia artificial». *La Vanguardia*, 16-12-2019.

⁶⁶ D. CARDON, *op. cit.*, pàg. 92.

Ara bé, més semblant o més diferent, el cas és que la IA pot influir sobre la vida humana de manera decisiva. Deixant de banda el debat sobre la naturalesa de la IA, es pot dir que és una tecnologia de la substitució de l'home amb la finalitat de «facilitar-li» la vida però també d'administrar-lo i d'explotar comercialment fins al darrer detall de la seva vida. Ho expliquem en l'últim apartat d'aquest article.

5. L'impacte de la IA: la substitució de l'ésser humà

El debat sobre el parentiu entre IA i IH no posa sobre la taula la qüestió més humanament transcendent: el significat real de la IA i el seu impacte profund en la nostra vida personal i social. Mentre estem en aquella discussió i ens enorgullim de la nostra especificitat inimitable, resulta que la IA, a poc a poc, ens guanya la partida del control de la nostra vida mental, personal i social, i ens va substituïnt en tots els terrenys: ens dispensa bons serveis a base de deixar-nos de banda. No estem fent un judici de valor sinó enunciant un fet.

La tècnica no és simplement un «saber fer» (Aristòtil), ni «ciència aplicada» (Descartes), ni un mitjà de transformació del món per mor del nostre «benestar» (Ortega). Semblen ingènues les tesis de la «neutralitat» o la «complementarietat» de la tècnica, que posen l'accent en l'ús bo o dolent que l'home en pot fer. Tota tècnica és una «realitat» i se'n planta al davant amb radical alteritat (Zubiri), sigui un martell o un *smartphone*. «Les coses artificials –diu Antonio González–, com les coses naturals, sorgeixen davant nostre i s'imposen sobre nosaltres».⁶⁷ Això fa de qualsevol tecnologia quelcom no del tot disponible ni controlable: mai no acabem de saber què donarà de si, quins efectes tindrà sobre nosaltres, la societat o l'entorn natural. Cap tecnologia es deixa dirigir del tot; com a màxim es pot «co-regir»: «L'ésser humà, com a aprenent de buixot –conclou González–, queda dominat i abraçat per la tècnica que ell mateix ha creat».⁶⁸

La tècnica no és, per tant, un recurs que canvia el món sense alterar-nos a nosaltres mateixos, sinó que s'insereix en la nostra vida i la modifica, afecta les nostres relacions i ho fa en la mesura en què resulta útil i eficaç. Per dir-ho una mica bruscament, la tècnica té un poder «infecció»: penetra la nostra vida i la canvia profundament. Més que un recurs per viure, constitueix el «medi» en què vivim: «Esdevinguda un Univers de mitjans, la Tècnica és de fet el *medi* de l'home» –diu Jacques Ellul–, és «un conjunt coherent que ens encotilla a tot arreu, s'introdueix en nosaltres i no ens en podem

⁶⁷ A. GONZÁLEZ, «Corregir la tècnica». *Cuadernos de Filosofía Latinoamericana*, Vol. 34, Núm. 109, 2013, Bogotá, pàg. 96.

⁶⁸ *Ibid.*, pàg. 98.

desfer».⁶⁹ La tècnica és una nota constitutiva o un «factor determinant» de l'actual sistema social (econòmic, polític, cultural) tecnoliberal i constitueix per si mateixa el «sistema tècnic», un complex de recursos i processos que s'impliquen recíprocament, s'alimenten els uns als altres i responen a inèrcies i dinamismes específics: sembla quelcom dotat de vida pròpia.

Ignasi Boada assenyala com a trets específics del sistema tècnic la simplificació matemàtica del real amb vistes a la seva dominació, l'esperit d'eficàcia, l'autonomització en relació amb el control humà, la permanent inestabilitat que introdueix en les nostres vides, sempre devaluades sota l'impuls de la «innovació» i el «progrés», l'acceleració que imposa a tot allò que fem, o el seu caràcter totalitari, perquè no deixa res fora del seu embat.⁷⁰ Doncs bé, si la IA és el vaixell insígnia de l'actual sistema tecnoliberal i la major promesa d'eficàcia i benestar que mai ha rebut la humanitat, és lògic que hi refluïxin de manera eminent els altres trets del sistema tècnic i que ocasioni un trasbals insòlit en la nostra existència. La IA és una força dominant dins la societat global, informacional i capitalista «tecnoliberal» (Sadin) del segle XXI. Aquest poder desbordant de la IA apunta cap a una creixent substitució de l'ésser humà amb vistes, molt sovint, al seu control i a la seva explotació comercial. La IA és una tecnologia mai del tot disponible per part de l'home, i li va prenent el lloc. «Hem dissenyat un sistema que ens descarta», sentència Nicholas Carr.⁷¹

La substitució de què parlem és produeix almenys de tres maneres:

a) La IA substitueix el treball humà

Les previsions varien molt segons qui parli: pronostiquen en algun cas que podrien desaparèixer fins al 85% dels llocs de treball actuals per l'aplicació d'IA en tots els sectors productius.⁷² Els primers a ser substituïts seran els obrers que fan feines molt rutinàries, però aviat també els qui fan tasques de formació, disseny o direcció. Com és lògic, molts treballs que requereixen intel·ligència seran realitzats per la IA. Moltes persones esdevindran auxiliars de màquines intel·ligents, els seus entrenadors i supervisors,⁷³ o es limitaran a sancionar les decisions que ha pres un algorisme. Com més sofisticades són les màquines, menys feina queda per a nosaltres.

⁶⁹ J. ELLUL, *Le Systeme techicien*. París: Le cherche midi, pàg. 49 i 54.

⁷⁰ I. BOADA, *La tecnologia com a forma de vida*. Vic: ISCRV, 2001.

⁷¹ N. CARR, *Atrapados*, op. cit., posició a Kindle: 3315.

⁷² DIV. AUTORS, *Intelligence artificielle, 4. Les algorithmes vont-ils tuer l'emploi?* París: Flammarion, pàg. 65 ss.

⁷³ Fins i tot els qui parlen de col·laboració home-màquina imaginem un futur on els homes són simples facilitadors dels treballs de les màquines. Cf. H. DAUGHERTY i H. JAMES WILSON, *Human+Machine: Reimagining Work in de Age of AI*. Harward Busines Rewiew Press, 2018.

Una IA no dubta, no es cansa, és «objectiva», no cobra un sou i no reivindica millores laborals. En canvi, els humans som fal·libles, poc fiables, «subjectius», reivindicatius, cal que ens paguin, ens posem malalts: «Sigui quina sigui la forma d'organització –diu Sadin–, hi ha un factor que, a més de generar costos, oposa una força d'inèrcia, comet errors, contesta les decisions: l'humà».⁷⁴ El 2017 un executiu de Goldman Sachs justificava l'acomiadament de 600 *traders* financers per deixar-ne només dos: «Ja saben vostès, 600 persones ocupen molt espai». La compravenda als mercats financers ja depèn en gran mesura de sistemes intel·ligents automàtics. No hi ha res més productiu per a moltes empreses que desfer-se de persones.

«La IA ens farà bascular vers una altra civilització on el treball i els diners podrien desaparèixer».⁷⁵ Hom parla obertament de la «societat post-treball» on se suposa que la gent viurà ociosa i es dedicarà a activitats culturals o solidàries. S'aventura que tot això es podrà finançar amb els rèdits generats per un sistema econòmic farcit d'IA. No ens ha d'estranyar que molts grans empresaris de la IA demanin una renda bàsica universal per subvencionar la vida de molts extreballadors que seran «consumidors» en estat d'insatisfacció perpètua, més que no pas artistes, filantrops o contemplatius. A mitjans del segle XX, encara sense IA, Hanna Arendt valorava la possibilitat d'una societat on els sistemes automàtics haurien eliminat el treball humà: «Es tracta d'una societat de treballadors que està a punt de ser alliberada de les traves de la feina, i aquesta societat desconeix aquestes altres activitats més elevades i significatives per a les quals mereixeria guanyar-se aquesta llibertat. [...] Ens enfrontem amb la perspectiva d'una societat de treballadors sense feina, és a dir, sense l'única activitat que els queda. Està clar que res podria ser pitjor».⁷⁶

Darrere del mite de l'alliberament del treball gràcies a la robotització, s'hi amaga el prejudici que el treball és *sempre* una desgràcia que cal evitar. Però també pot existir el treball com a ofici o professió realitzat en condicions dignes, salvat de l'explotació capitalista i de la mecanització del treballador. Si respon a les aptituds i l'interès de les persones, el treball «representa una part de la dignitat humana»,⁷⁷ estructura la vida, l'orienta, li dona un sentit, una funció social. El treball esdevé aleshores una forma d'estar en la realitat, una forma de coneixement i d'autoconeixement. Pot ser l'autoexpressió de l'ésser humà: treballant podem manifestar el que som, posar-nos a prova, assumir reptes, donar de si, col·laborar amb els altres, responsabilitzar-nos-en. «Massa sovint –lamenta Carr– l'automatització ens allibera d'allò que ens fa sentir-nos lliures».⁷⁸

⁷⁴ E. SADIN, *op. cit.* (IA), pàg. 147.

⁷⁵ L. ALEXANDRE i J.M. BESNIER, *op. cit.*, posició a Kindle: 856.

⁷⁶ H. ARENDT, *La condició humana*. Barcelona: Paidós, 2009, pàg. 15.

⁷⁷ E. SADIN, *op. cit.* (IA), pàg. 154.

⁷⁸ M. CARR, *op. cit.* (AT), posició a Kindle: 203.

b) La IA substitueix les capacitats humanes

Si la IA elimina feines i descarta treballadors, és perquè substitueix les nostres capacitats en tots els àmbits de la producció i l'administració. I ho fa també en la vida quotidiana, en les relacions personals, l'oci, la llar, la creació artística, la cura dels altres. De quines capacitats estem parlant? No pas de la nostra força física o de la capacitat de càlcul, com feien altres tecnologies: una excavadora o una calculadora, per exemple. La IA obté informació, la selecciona, la classifica i l'analitza, ens fa propostes, ens suggereix decisions o les pren directament en lloc nostre, executa tasques, es comunica amb nosaltres i ens suplanta en la comunicació amb els altres. Deixem d'usar moltes capacitats mentals perquè les exerceixi un sistema intel·ligent que iguala o millora els nostres resultats. A quin preu? L'apunta l'historiador de la tecnologia George Dyson: «I si el cost de tenir màquines que pensen és tenir gent que no pensa?».⁷⁹

Si la màquina fa les coses millor que l'home, aleshores és millor reduir al mínim la intervenció humana. La IA, per exemple, fa el diagnòstic i el metge el signa i el comunica al pacient. Haurà d'anar amb compte: si accepta el diagnòstic de la màquina i aquesta no l'encerta, ningú li retraurà res; com que la màquina encerta gairebé sempre, si el metge contradiu la màquina, se la juga. Eric Sadin parla de l'emergència d'un «nou poder *alétheic*», una espècie de «demiürg» que «personifica una forma de superjò dotat en tot moment de la intuïció de la veritat i que orienta el curs de les nostres accions individuals cap al millor dels mons possibles».⁸⁰ Un poder quasi omniscient i opac: en molts casos, veiem les conclusions a què arriba l'algorisme però no en podem conèixer les raons, amagades en el garbuix d'una computació fulgurant on entren en joc centenars de milions de dades i de possibilitats. No és estrany que, cada cop més, es reclami «transparència» a la IA.

c) La IA desposseeix els humans de les seves capacitats.

A més de substituir-les, la IA ens pren facultats i ens imposa les seves. El lema empresarial d'IBM, «Per un planeta més intel·ligent», ve a dir que si l'intel·ligent és el planeta, potser ja no cal que ho siguem cadascú de nosaltres. Hom parla d'*smartphones*, de ciutats intel·ligents, de cotxes intel·ligents, d'*Internet of things* o de la *www* (World Wide Web) com a *intel·ligència col·lectiva*.

La desposseïció de les nostres facultats es produeix de dues maneres. Primerament, perquè ens acostumem a no fer-les servir i les delegem en sistemes artificials: el GPS ens orienta en ruta, Tinder ens busca parella, Google ens facilita qualsevol cosa que cerquem, el pilot automàtic porta l'avió, el sistema EMR guia el tractament mèdic d'un

⁷⁹ Citat a N. CARR, *op. cit.* (AT), posició a Kindle: 1876.

⁸⁰ E. SADIN, *La siliconización del mundo*. Buenos Aires: Caja Negra, pàg. 37.

pacient. Lliurar-se a la multitud de serveis de la IA es paga en forma de desatenció, complaença acrítica, disminució de la memòria, desactivació del raonament profund, pèrdua dels hàbits que ens empoderen i dificultat per crear-ne de nous.

A més de perdre capacitats, en segon lloc, l'home adquireix paulatinament la *forma mentis* més adequada als sistemes d'IA. Carr ho resumeix dient que «programem els nostres ordinadors i, posteriorment, ells ens programen a nosaltres».⁸¹

La *virtualització* tradueix el real al format en què és computable, comunicat sense límits de lloc i de temps i exhibit en una pantalla. Quedem sotmesos, diu Sadin, a «una capa xifrada digital mitjancera en la nostra relació amb els fets i les coses [...]. El pas de la petjada analògica a la manipulació informatitzada indueix la pèrdua de visibilitat i de fricció sensible amb les coses, per introduir jocs d'interferència *imperceptibles* i *automatitzats*».⁸² El paisatge queda substituït pel mapa del GPS; l'amic, pel seu compte d'Instagram; el jo personal, pel jo públic a Facebook o Twitter; la conversa personal, per la relació de missatges de *WhatsApp*; el debat polític, per les polèmiques a la xarxa social; la visita real a la Sagrada Família, per una visita virtual. La realitat sentida és bescanviada per un mar de símbols; el món és l'artifici filtrat i personalitzat que la pantalla ens brinda.

La *dispersió* i la *distracció* constants que ens imposen els algorismes amb la finalitat d'impulsar el consum i facilitar la detecció de patrons en la nostra praxi digital dificulten el pensament lineal, concentrat, circumstanciat, pausat, complex, profund. La multitasca ens distreu, l'estimulació plural no ens deixa reposar enlloc, els *links* dinàmics ens atrauen i mouen sense parar la ment i els sentits perquè nosaltres tampoc no parem de subministrar dades sobre nosaltres mateixos mitjançant clics de ratolí o tocs de pantalla. El resultat és la substitució d'una manera de pensar per una altra: «Calmada, concentrada, sense distraccions, la ment lineal està sent desplaçada per una nova classe de ment que vol i necessita rebre i disseminar informació en esclats curts, descoordinats, sovint solapats, com més ràpidament millor».⁸³

La *recurrència* com a clau d'interpretació de la realitat i el *filtre-bombolla* ens redueixen al que hem sigut fins ara, ens mantenen a prop d'aquells que pensen i actuen com nosaltres, s'anticipen als nostres desitjos suposant que replicarem sempre conductes passades, dificulten que coneguem alternatives i que arribem a pensar diferent. Ens compliquen, en definitiva, anar més enllà de nosaltres mateixos i del cercle dels «nostres» per albirar noves possibilitats: la IA invasiva de la vida personal dificulta, en definitiva, l'exercici de la racionalitat.

⁸¹ N. CARR, *Superficiales. ¿Qué está haciendo internet con nuestras mentes?* Madrid: Taurus, 2011, posició a Kindle: 3384.

⁸² E. SADIN, *La humanidad aumentada*. Buenos Aires: Caja negra Ed., pàg. 41.

⁸³ N. CARR, *Superficiales, op. cit.*, posició a Kindle: 171.

Finalment, l'*ergonomia* o les *interfícies* amables naturalitzen les noves tecnologies, les incorporen al nostre cos, les associen als nostres gestos més senzills, que passen a integrar-se en un procés aliè que no dirigim. Un dels lemes de l'iPad d'Apple és «Aparta la tecnologia del teu camí»: es tracta que no notem la tecnologia perquè aquesta s'ha encarnat en hàbituds que posem en pràctica de manera molt simple i inconscient. El pas següent serà la inserció de dispositius intel·ligents imperceptibles dins del nostre cos per facilitar la monitorització de les nostres dades biològiques i l'orientació permanent d'una existència «saludable», ajustada a uns estàndards que justificaran la autoimposició de pautes de conducta i de consum més sans que algú ens facilitarà a un mòdic preu. Com més suau i amable és la presència de la IA en les nostres vides, més exerceix el seu caràcter totalitari.

d) La IA substitueix l'ésser humà en les seves decisions

Si assignem tasques cada cop més complexes a la IA, és lògic que incloguin resolucions de caràcter típicament moral. A poc a poc, la IA anirà assumint les actuacions humanes que demanen una major inversió de consciència, responsabilitat i llibertat: decretar la condemna d'un acusat o la llibertat condicional d'un presoner, decidir el moviment inversor de milions d'euros a la borsa, la contractació o l'acomiadament d'un treballador, l'ordenació de la feina, l'elecció de parella, l'orientació d'un discurs polític o l'atac amb drons contra un objectiu militar, per citar només alguns exemples. Se suposa també aquí que la màquina no s'altera, no té fòbies ni manies personals o dubtes, i que fins i tot pot ser èticament més ponderada i equitativa que les persones. Es passa per alt que els algorismes incorporen «els prejudicis, les equivocacions i els biaxos humans» dels seus creadors per convertir-se, segons C. O'Neil, en autèntiques «armes de destrucció matemàtica» que ens acaben subjugant.⁸⁴

Sadin diu que «el *superjò computacional* decideix per nosaltres, sota el “postulat tecnològic fonamental” que hi ha una deficiència humana fonamental que serà salvada pels poders de la IA, que augmenten i varien constantment».⁸⁵ S'opera així a poc a poc la dissolució del subjecte modern com a ésser singular i lliure, reflexiu, capaç de prendre decisions, i responsable dels seus actes. Sembla que retrocedim a la «minoria d'edat» preil·lustrada que, segons Kant, hauríem d'haver superat amb l'ús crític de la raó. Silicon Valley «ha destruït –sosté Sadin– el principis fundadors de l'humanisme europeu que afirmen l'autonomia del judici i la lliure elecció i que indueixen un corollari: el principi de responsabilitat i els drets de les societats a decidir el seu destí comú».⁸⁶

⁸⁴ C. O'NEIL, *Armas de destrucción matemática*. Madrid: Capitán Swing, 2017, pàg. 11.

⁸⁵ E. SADIN, *La siliconización del mundo*, pàg. 37.

⁸⁶ *Ibid.*, pàg. 40.

I no tan sols les resolucions més conscients, també les petites decisions més intrascendents se'ns prenen. La IA aplicada a internet té «l'objectiu de crear i perfeccionar una teoria sobre qui ets, el que faràs i el que desitjaràs a continuació».⁸⁷ L'objectiu declarat dels fabricants d'IA és avançar-se a les nostres necessitats, estalviar-nos les preguntes i les demandes: «Google ja considera la cognició humana com quelcom gastat i inexacte, un procés biològic que és millor que faci un ordinador. “Crec que d'aquí a uns anys la majoria de les cerques seran contestades sense ni tan sols haver de preguntar”, afirma Ray Kurzweil».⁸⁸ No caldrà esperar tant: el mòbil ens recorda els aniversaris i ens planifica el dia, el frigorífic que es va buidant encomana per internet els productes que ens falten, el biberó intel·ligent indica als pares la millor inclinació de la mà, la polsera d'activitat ens programa les millors rutines esportives i l'horari d'anar a dormir, la impressora encomana els cartutxos de tòner, els llums de la casa s'encenen i s'apaguen sols quan cal, Google Instant prediu el que vols comprar i et selecciona les millors ofertes, Netflix sap quines sèries vols veure. Es consagra en definitiva un «home no-pràctic», que queda expropiat de la seva praxi completa; aquesta ja no li pertany, l'executa en molts casos però ja no és seva sinó impulsada, orientada, ordenada i quasi decidida per sistemes automàtics intel·ligents. Es despulla progressivament les persones de la seva autonomia. L'ésser humà, a base de fer-se a si mateix cada cop més prescindible, corre el risc de perdre's a si mateix. No cal que temem una invasió de robots maliciosos: ja ens haurem rendit i batut en retirada molt abans davant una IA hegemònica i invasiva.

Conclusió

Hem passat revista a les principals maneres de concebre la IA: l'equiparació, la superació, la distinció o la imitació, i hem reflexionat sobre la seva tendència imparabile a substituir l'home i les seves capacitats més característiques.

Cal reconèixer les aportacions positives que la IA ja està fent en molts terrenys; tots ens n'hem aprofitat personalment i els seus beneficis són evidents en l'economia, les comunicacions, els serveis públics o la salut. No hem de renunciar-hi, per descomptat. Però cal veure quin preu haurem de pagar en termes antropològics si la IA acaba ocupant *tot* el conjunt de la nostra vida personal i social i si es desenvolupa sense un control social democràtic i lúcid. Serà tan important evitar que els algorismes d'IA gestionin la nostra vida personal al servei d'interessos espuris (polítics, comercials, etc.), com demarcar i regular els àmbits de col·laboració entre la IA i les persones d'una manera socialment profitosa. La IA pot ser una «eina» al servei de la societat, però no pot ser una

⁸⁷ E. PARISER, *El filtro burbuja*. Madrid: Taurus, 2017, posició a Kindle: 158.

⁸⁸ N. CARR, *op. cit.* (AT), posició a Kindle: 3036.

«arma» descontrolada al servei dels poders econòmics i polítics. Ha de contribuir a l'eficiència econòmica, però sobretot i molt principalment a la justícia.

Impressiona veure com la matemàtica Cathy O'Neil descriu el desplegament als EUA de la IA associada al Big Data, aplicada a la publicitat individualitzada i depredadora, que exprimeix el consumidor fins al darrer dòlar; a la predicció de delictes, que acaba penalitzant els més pobres i marginats; a la planificació dels horaris dels treballadors, que els priva d'una vida personal i familiar ordenada; a la prevenció de riscos financers, que estigmatitza les classes baixes i justifica que paguin més interessos que ningú o es quedin sense la cobertura de les assegurances; a les campanyes electorals, basades en la microsegmentació de l'electorat per enviar-li missatges i *fake news* personalitzats. Tot plegat en nom de l'eficiència i la racionalitat, quan al darrere hi ha la voracitat capitalista i l'ambició dels poderosos. Els algorismes d'IA esdevenen així «armes de destrucció matemàtica» caracteritzades pel seu abast universal, la seva opacitat i el seu poder de fer mal, aliments de totes les dinàmiques de la desigualtat.⁸⁹

L'*entusiasme* ingenu amb la IA revela una profunda desconfiança en relació amb les persones, el rebuig de la finitud, la corporalitat i la fal·libilitat de l'ésser humà; posa de manifest, en paraules de H. Arendt, «el desig d'escapar de la condició humana», tan característic del nostre temps. Quan l'educació i la política ja no semblen recursos a qui confiar l'orientació i la millora de la nostra existència individual i social, apareix la IA com a pilar d'una utopia tecnològica i salvadora, la sospitosa promesa de la ideologia «tecnoliberal» hegemònica entre les elits polítiques i econòmiques.

Tal i com es va desplegant dins la societat tecnoliberal, la IA promet el guiatge científic i complet de les nostres vides quan el que es persegueix molt sovint és el control polític, la colonització comercial de tots els espais de la nostra existència. Sadin diu que s'aspira a convertir la IA en la «indústria de la vida», la darrera etapa d'un capitalisme salvatge que fa negoci amb les nostres dades personals.

Les sensacions simultànies i falsament compatibles de tenir-ho tot a les mans i d'estar en les millors mans pot anar lligada al major dels esclavatges i a la pèrdua gradual de la nostra humanitat i la nostra llibertat. Per això el debat sobre la IA té una forta dimensió política i caldrà fer-lo en moltes instàncies socials i de manera molt interdisciplinària, lúcida, participativa i democràtica. Per *agafar el timó* de la IA ens haurem d'encomanar novament a la política i l'educació com a recursos per a la preservació de la vida humana en la seva integritat i dignitat. La dignitat humana i la justícia, no pas el negoci, són les finalitats principals del sistema polític i econòmic.

Davant la IA, la filosofia ha d'estar sempre de guàrdia, evitant condemnes taxatives i irracionals, però exercint el seu característic poder d'anàlisi i revisió crítica de les actuacions i dels discursos humans. Ens pot ajudar a entendre en què consisteix el supo-

⁸⁹ C. O'NEIL, *op. cit.*

sat caràcter «intel·ligent» de la IA a mesura que vagi evolucionant i a esbrinar tot allò que es posi en joc quan interfereixi i moduli la nostra praxi. La filosofia pot contribuir al discerniment ètic de tot allò relacionat amb la IA, dels guanys que ens proporciona i els riscos que suposa. És obligatori continuar aquestes reflexions en diàleg amb totes les disciplines vinculades a la recerca i el desenvolupament de la IA. Per descomptat que no es tracta de fer desaparèixer la IA, sinó de «co-regir-la» i de supeditar-la a les més pregoneres necessitats humanes.⁹⁰

Joan Albert VICENS
Facultat de Filosofia (URL)
joanalbert.vicens@gmail.com

Article rebut: 30 de setembre de 2020. Article acceptat: 1 de febrer de 2021

⁹⁰ L'esborrany d'aquest article fou objecte de lectura i debat amb els estudiants de l'assignatura «Veritat i Raó», del Màster de recerca de Filosofia i Humanitats de la Facultat de Filosofia de la URL, el curs 2019-20. Els seus comentaris i objeccions, que vull agrair aquí, em van servir per esmenar-lo i aclarir-lo.

INTELIGENCIA ARTIFICIAL: APROXIMACIÓN DESDE UNA TEOLOGÍA AMIGA DE LA CIENCIA

Juan Ramón LA PARRA

Resumen

El diálogo de la teología con la búsqueda de una inteligencia artificial se centra en la inteligencia artificial «fuerte», que puede englobarse dentro del transhumanismo tecnocientífico. El primer ámbito de interés es el de la teología fundamental, que considera las fuentes para el diálogo, el carácter de fe secular del transhumanismo tecnocientífico y los interrogantes principales ante los que se sitúa el hombre. El segundo ámbito es el de la antropología teológica, con la consideración del hombre creado a imagen de Dios, colaborador de la creación (incluyendo aquí el desarrollo técnico) y herido por el pecado; además, se reconoce la salvación como un don gratuito inseparable de la condición carnal, y se contempla el horizonte escatológico, en el que se sitúan muchas de las profecías de la especulación tecnocientífica.

Palabras clave: teología, inteligencia artificial, fe secular, teología fundamental, antropología teológica.

Artificial Intelligence: approach from a science-friendly theology

Abstract

Dialog between theology and the quest for AI is focused on «strong» AI, which can be included as a part of technoscientific transhumanism. The first field of interest is fundamental theology, which reflects about the sources for dialog, technoscientific transhumanism as a secular faith, and the human main questions. The second field is theological anthropology, considering the human being as created in God's image, collaborator of creation (including technical advances) and wound by sin; moreover, we should acknowledge gratuity and carnality in salvation, and envision an eschatological horizon, that can be related most of prophecies of technoscientific speculation.

Keywords: theology, artificial intelligence, secular faith, fundamental theology, theological anthropology.

1. Introducción¹

1.1. Un discurso teológico en diálogo con la ciencia

En estas páginas se intenta poner en contacto la teología con la búsqueda de la inteligencia artificial, lo cual supone un evidente contraste: buscamos el diálogo entre una disciplina con siglos de historia, heredera de una sabiduría que hunde sus raíces en la misma experiencia original del ser humano, y una investigación tecnocientífica que solo se extiende en el arco de unos decenios, con un desarrollo vertiginoso en los tiempos recientes.

Desde el principio explicitamos que la perspectiva de estas páginas será esencialmente teológica, y específicamente cristiana, pero sin dar nada por supuesto, procurando que nuestro lenguaje sea claro y accesible a aquellos que no se hallen familiarizados con la terminología teológica, buscando por encima de todo el poder comunicar efectivamente. Partimos de la búsqueda de una colaboración humilde, en el camino de la interdisciplinariedad,² y el deseo de integración del conocimiento.³

De este modo, se pretende ayudar a una reflexión que, en las relaciones de la teología respecto de la ciencia y la técnica, deje atrás el conflicto y la indiferencia (o la independencia), para exponerse al diálogo, buscando como meta una auténtica integración.⁴ Por esto planteamos una teología que sea «amiga» de las ciencias, sin renunciar a plantear cuestiones clave, pero evitando la transgresión epistemológica.⁵

En este diálogo que busca la integración, la teología debe situarse con humilde atención ante los avances científicos y técnicos, que presentan continuas novedades, pero no debe precipitarse y, ante todo, no puede renunciar a la audacia de plantear cuestiones que orienten la mirada hacia la auténtica estatura del hombre, evitando reduccionismos y espejismos en el horizonte, así como caer en la trampa de las ideologías. Se busca ejercitar lo que Joseph Ratzinger-Benedicto XVI denomina «razón abierta».⁶

¹ El origen de este texto se halla en una intervención en las Jornadas de Filosofía e Inteligencia Artificial *Filosof-IA*, que tuvieron lugar en la sede de la Facultat de Filosofia de Catalunya (URL), los días 27 y 28 de noviembre de 2019. Esta intervención respondía a la solicitud de incorporar la perspectiva de la teología en diálogo con la ciencia en el marco de unas jornadas filosóficas. Si bien el texto ha sido completamente reelaborado, se ha mantenido el mismo estilo de discurso para su publicación en esta revista filosófica.

² Cf. FRANCISCO, Exhortación apostólica *Evangelii Gaudium*, 134, y Constitución apostólica *Veritatis Gaudium*, 4c.

³ Nos situamos en la perspectiva de la búsqueda de un saber unitario y orgánico, convencidos de que esto es posible, tal como aparece en la encíclica *Fides et Ratio*, 85, de Juan Pablo II (1998).

⁴ Ian Barbour sitúa estos cuatro estadios en la relación entre ciencia y religión: conflicto, independencia, diálogo e integración. Cf. IAN BARBOUR, *El encuentro entre ciencia y religión*, entre otros.

⁵ Cf. N. BALLANTYNE, «Epistemic Trespassing».

⁶ Cf. M. CANTOS APARICIO, *Razón abierta: la idea de universidad en J. Ratzinger / Benedicto XVI*.

1.2. Centrado especialmente en la IA «fuerte»

La inteligencia artificial es un tema de gran actualidad,⁷ y es evidente que lo va a continuar siendo, junto con la cuestión del transhumanismo, con la que presenta campos de intersección.⁸

Puede decirse que hay dos maneras de entender la IA, o dos tipos de IA: la «débil» y la «fuerte», en una distinción acuñada por J. Searle. La IA «débil» (*weak*) es la ciencia-ingeniería que permite diseñar programas, máquinas y programarlas, para hacer cosas que se considera que requieren de inteligencia para ser hechas; es específica, no general.

La IA «fuerte» o «dura» busca replicar, e incluso superar, la inteligencia humana, planteando la llegada de un momento histórico en el que se produzca el punto de inflexión, conocido habitualmente como «singularidad», en un término tomado de la astrofísica. Sus partidarios afirman que tales máquinas no solamente podrán simular o imitar la inteligencia, sino que *serán verdaderamente* «inteligentes». Lograr una máquina con una «inteligencia» equiparable a la humana es el «Santo Grial» de la IA. Los partidarios de la IA fuerte plantean incluso que puede darse un cambio en el sustrato sobre el que se sitúa la inteligencia.⁹

El desarrollo de sistemas que puedan realizar eficientemente tareas asociadas al razonamiento (esto es, una IA «débil») parece no suscitar *a priori* interrogantes teológicos, dejando de lado las cuestiones morales. Al respecto, hallamos un precedente insigne en Ramon Llull, y su *Ars Magna*, como intento de crear un sistema mecánico que imitase el razonamiento lógico.¹⁰

Ahora bien, plantear la búsqueda de una «inteligencia artificial» en su versión «fuerte», con máquinas igualando o superando la inteligencia humana, lo cual supone una particular mirada sobre el hombre, viene precedido por la «predicación» de sus propios «profetas», y su particular literatura «apocalíptica», que anuncia unos posibles escenarios futuros, incluyendo promesas de salvación o terribles vaticinios, esto requiere una

⁷ Para una aproximación general, cf. M. A. BODEN, *Inteligencia artificial*; R. LÓPEZ DE MÁNTARAS y P. MESEGUER, *Inteligencia artificial*; R. LÓPEZ DE MÁNTARAS, «Present i futur de la intel·ligència artificial»; H. BRIGHTON y H. SELINA, *Inteligencia artificial: una guía ilustrada*. Con respecto a la relación entre inteligencia artificial y la teología, pero con una buena perspectiva introductoria respecto a la IA: M. DOROBANTU, «Recent advances in Artificial Intelligence (AI) and some of the issues in the theology & AI dialogue»; J. L. RUIZ DE LA PEÑA, *Crisis y apología de la fe: Evangelio y nuevo milenio* (cap. IV: «El desafío antropológico»); F. J. GÉNOVA OMEDES, *Anne Foerst: dimensión religiosa de la búsqueda de inteligencia artificial* (primera parte, especialmente el capítulo tercero).

⁸ Con respecto al transhumanismo, nos centramos en la cartografía presente en A. J. DIÉGUEZ LUCENA, *Transhumanismo*, a la que nos referiremos en diversas ocasiones en las páginas que siguen. No hemos querido entrar en discusiones entre transhumanismo y posthumanismo, sino que remitimos al primero a nivel general, teniendo en cuenta la complejidad de la cuestión, que no es objeto de nuestro trabajo presente.

⁹ Cf. R. LÓPEZ DE MÁNTARAS, «Present i futur de la intel·ligència artificial», pp. 9ss.

¹⁰ Cf. E. COLOMER, «De Ramon Llull a la moderna informàtica»; Id., «Anne Foerst. El encuentro entre la teología e inteligencia artificial», p. 90.

consideración diferente. Indudablemente, este ámbito resulta más interesante para la teología, que, aun dentro de sus pausados ritmos, lleva ya años atenta a la cuestión de la IA. Dos figuras pioneras en este terreno, ambas informáticas y teólogas, pero muy diferentes entre sí, son la alemana A. Foerst¹¹ y la americana N. Herzfeld.¹² En lengua española, no podemos dejar de mencionar a J. L. Ruiz de la Peña.¹³

De hecho, la búsqueda de la IA «fuerte» puede situarse dentro del «transhumanismo», considerado en su dimensión tecnocientífica, si seguimos la clasificación de A. Diéguez. Por ello, muchas de las reflexiones que se están planteando con respecto al multiforme tema del transhumanismo, cada vez más presente en los debates y en el imaginario, pueden ser de interés para la reflexión teológica acerca de la IA.¹⁴

1.2.1. Basado en la Teología Fundamental y la Antropología Teológica

En estas líneas se dejan a un lado, deliberadamente, las cuestiones morales (que también podrían interpelar a la teología, en otra de sus dimensiones), para centrarnos en las preguntas que se sitúan preferentemente en los campos de la teología fundamental y de la antropología teológica. Aclaremos cuáles son estos campos.

Por un lado, la «teología fundamental»¹⁵ plantea las bases del conocimiento teológico, y afronta las relaciones entre la fe y la razón, contemplando al hombre como constitutivamente abierto para acoger la manifestación (Revelación) del Misterio. Busca «dar razón de la esperanza» que hay en nosotros.¹⁶

¹¹ En 1998 publicó un famoso artículo: A. FOERST, «Cog, a Humanoid Robot, and the Question of the Image of God», que resultó muy polémico. Posteriormente, en 2004, A. FOERST, *God in the machine: what robots teach us about humanity and God*, y después apenas ha vuelto a publicar sobre la cuestión.

¹² N. HERZFELD, *In our image*, de 2002, y varios trabajos posteriores.

¹³ J. L. Ruiz de la Peña (1937-1996), insigne asturiano, es una figura básica de la antropología teológica en lengua española. Con respecto a nosotros, destacan J. L. RUIZ DE LA PEÑA, *Imagen de Dios: antropología teológica fundamental*; *Id.*, «El desafío antropológico»; si su muerte no hubiera acaecido tan «pronto», con 59 años, seguramente hubiera hecho notables aportaciones a la cuestión. Recientemente, en nuestro contexto, hallamos los trabajos de F. J. Génova, que estudia las aportaciones de Anne Foerst, entre otros; cf. F. J. GÉNOVA OMEDES, *Anne Foerst: dimensión religiosa de la búsqueda de inteligencia artificial*; *Id.*, «Anne Foerst. El encuentro entre la teología e inteligencia artificial»; *Id.*, F. J. GÉNOVA OMEDES, «Inteligencia artificial i teologia». También sería necesario tener en cuenta a J. Sánchez-Cañizares («La Inteligencia Artificial vista desde la filosofía y la teología»).

¹⁴ Cf. A. J. DIÉGUEZ LUCENA, *Transhumanismo*. Este autor distingue entre dos tipos de transhumanismo, el «cultural» y «tecnocientífico», dentro del cual engloba tanto el vinculado a la inteligencia artificial como el vinculado a la biología y la medicina (biomejoramiento). En una línea parecida se halla el trabajo previo: H. TIROSH-SAMUELSON, «Transhumanism as a secularist faith», que encabeza un número monográfico de *Zygon*[®] sobre el transhumanismo. N. López Moratalla considera la IA y el transhumanismo como dos caras de la misma moneda, aunque sus desarrollos impliquen aparentemente expectativas de signo contrario; cf. N. LÓPEZ MORATALLA, *Inteligencia Artificial ¿Conciencia Artificial?*, p. 10.

¹⁵ Entre las obras de referencia, cf. S. PIÉ I NINOT, *Teología fundamental*.

¹⁶ Cf. 1 Pe 3, 15, la «carta magna» de la teología fundamental: «Más bien, glorificad a Cristo el Señor en vuestros corazones, dispuestos siempre para dar explicación (*apología*) a todo el que os pida una razón (*logos*) de vuestra esperanza».

Por otro lado, la «antropología teológica», en su sentido contemporáneo, busca ser una reflexión sobre el hombre a la luz de la fe cristiana, y agrupa los antiguos tratados de creación y pecado, la salvación por la gracia de Cristo y el final del hombre y de la historia (la «escatología»).¹⁷

Consideramos que estos son dos de los principales campos de los estudios teológicos que nos pueden ofrecer perspectivas que arrojen luz en el diálogo acerca de la inteligencia artificial. Todo ello, teniendo además en perspectiva las aportaciones del diálogo teología-ciencias (que, en cierto modo, se podría situar en el área de influencia de la teología fundamental).

Por último, debe reconocerse que, aunque no se hallen estrictamente dentro de la teología, los trabajos que se refieren al tema realizados dentro de los denominados «estudios religiosos» ofrecen numerosos elementos de interés y contribuyen a situar las preguntas.

2. Elementos de base para la reflexión

2.1. Unas fuentes poco convencionales para la reflexión teológica

Como se ha indicado, la teología fundamental reflexiona sobre las bases del conocimiento teológico, que presentan una dimensión objetiva, la Revelación (que hallamos asociada a la Escritura, la Tradición y, a su servicio, el Magisterio eclesial), junto con la dimensión subjetiva (la fe, de la que participan los teólogos a lo largo de la historia).

Ahora bien, si se atiende a los desafíos que presenta la inteligencia artificial a la teología, para elaborar una reflexión no basta con tener en cuenta los actuales conocimientos científicos y tecnológicos como base para la discusión, junto con las citadas «bases» propias del conocimiento teológico. Hay otros campos, menos sólidos si se quiere, pero que adquieren un protagonismo inusitado: la denominada «ciencia especulativa», que a menudo se presenta en el formato de la divulgación científica, y la «ciencia ficción».¹⁸ Sin duda, son estos unos peculiares lugares para el desarrollo de la reflexión teológica, incluso en el terreno de la relación entre teología y ciencias.

En primer lugar, la «ciencia especulativa» consiste en la reflexión científica acerca del futuro del progreso científico, partiendo de la realidad presente. Dentro de este ámbito hallamos las propuestas de numerosos autores (a menudo reputados expertos en sus trabajos científicos o técnicos) que, partiendo de sus conocimientos científicos actuales, realizan extrapolaciones sobre el futuro, no pocos de ellos con una fe ciega en el

¹⁷ En el campo de habla española, destaca la obra de J. L. Ruiz de la Peña y de L. F. Ladaria.

¹⁸ Cf. M. W. DELASHMUTT, «A better life through information technology? The techno-theological eschatology of posthuman speculative science».

mito ilustrado del progreso (en este caso, científico y tecnológico), planteando una esperanza en un futuro tecnológicamente superior. En este terreno, aunque con grandes diferencias entre ellos, hallamos matemáticos, informáticos, científicos, ingenieros y filósofos como A. Turing,¹⁹ M. Minsky, H. Moravec,²⁰ F. Tipler, R. Kurzweil²¹ o, salvando las distancias, N. Bostrom.²²

Como se ha indicado previamente, entre los defensores de la IA «fuerte», vinculados a la vertiente tecnocientífica del transhumanismo, hallamos a diversos «profetas», que ofrecen promesas de futuro y, en algunos casos, exhortaciones para vivir el presente. Muchos de ellos son personas geniales, que destacan enormemente por sus capacidades técnicas en su campo, pero que dan rienda suelta a su imaginación cuando se trata de hablar del futuro, en relación con lo que acabamos de denominar «ciencia especulativa», hasta realizar en algunos casos propuestas descabelladas.²³ Se trata de figuras que acaban desempeñando un profetismo «ambiguo»,²⁴ y que presentan incluso paralelismos con populares figuras del evangelismo actual.²⁵

En segundo lugar, la denominada «ciencia ficción», a diferencia del anterior discurso, es manifiestamente ficticia, y acostumbra a buscar el entretenimiento, aunque a menudo contiene formas de crítica social. En este campo hallamos escritores, como I. Asimov o P. K. Dick, junto con cineastas y autores de series,²⁶ e incluso creadores de videojuegos o de entornos de realidad virtual. La ciencia ficción se muestra como un terreno fecundo para el planteamiento de cuestiones clave, tanto de un supuesto futuro

¹⁹ A. M. TURING, «Computing Machinery and Intelligence». Traducido al español.

²⁰ H. MORAVEC, *Mind children: the future of robot and human intelligence*; H. MORAVEC, *Robot: mere machine to transcendent mind*. El primero se halla traducido al español.

²¹ R. KURZWEIL, *The Age of Spiritual Machines: when computers exceed human intelligence*; R. KURZWEIL, *The singularity is near when humans transcend biology*, entre otros. Ambos tienen traducción en lengua española.

²² N. BOSTROM, *Superintelligence: paths, dangers, strategies*. Traducido al español.

²³ Este contraste es puesto dramáticamente de manifiesto en A. J. DIÉGUEZ LUCENA, *Transhumanismo*, p. 85, donde se cita una entrevista a D. Hofstadter en *American Scientist*, en la que se compara las obras de autores como H. Moravec o R. Kurzweil con la mezcla de buena comida con excrementos de perro hasta hacerlos indistinguibles.

²⁴ Cf. A. J. DIÉGUEZ LUCENA, «Los profetas ambiguos». R. Geraci, a partir de diversas obras de «ciencia especulativa», ofrece una lista de las promesas de los profetas del ciberespacio: sociedad igualitaria, satisfacción de las necesidades, felicidad, e incluso mejor vida sexual; cf. R. M. GERACI, «Apocalyptic AI: Religion and the Promise of Artificial Intelligence», p. 149.

²⁵ Cf. R. COLE-TURNER, «The Singularity and the Rapture: Transhumanist and Popular Christian Views of the Future», en donde se comparan las promesas del transhumanismo tecnocientífico con las de diversos evangelistas de la América contemporánea, especialmente por lo que respecta a la cuestión escatológica.

²⁶ En el tercer capítulo de F. J. GÉNOVA OMEDES, *Anne Foerst: dimensión religiosa de la búsqueda de inteligencia artificial*, bajo el título de «Mitología», se analizan las propuestas de la literatura y el cine, antes de tratar la historia de la IA.

como del presente, al que se enfrenta con una gran libertad, debido a la distancia que plantea el propio género.²⁷

2.2. Una «fe» secular

La teología fundamental atiende a la fe, como principio subjetivo del conocimiento teológico, que puede ser vista, dentro del cristianismo, como relación con Dios por medio de Jesucristo, en términos de confesión, confianza, camino de vida y horizonte escatológico. Esta relación se sitúa en consonancia con la estructura antropológica, constitutivamente abierta al absoluto.²⁸ La consideración teológica de la fe nos puede ser de ayuda para considerar el tema que nos ocupa. Asimismo, no podemos olvidar la relación que se establece entre religión y tecnología, que incluso puede acabar dotando a esta última de rasgos religiosos.²⁹

Si tenemos en cuenta que los avances tecnológicos son, en cierto modo, un «espejo» en el que mirarnos, la dimensión «religiosa» presente en la IA implica que esta puede ser tomada como referencia para reflexionar acerca de la experiencia religiosa.³⁰

A pesar de que la mayoría de los defensores de la búsqueda y profecía de la IA «fuerte» (como parte del transhumanismo tecnocientífico) se confiesen ateos y lo consideren como una extensión del humanismo secular, sin embargo esta postura aparece con características propias de una fe, pero una «fe secular», sin Dios, en la que los rasgos religiosos se entremezclan con las propuestas científicas.³¹ Esta hibridación podría ser considerada como característica del mundo «post-secular» (entendido como aquel que se sitúa más allá de los presupuestos ilustrados). Presenta un fuerte impulso escatológico, que confía en los esfuerzos humanos, y no en la intervención divina, eso sí, y que se tratará más adelante. Posee sus propios dogmas, como la innovación tecnológica, ofrece relatos de sentido, y busca nuevos adeptos. De hecho, puede considerarse que el

²⁷ Sobre el valor de la ciencia ficción respecto al tema que nos ocupa, cf. R. M. GERACI, «Robots and the Sacred in Science and Science Fiction: Theological Implications of Artificial Intelligence»; *Id.*, «The popular appeal of apocalyptic AI», entre otros textos del mismo autor.

²⁸ S. PIÉ I NINOT, *Teología fundamental*, capítulos 2 («El hombre abierto a la cuestión de Dios») y 5 («La fe, principio subjetivo del conocimiento teológico»).

²⁹ Cf. el ya «clásico» D. F. NOBLE, *The religion of technology: the divinity of man and the spirit of invention*; existe traducción española.

³⁰ Cf. M. DOROBANTU, «Recent advances in Artificial Intelligence (AI) and some of the issues in the theology & AI dialogue», p. 14. La idea del «espejo» se halla desarrollada en P. HEFNER, «Technology and Human Becoming».

³¹ El transhumanismo tecnocientífico acostumbra a entrar en conflicto con las religiones tradicionales de Occidente, mientras que puede encontrar puntos de contacto con los denominados Nuevos Movimientos Religiosos (y los mormones), como se explicita en H. TIROSH-SAMUELSON, «Transhumanism as a secularist faith».

progreso en la IA supone un «re-encantamiento» del mundo.³² Hay quien ha señalado incluso los nexos entre determinadas propuestas actuales y la filosofía hegeliana.³³

3. El surgir de los interrogantes

Plantear la búsqueda de una «inteligencia artificial» conlleva el reclamo de una reflexión sobre qué es la inteligencia humana, en particular, y qué es el hombre en general.³⁴

La misma expresión «inteligencia artificial» permite orientar el planteamiento de algunos interrogantes, ante los que puede situarse la reflexión teológica: ¿Qué es el hombre? ¿Qué es la inteligencia y qué significa que el hombre esté dotado de ella? ¿Qué nos hace humanos? ¿Qué nos hace diferentes de las máquinas?³⁵

Atendiendo a la famosa advertencia de Gregorio de Nisa: «los conceptos crean ídolos, solo el asombro conoce»,³⁶ podemos elevar nuestra pregunta partiendo del estupor, como hace el salmista ante el Señor:

«¿Qué es el hombre para que te acuerdes de él,
el ser humano, para mirar por él?
Lo hiciste poco inferior a los ángeles,
lo coronaste de gloria y dignidad;
le diste el mando sobre las obras de tus manos.
Todo lo sometiste bajo sus pies.» [Salmo 8, 5-7]

³² H. TIROSH-SAMUELSON, «Transhumanism as a secularist faith»; M. Y. CHAUDHARY, «Augmented Reality, Artificial Intelligence, and the Re-enchantment of the World»; R. M. GERACI, «Cyborgs, Robots, and Eternal Avatars»; R. M. GERACI, «The popular appeal of apocalyptic AI». En estos dos últimos trabajos, R. Geraci trae a colación, respecto al tema que nos ocupa, la curiosa expresión «*authentic fakes*» para referirse a prácticas fraudulentas que realizan funciones religiosas reales (promueven un sentido de transcendencia, generan comunidades, establecen una moral, orientan la acción humana), basado en D. CHIDESTER, *Authentic fakes: religion and American popular culture*.

³³ Cf. M. E. ZIMMERMAN, «The Singularity: A crucial phase in divine self-actualization?». En este caso, el autor se refiere a Ray Kurzweil.

³⁴ De hecho, el físico y teólogo J. Sánchez Cañizares se pregunta, después de haber analizado la cuestión detenidamente, si la expresión «Inteligencia Artificial» no será un oxímoron. Cf. J. SÁNCHEZ CAÑIZARES, «*La Inteligencia Artificial vista desde la filosofía y la teología*». También hallamos un rechazo a aplicar la noción de inteligencia a las máquinas en T. PETERS, «Where There's Life There's Intelligence»; Id., «Intelligence? Not Artificial, but the Real Thing!».

³⁵ P. Benanti pone de manifiesto que es clave reflexionar sobre qué es el hombre y en qué se diferencia de una máquina. Cf. P. BENANTI, *Le macchine sapienti*, capítulo II. Iguamente, N. LÓPEZ MORATALLA, *Inteligencia Artificial ¿Conciencia Artificial?*, p. 10.

³⁶ Cf. GREGORIO DE NISA, *De vita Moysis* [PG 44, col. 377B]; *Contra Eunomium* [PG 54, col p. 604B–D]; *In Cantica Cantorum*, Hom. XII [PG 44, col. 1028D].

3.1. «Conócete a ti mismo»

La pregunta *¿qué es el hombre?* ha atravesado los siglos, resonando en las diversas manifestaciones de la religión y la cultura.

En la Antigüedad hallamos el famoso lema «conócete a ti mismo»,³⁷ que se hallaba en el frontispicio del templo de Apolo en Delfos, y que Platón pone en boca de su maestro Sócrates.³⁸ Podría decirse que el adagio delfico se erige como tarea vital del hombre. La urgencia de este imperativo ha recorrido los siglos desde la Antigüedad, pasando por el gran Agustín, que inaugura el «socratismo cristiano»,³⁹ hasta llegar a Juan Pablo II, que sitúa el «conócete a ti mismo», de manera sorprendente, al inicio de su encíclica dedicada a las relaciones entre la fe y la razón, vinculándolo a las diversas tradiciones religiosas.⁴⁰

Sin duda, puede decirse que el trabajo en el terreno de la inteligencia artificial nos ha ayudado a conocernos mejor a nosotros mismos. Los pioneros de la IA, presentes en la conferencia de Dartmouth College en 1956, se proponían, entre otras cosas, una mejor comprensión de la inteligencia humana. De hecho, los avances en IA, especialmente aquellos que han partido de la referencia a la inteligencia humana, han contribuido —como es sabido—, podríamos decir que de manera sinérgica, a avanzar en nuestra comprensión del conocimiento humano.

3.2. La pregunta por el sentido

Ahora bien, el recorrido iniciado por el «conócete a ti mismo» nos lleva a ver al hombre como aquel que se plantea la pregunta sobre el sentido, pregunta que de algún modo reúne los grandes interrogantes del hombre,⁴¹ que podríamos agrupar sintéticamente en la tríada amor-muerte-futuro.⁴²

Consecuentemente, esta totalizante pregunta por el sentido resuena ante la pretensión de alcanzar una IA «fuerte». ¿De qué manera puede plantearse la existencia de tal IA, en relación con el imperativo del «conócete a ti mismo» y la pregunta por el sentido de la propia existencia?⁴³

³⁷ *Nosce te ipsum: γνῶθι σεαυτόν - gnóthi seautón.*

³⁸ PLATÓN, *Protágoras*, 343b.

³⁹ Cf. É. GILSON, *L'Esprit de la philosophie médiévale*, p. 214ss. Agustín de Hipona parte de la introspección para avanzar en el discurso teológico, como sucede, a modo de ejemplo, en su reflexión sobre la Trinidad.

⁴⁰ Cf. JUAN PABLO II, *Fides et Ratio*, 1.

⁴¹ Cf. JUAN PABLO II, *Fides et Ratio*, 1.26.33, especialmente.

⁴² Cf. K. RAHNER, *Curso fundamental sobre la fe*, 344-348, entre otros textos de este autor.

⁴³ En la película *Blade Runner*, de Ridley Scott (1982), el policía Rick Deckard, ante la muerte del último robot androide («replicante»), empatizando con los sentimientos de algunos de ellos, exclama: «No sé por qué me salvó la

Curiosamente, la pregunta de qué conciencia puede tener una máquina, que es una cuestión clave, había sido ya reconocida en 1950 por el mismísimo A. Turing.⁴⁴

3.3. Apertura al Misterio y deseo de infinito

Además, la reflexión teológica contempla al hombre como estructuralmente abierto al Misterio (a lo que podríamos denominar «la cuestión de Dios»; *potentia oboedientialis*), y también como marcado por el deseo del infinito (*deseo de Dios*).⁴⁵ Este deseo sitúa al hombre en una inquietud que sobrepasa su propia capacidad de respuesta y, en palabras de Agustín, solo halla su reposo en el Eterno.⁴⁶

Alto es, pues, el horizonte que pone en movimiento al inquieto corazón del hombre. De él toma fuerzas, sin duda, la búsqueda de la inteligencia artificial, y es preciso no olvidarlo ni diluirlo, ante determinados planteamientos de la IA «fuerte», que corren el riesgo, como mínimo, de rebajar la estatura del hombre y su deseo, censurándolo o reduciéndolo a la hora de plantear el objetivo a cumplir,⁴⁷ con la ilusión incluso de su manipulación: «¿Qué queremos desear?».⁴⁸

4. El hombre creado por Dios

El relato del Génesis ofrece datos de interés para el tema que nos ocupa.⁴⁹ A partir del texto bíblico, buscamos elementos que respondan a la pregunta antropológica. Se

vida. Quizá en esos últimos momentos él amó la vida con más intensidad que nunca, no solo su vida, la de cualquiera, mi vida. Y lo único que quería eran las mismas respuestas que el resto de nosotros: ¿de dónde vengo?, ¿a dónde voy?, ¿cuánto tiempo me queda? Todo lo que podía hacer era quedarme allí y verlo morir», citado en N. LÓPEZ MORATALLA, *Inteligencia Artificial ¿Conciencia Artificial?*, p. 80.

⁴⁴ Cf. A. M. TURING, *¿Puede pensar una máquina?*, pp. 46ss (la versión original es «Computing Machinery and Intelligence»); cf. J. L. RUIZ DE LA PEÑA, «El desafío antropológico», pp. 161ss.

⁴⁵ Cf. S. PIÉ I NINOT, *Teología fundamental*, capítulo II, pp. 65-97.

⁴⁶ Es el famoso tema del *Cor inquietum* de Agustín: «Nos hiciste, Señor, para ti, y nuestro corazón está inquieto, hasta que descansa en ti» (*Confesiones* I, 1, 1)

⁴⁷ «El corazón del hombre tiene demasiada nostalgia de belleza, de amor, de verdad, de felicidad. Por eso hemos tratado de rellenar el vacío con las utopías de crear superhombres o de reprogramar la Humanidad», en: N. LÓPEZ MORATALLA, *Inteligencia Artificial ¿Conciencia Artificial?*, p. 38.

⁴⁸ Esta pregunta aparece en Y. N. HARARI, *De animales a dioses*, p. 454. Recogido en A. J. DIÉGUEZ LUCENA, *Transhumanismo*, p. 29, donde se remite a Ortega, quien «ya había llegado hace tiempo a la conclusión de que la pregunta real a la que nos enfrentamos es “¿Qué queremos desear?”»; en los capítulos IV y V, A. J. Diéguez retoma la referencia a Ortega y la crisis del deseo.

⁴⁹ Dos textos que ofrecen un marco de referencia católico a nivel bíblico y teológico respecto a la cuestión antropológica a partir del relato del Génesis son, respectivamente: PONTIFICIA COMISIÓN BÍBLICA, *«Qué es el hombre?»: un itinerario de antropología bíblica* (2020); COMISIÓN TEOLÓGICA INTERNACIONAL, *Comunión y servicio: la persona humana creada a imagen de Dios* (2004).

parte además de la consideración de que la tecnología está siempre ligada a una antropología de referencia.⁵⁰

Somos conscientes de que estos elementos se hallan vehiculados por imágenes propias de su contexto original (el Oriente Medio de varios siglos antes de Cristo), que necesitan ser interpretadas adecuadamente más allá de la literalidad, pero no desechadas sin más como mitológicas o infantiles.

La fe bíblica nos presenta al hombre como un ser querido especialmente por Dios, que lo ha arrancado de la nada y lo sostiene en la existencia (*creatio ex nihilo*). El hombre (1) ha sido creado «a imagen de Dios» y (2) se le ha encomendado la custodia y dominio de la creación, (3) pero se halla herido por aquello que denominamos «pecado». Estos tres datos permiten plantear retos en el diálogo entre la teología y la búsqueda de la IA.⁵¹

4.1. Creado «a imagen de Dios»

4.1.1. La antropología cristiana de la imagen

El hombre se distingue de las otras creaturas por haber sido creado «a imagen de Dios», lo cual le hace poseedor de una forma particular de inteligencia (partícipe de la inteligencia de Dios),⁵² así como sujeto de conciencia y libertad.⁵³

Tradicionalmente, la teología cristiana se ha servido del «alma» para referirse al principio vital del hombre, que desborda la mera materialidad,⁵⁴ sin dejar por ello de afirmar el valor del cuerpo, en una unidad dual.⁵⁵

De hecho, los investigadores y «profetas» de la IA fuerte corren el riesgo de menospreciar la corporeidad, y caer en un nuevo gnosticismo de corte tecnológico, al reducir al ser humano a información, eventualmente transferible a otros sustratos distintos del biológico, lo cual presenta evidentes resonancias gnósticas.⁵⁶

⁵⁰ Cf. P. BENANTI, *Le macchine sapienti*, p. 91.

⁵¹ F. J. Génova («Intel·ligència artificial i teologia», p. 28) identifica en estos tres elementos las tres grandes líneas que podrían sintetizar los retos de la IA a la teología.

⁵² Cf. GS 15. Podríamos decir que, si todo fue creado por medio del *Logos* de Dios, Jesucristo, el hombre es asimismo sujeto de un *logos* particular.

⁵³ Cf., por ejemplo, *Gaudium et Spes*, 15-17.

⁵⁴ En torno a las controversias que suscitó el evolucionismo, se afirmó la creación del alma inmediatamente por Dios. Cf. Pío XII, *Humani Generis*, AAS 42 (1950), 575, DH 3896.

⁵⁵ Según la expresión «corpore et anima unus», GS 14. Cf. G. M. TOMÁS Y GARRIDO y E. POSTIGO SOLANA, *Bioética personalista: ciencia y controversias*, capítulos I y III.

⁵⁶ N. HERZFELD, «Human-Directed Evolution. A Christian perspective», pp. 592-593. El elemento corpóreo ha sido clave en las propuestas críticas del filósofo H. Dreyfus, como se hizo presente en H. L. DREYFUS, *What computers can't do: the limits of artificial intelligence*; más tarde apareció Id., *What computers still can't do: a critique of artificial reason*.

Además, la teología cristiana se ha apoyado sobre el concepto de «persona», que se ha ido enriqueciendo a lo largo de la historia, siendo usado inicialmente para Dios, para después ser fecundamente aplicado al hombre. Modernamente, hay quienes buscan ampliar el alcance de este término, no solamente a los animales, sino a las máquinas «inteligentes».

«Imagen de Dios», «alma» y «persona» son términos no exentos de debates y controversias, aunque nos permiten salvaguardar la dignidad y la singular riqueza de ser humanos, e introducen una mirada humilde frente a los logros de la técnica, ya que nos llevan a preguntarnos qué analogías o paralelismos establecer frente a una eventual IA «fuerte».

La reflexión teológica nos alerta contra el riesgo que suponen los reduccionismos: reducir el hombre a una máquina, su cerebro a un ordenador, y la historia personal a un conjunto de datos.⁵⁷ Y es que, ante el reto que supone la IA «fuerte», con el planteamiento de una nueva etapa, posterior a la «singularidad», la teología aporta una dosis de realismo y, a la vez, una defensa de lo que supone ser humanos, ante la pretensión de fabricar una inteligencia general que se equipare a la nuestra e incluso nos supere.

4.1.2. La fecundidad de la noción de «imagen de Dios» ante la IA

La creación «a imagen de Dios», que aparece sucintamente en el relato del Génesis, ha supuesto una enorme fecundidad teológica y ha sido objeto de controversias en la teología reciente. Además, la noción del hombre como creado «a imagen de Dios» resulta fecunda a la hora de plantear la creatividad humana (a la que se dedica el apartado siguiente), incluyendo el desarrollo de la IA.

Con respecto al tema que nos ocupa, la teóloga e informática N. Herzfeld ha puesto en paralelo tres distintas dimensiones de la imagen de Dios:⁵⁸ la substantiva (presente en la teología clásica, vinculada a la racionalidad), la funcional (propuesta por exégetas contemporáneos, vinculada al papel de «dominio» del hombre en la creación, como representante de Dios) y la relacional (que remonta a K. Barth, vinculada a Dios Trinidad).

En trabajos posteriores donde retoma sus propuestas, N. Herzfeld establece un interesante paralelismo entre las tres dimensiones de la «imagen de Dios» y las distintas fases del desarrollo de la inteligencia artificial: IA simbólica, IA funcional, IA vinculada a la corporeidad (*embodied*).⁵⁹

⁵⁷ Cf. F. J. GÉNOVA OMEDES, «Intel·ligència artificial i teologia», pp. 28-29.

⁵⁸ Cf. N. HERZFELD, *In our image*; esta división está presente también en muchos de sus trabajos posteriores, citados más adelante.

⁵⁹ Cf. N. HERZFELD, «A new member of the family? The continuum of being, artificial intelligence, and the image of God», con interesantes referencias a otras religiones; N. HERZFELD, «In Whose Image? Artificial Intelligence and the *Imago Dei*». La idea de la vinculación de la IA corporeidad, como *embodiment*, había sido planteada por el citado filósofo H. Dreyfus.

Otros autores han planteado también un modelo de carácter dinámico, con un horizonte cristológico, que tiene en cuenta a Cristo como arquetipo de la imagen de Dios y, por tanto, como fuente antropológica.⁶⁰ En línea con esto, también se subraya, siguiendo a Cristo, su amor libre como elemento clave que puede servir para plantear cuestiones a la IA.⁶¹

4.2. Colaborador de la Creación

Volviendo al relato bíblico, Dios sitúa al hombre como «colaborador», llamado a cuidar y hacer fecundo el cosmos creado.

«Sed fecundos y multiplicaos, llenad la tierra y sometedla; dominad los peces del mar, las aves del cielo y todos los animales que se mueven sobre la tierra» [Gn 1, 28].

En esta vocación podemos situar los desarrollos de la técnica, vistos como una cooperación del hombre con la divinidad. Pero aquí también podemos situar la creatividad humana, que permite desarrollar la música, la literatura o la pintura.

De hecho, se ha planteado la búsqueda de la IA, entre otras, como una consideración del hombre como «co-creador»,⁶² y aparece la cuestión de que, si nosotros hemos sido creados a imagen de Dios, a imagen *de quién* «creamos» nosotros. ¿Qué significa crear a *nuestra* imagen?⁶³ Debemos ser cuidadosos al considerar nuestra actividad creativa, para no acabar reduciendo, de manera refleja, la visión del hombre.

Hay quien podría describir determinados escenarios de la IA o el transhumanismo como «jugar a ser Dios».⁶⁴ Ahora bien, desde la teología es necesario advertir que existe una radical diferencia entre la creación por parte de Dios (*creatio ex nihilo*, que arranca de la nada y sostiene en el ser), y la «cooperación creativa» del hombre que, en todo caso, se encontraría en un nivel mucho más humilde.

Dicho esto, debemos recordar una de las principales cuestiones que se plantean ante una eventual IA «fuerte»: la de su eventual «creatividad» propia. No se trata de una objeción nueva, y de hecho ya había sido anticipada por los precursores y por los «padres» de la IA.⁶⁵

⁶⁰ Cf. M. S. BURDETT, «The Image of God and Human Uniqueness: Challenges from the Biological and Information Sciences».

⁶¹ Cf. I. DELIO, «Artificial Intelligence and Christian Salvation: Compatibility or Competition?».

⁶² Al respecto, podemos destacar la expresión «co-creador creado», de P. Hefner; cf. V. LORRIMAR, «The scientific character of Philip Hefner's "created co-creator"».

⁶³ Cf. los diversos textos citados de N. Herzfeld. Recientemente, cf. G. ZARKADAKIS, *In our own image: will artificial intelligence save or destroy us?*

⁶⁴ A. Diéguez es sumamente crítico con esta expresión. A. J. DIÉGUEZ LUCENA, *Transhumanismo*, pp. 135ss.

⁶⁵ Es conocida como la «objeción de Lady Lovelace». Ada Lovelace, que vivió en la primera mitad del siglo XIX, es considerada la iniciadora de la programación y una precursora de la IA. Esta objeción fue recogida en 1950 por A. Turing (*¿Puede pensar una máquina?*, p. 46ss.; la versión original es «Computing Machinery and Intelligence»). Cf. J. L. RUIZ DE LA PEÑA, «El desafío antropológico», p. 161ss.

Podemos hacer nuestra la llamada de humildad del salmista cuando, al hablar de los ídolos de las diversas naciones, advierte:

«Nuestro Dios está en el cielo, lo que quiere lo hace.
Sus ídolos, en cambio, son plata y oro, hechura de manos humanas:
tienen boca, y no hablan; tienen ojos, y no ven;
tienen orejas, y no oyen; tienen nariz, y no huelen;
tienen manos, y no tocan; tienen pies, y no andan; no tiene voz su garganta:
que sean igual los que los hacen, cuantos confían en ellos».
[Salmo 113b, 11-16 (114, 3-8)]

4.3. La cuestión del pecado y mal

En tercer lugar, después de describir el origen del hombre, el libro del Génesis, en sus capítulos 3 a 11, narra la disgregación de la situación inicial a causa de la irrupción del pecado, a través de conocidos relatos: el que corresponde a Adán y Eva (con la serpiente y la expulsión del paraíso, que supone la ruptura de la amistad con Dios, y la sospecha entre el hombre y la mujer); la historia de Caín y Abel (la discordia entre hermanos); la torre de Babel (la división de los pueblos a través de las lenguas), hasta llegar a la situación que precede al Diluvio Universal. Por esta caída, se ve alterada la situación originaria, y el mal entra en el mundo, y con él la discordia (con Dios y entre los hombres), la muerte, la dureza en el trabajo (y el estudio), o el dolor en el parto.

No podemos negar que, más allá de los detalles de los relatos, el Génesis ilumina la situación actual del hombre. Este «lado oscuro» es reconocido en las diversas tradiciones religiosas y de pensamiento, y una mirada realista hacia el hombre nos lleva a no censurarlo.⁶⁶

Esto no puede soslayarse ante los diversos avances tecnológicos y, en concreto, ante la búsqueda de una inteligencia artificial. Además, curiosamente, se plantea que la IA y, más ampliamente, el transhumanismo contribuirían a solucionar los elementos negativos que hemos enumerado como consecuencia de la caída. No faltan quienes (consciente o inconscientemente) plantean una suerte de retorno a una situación prelapsaria, e incluso su superación: más inteligentes, más rápidos, unidos por una nueva forma de lenguaje, e incluso elevando finalmente una torre de Babel para erigirnos en dioses.

No podemos tener una visión ingenua, ni ante la realidad humana, ni ante el supuesto horizonte de una IA «fuerte», que poseyera «algo así» como nuestra libertad.

⁶⁶ «Las tradiciones religiosas y de pensamiento han reconocido siempre que no todo es luminoso y encomiable en nuestra naturaleza; que hay en ella un lado oscuro que hemos de aprender a controlar, y que adolecemos de numerosas deficiencias de diseño que condicionan fuertemente nuestra vida, sobre todo a edades avanzadas», A. J. DIÉGUEZ LUCENA, *Transhumanismo*, pp. 146-147.

Resuena a menudo la pregunta de cómo controlar a nuestras propias «creaciones». Hay autores que incluso plantean que la desobediencia de Adán al Creador puede ser un aviso de lo que podría suceder al hombre como «creador» de una IA.⁶⁷ De todos modos, no hace falta especular con el futuro, pues los avances actuales de la IA ya suponen perspectivas cuanto menos inquietantes.⁶⁸

5. Salvación gratuita y encarnada

El cristianismo ofrece, como es sabido, elementos para plantear la salvación del hombre caído, restaurándolo y llevándolo a plenitud gracias a Jesucristo, Dios hecho hombre. Al respecto, encontramos diversas nociones, como salvación, redención, divinización, gracia.⁶⁹

La teología cristiana siempre ha defendido el papel de la carnalidad en el destino último del hombre, desde los primeros siglos de nuestra era, en los que se enfrentó al gnosticismo, que propugnaba una salvación que venía del conocimiento y que era una liberación de la materialidad.⁷⁰

Ante quienes proponen que solo la mente debe ser salvada, debemos recordar que el elemento fundamental del cristianismo es que el *Logos* de Dios se ha hecho carne en Jesucristo.⁷¹ Además, recuérdese, teniendo presente el símbolo apostólico, que en el cristianismo, la salvación final está asociada a la resurrección de la carne, que es mucho más que una revivificación del cuerpo.

Además, frente a la pretensión humana de alcanzar la salvación por las propias fuerzas, el cristianismo ha reconocido el carácter gratuito de esta salvación, con figuras como Pablo y Agustín.⁷²

⁶⁷ Cf. G. ZARKADAKIS, *In our own image: will artificial intelligence save or destroy us?* El autor recuerda que Adán recibió instrucciones precisas sobre lo que hacer y evitar, y aún así desobedeció.

⁶⁸ Cf. las perspectivas que ofrecen, entre muchos otros, P. BENANTI, *Le macchine sapienti*; M. DOROBANTU, «Recent advances in Artificial Intelligence (AI) and some of the issues in the theology & AI dialogue».

⁶⁹ No es posible detallar aquí la riqueza de estos términos, y su complejidad a lo largo de la historia de la teología. Puede consultarse un buen diccionario teológico, como J. Y. LACOSTE, *Diccionario Akal crítico de Teología*.

⁷⁰ Al respecto, destacan Ireneo de Lyon y Tertuliano (c. 160-c. 220). Es famosa la expresión de Tertuliano, «*caro salutis est cardo*» («la carne es el quicio de la salvación»), en *De resurrectione carnis*—o *De resurrectione mortuorum*—, VIII, 2 (PL 2, 806; CCL 2, p. 931).

⁷¹ Cf. I. DELIO, «Artificial Intelligence and Christian Salvation: Compatibility or Competition?», p. 39.

⁷² Hallamos aquí la dialéctica paulina entre la fe y las obras, así como la posterior controversia entre Agustín y Pelagio, en el primer tercio del siglo v, que ha marcado la historia de la teología cristiana occidental. El mismo papa Francisco, en numerosas ocasiones, ha denunciado lo que él considera la presencia del gnosticismo y el pelagianismo en el mundo actual. Cf. FRANCISCO, Exhortación apostólica *Evangelii Gaudium*; carta apostólica *Gaudete et Exultate* (19 de marzo de 2018); CONGREGACIÓN PARA LA DOCTRINA DE LA FE, *Carta «Placuit Deo» sobre algunos aspectos de la salvación cristiana* (22 de febrero de 2018).

Las antiguas tentaciones han reaparecido a lo largo de la historia, y las encontramos en algunos de los «profetas» de la IA «fuerte» (como H. Moravec y R. Kurzweil). Se rechaza lo físico para abrazar lo «virtual».⁷³

Una salvación, obtenida por medios técnicos, que supuestamente nos libere de la carne, para reducir nuestra historia a datos, no será verdadera salvación, porque no responde a la totalidad del hombre, ni al deseo de plenitud presente en su corazón. De hecho, más bien puede aparecer como una condena. Vemos aquí reactualizado el peligro constante de la antigua gnosis.

6. Escatología

Si hay un terreno en el cual la teología se pone especialmente en juego ante la idea de una IA «fuerte» y el profetizado advenimiento de la «singularidad», es el de la Escatología, esto es, la reflexión a partir de los datos revelados sobre las realidades últimas (*novissima*, «postrimerías»): el fin de la historia y el destino final del hombre (el juicio, la resurrección de la carne, el cielo y el infierno).

Es necesario reconocer la limitación de las imágenes y concepciones que podemos proponer al respecto de «lo que está por llegar», aunque se basen en la Escritura, o hayan sido explotadas por el arte o la predicación. Podría decirse que la escatología ha aprendido a ser humilde (especialmente a lo largo del siglo XX) y a hablar del «futuro» con cautela. Durante el siglo XX, la escatología ha pasado a ocupar un lugar central en la teología. Se han ido abandonando las concepciones previas que, en palabras de Y. Congar, caían en una «física de las postrimerías», para situar el horizonte escatológico dentro del discurso teológico.⁷⁴

6.1. Apocalíptica e IA

A menudo, estas imágenes «futuras» provienen de la denominada «literatura apocalíptica» judeocristiana, que se halla presente tanto en el Antiguo como en el Nuevo Testamento y en la denominada «literatura intertestamentaria». Esta literatura relata visiones espectaculares, supuestamente del porvenir, pero que ante todo buscarían aportar claves para vivir las dificultades del presente, especialmente en tiempos de persecución.

Con respecto a los paralelismos entre la literatura apocalíptica y la IA, destacan los trabajos del historiador de las religiones R. Geraci.⁷⁵ Según este autor, los elementos

⁷³ R. M. GERACI, «Apocalyptic AI: Religion and the Promise of Artificial Intelligence» (2008), p. 160.

⁷⁴ Cf. A. NITROLA, *Trattato di escatologia*.

⁷⁵ Cf., entre otros: R. M. GERACI, «Spiritual Robots: Religion and Our Scientific View of the Natural World» (2006); Id., «Apocalyptic AI: Religion and the Promise of Artificial Intelligence» (2008); Id., *Apocalyptic AI: visions of heaven in robotics, artificial intelligence, and virtual reality* (2010).

clave de la «literatura apocalíptica» pueden sintetizarse en tres: (1) la experiencia de alienación dentro del mundo presente, (2) la llegada de un Nuevo Mundo celestial, y (3) la transformación de los seres humanos para poder vivir en tal mundo. Esta transformación implica reconsiderar el elemento corporal.⁷⁶

Hallamos numerosos «profetas» de la inteligencia artificial «fuerte», defensores de la «singularidad»⁷⁷ que, extrapolando a partir de los avances tecnológicos actuales, ofrecen un panorama futuro en el que las máquinas y la IA ejercen un papel cuanto menos inquietante con respecto a la humanidad. Esto presenta notables paralelismos con la apocalíptica «judeocristiana». Siguiendo con los tres elementos citados:

(1) La alienación serían las restricciones del cuerpo humano, la enfermedad y la muerte.

(2) El nuevo mundo («Reino Virtual») llegaría con la singularidad (posiblemente con conflictos). Hay autores que plantean un futuro más apacible que otros, mientras que no es raro encontrar propuestas de tipo «milenario», con una escatología en dos estadios (un «reino de los robots», seguido por un «reino de la mente»), como H. Moravec.⁷⁸

(3) Se llegaría a una reconfiguración corporal para participar de este nuevo Reino, ya sea a través de ciborgs, robots o desligándonos de la dependencia a un substrato físico (como *software* y datos).⁷⁹

Es necesario poner de manifiesto que en la «IA apocalíptica» se halla ausente (o, más bien, deformado) un elemento fundamental de la literatura apocalíptica judía y cristiana: la intervención salvadora de Dios, que queda sustituida por los empeños humanos y por la evolución dirigida por la técnica.⁸⁰

6.2. ¿Un auténtico horizonte último y de plenitud?

No podemos negar que algunas de las más famosas «distopias» literarias y cinematográficas ya clásicas presentan predicciones que hoy nos parecen menos fantásticas y más inquietantes que cuando aparecieron. Como se ha indicado, la teología ayuda a poner de manifiesto que los defensores de la inteligencia artificial «fuerte» (y del transhumanismo, por tanto, en su vertiente tecnocientífica, al menos) parten de una base religiosa, aunque secularizada, pues ya no se remite a la intervención salvadora de Dios,

⁷⁶ Cf. R. M. GERACI, «Apocalyptic AI: Religion and the Promise of Artificial Intelligence», p. 138.

⁷⁷ Como H. Moravec o R. Kurzweil, pero atendiendo también aquí a los autores de ciencia ficción.

⁷⁸ Cf. A. J. DIÉGUEZ LUCENA, *Transhumanismo*, p. 60.

⁷⁹ R. M. GERACI, «Apocalyptic AI: Religion and the Promise of Artificial Intelligence» (2008).

⁸⁰ Cf. R. M. GERACI, «Apocalyptic AI: Religion and the Promise of Artificial Intelligence», p. 159.

sino a una evolución (en sentido amplio) que sacraliza los avances técnicos.⁸¹ ¿Pero se trataría de una verdadera escatología, ya que aún habría espacio para un «más allá» que respondiera a la cuestión del hombre en toda su radicalidad dando reposo a su corazón inquieto?

Con respecto a la especulación «científica» asociada a una búsqueda y promesa de una IA «fuerte», en línea con lo que se ha puesto de manifiesto al hablar del nexo entre IA y «apocalíptica», la teología ayuda a desenmascarar algunos errores: (1) la certeza de tipo positivista en las capacidades futuras que tendrá la tecnología de la información para facilitar una escatología tecno-teológica; (2) la aceptación acrítica del mito de la tecnología y del progreso tecnológico; (3) la idea de que la tecnología, como mito, puede facilitar la «preocupación última».⁸²

Perdurar supuestamente a través de la computación, en una suerte de «inmortalidad cibernética», es radicalmente diferente a la resurrección de la carne. Posponer la muerte es algo muy diferente a vencer definitivamente sobre ella. En el cristianismo, paradójicamente, la muerte es considerada como el enemigo último a ser vencido, a la vez que la misma muerte es el paso para la salvación eterna, alcanzada a través de la resurrección de la carne. Por tanto, el horizonte de plenitud en el cristianismo se encuentra más allá del proceso temporal e histórico (del que no puede desligarse la perdurabilidad a través de medios tecnológicos). La esperanza no consiste simplemente en más tiempo en este mundo, ni en la transmigración del cuerpo a otro soporte, sino en una nueva creación que incluye cuerpo y alma.⁸³

7. Conclusión

A través de un recorrido por algunos aspectos de la teología (teología fundamental y antropología teológica), se han presentado una diversidad de cuestiones que plantean interrogantes a la búsqueda de la IA, especialmente en el caso de la IA «fuerte», que se ha situado como vinculada al transhumanismo tecnocientífico.

En primer lugar, se ha tomado en consideración la singularidad de las fuentes a considerar, así como las características religiosas (secularizadas) de la búsqueda de la IA «fuerte».

⁸¹ A. Diéguez (A. J. DIÉGUEZ LUCENA, *Transhumanismo*, p. 27), refiriéndose al transhumanismo, pero con términos compatibles con una IA «fuerte», afirma: «La vida eterna no está en el más allá, sino aquí mismo, al alcance de nuestra mano, y es la tecnología la que puede proporcionárnosla. No es un premio a toda una vida haciendo el bien, o aceptando con humildad la voluntad divina, es algo que nos merecemos todos, por la sencilla razón de que estamos cerca de alcanzar el conocimiento suficiente para obtenerlo».

⁸² Cf. M. W. DELASHMUTT, «A better life through information technology? The techno-theological eschatology of posthuman speculative science», p. 284.

⁸³ Cf. M. W. DELASHMUTT, «A better life through information technology? The techno-theological eschatology of posthuman speculative science», p. 285; N. HERZFELD, «Human-Directed Evolution. A Christian perspective», p. 594.

Seguidamente, se ha puesto de manifiesto el valor de las grandes preguntas y el deseo del hombre, que no pueden ser censurados ni diluidos al plantear la búsqueda de una inteligencia artificial.

Más adelante, la creación del hombre a «imagen de Dios» y la colaboración del hombre, al modo humano, con la creación divina, nos ponen ante la cuestión de la creatividad humana y «artificial».

Finalmente, el planteamiento del escenario de la salvación, gratuita y en la carnalidad, y de la escatología, que se ve vehiculado por imágenes apocalípticas, muestra cómo algunas propuestas de los partidarios de la IA «fuerte» y de la «singularidad» son, en realidad, fruto de la secularización de elementos religiosos, cuyo trasfondo la teología debe ayudar a desvelar.

Ante el hombre marcado por el desafío tecnológico, la teología, como antigua sabiduría que busca poner de manifiesto la estatura de la vida humana, llamada a la Vida, con mayúsculas, tiene como tarea recordar al hombre su origen y destino, sin ceder a espejismos que pueden resultar engañosos. Podemos concluir citando al poeta T. S. Eliot, que en 1934 escribía:

«¿Dónde está la vida que hemos perdido viviendo?
¿Dónde está la sabiduría que hemos perdido en conocimiento?
¿Dónde está el conocimiento que hemos perdido en información?»
[T. S. ELIOT, *Los coros de «la Piedra»* (1934)]⁸⁴

Bibliografía

- BALLANTYNE, N., «Epistemic Trespassing». *Mind*, 128, 2019, pp. 367-395.
- BENANTI, P., *Le macchine sapienti: intelligenza artificiale e decisioni umane*. Milán: Marietti 1820, 2018.
- BODEN, M. A., *Inteligencia artificial*. Noema. Madrid: Turner, 2017.
- BOSTROM, N., *Superintelligence: paths, dangers, strategies*. Oxford: OUP, 2017^l.
- BRIGHTON, H.; SELINA, H., *Inteligencia artificial: una guía ilustrada*. Madrid: Tecnos, 2019.
- BURDETT, M. S., «The Image of God and Human Uniqueness: Challenges from the Biological and Information Sciences». *The Expository Times*, 127, 2015, pp. 3-10.
- CANTOS APARICIO, M., *Razón abierta: la idea de universidad en J. Ratzinger / Benedicto XVI*. Fe y razón. Estudios y ensayos 2, Madrid: UFV-BAC, 2015.
- CHAUDHARY, M. Y., «Augmented Reality, Artificial Intelligence, and the Re-enchantment of the World», *Zygon*[®], 54, 2019, pp. 454-478.
- CHIDESTER, D., *Authentic fakes: religion and American popular culture*. Berkeley, Calif.: University of California Press, 2005.
- COLE-TURNER, R., «The Singularity and the Rapture: Transhumanist and Popular Christian Views of the Future». *Zygon*[®], 47, 2012, pp. 777-796.
- COLOMER, E., «De Ramon Llull a la moderna informàtica». *Estudios Lulianos*, 1979, pp. 113-135.

⁸⁴ T. S. ELIOT, *Poesías reunidas 1909-1962*, 169.

- COMISIÓN TEOLÓGICA INTERNACIONAL, *Comunión y servicio: la persona humana creada a imagen de Dios; En busca de una ética universal: nueva perspectiva sobre la ley natural*. Biblioteca de autores cristianos. Documentos 40, Madrid: BAC, 2009.
- DELASHMUTT, M. W., «A better life through information technology? The techno-theological eschatology of posthuman speculative science». *Zygon*®, 41, 2006, pp. 267-288.
- DELIO, I., «Artificial Intelligence and Christian Salvation: Compatibility or Competition?». *New Theology Review*, 16, 2003, pp. 39-51.
- DIÉGUEZ LUCENA, A. J., *Transhumanismo: la búsqueda tecnológica del mejoramiento humano*. Barcelona: Herder, 2017.
- , «Los profetas ambiguos». *Claves de razón práctica*, 2018, pp. 22-31.
- DOROBANTU, M., «Article Review: Recent advances in Artificial Intelligence (AI) and some of the issues in the theology & AI dialogue». *ESSSAT News & Reviews*, 29, 2019, pp. 4-17.
- DREYFUS, H. L., *What computers can't do: the limits of artificial intelligence*. Harper Colophon Books, Nueva York: Harper & Row, 1979^{Rev.}
- , *What computers still can't do: a critique of artificial reason*. Cambridge, Mass. / Londres: MIT Press, 1992^{III}.
- ELIOT, T. S., *Poesías reunidas 1909-1962*. Madrid: Alianza, 1999.
- A. FOERST, A., «Cog, a Humanoid Robot, and the Question of the Image of God». *Zygon*®, 33, 1998, pp. 91-111.
- , *God in the machine: what robots teach us about humanity and God*. Nueva York: Dutton, 2004.
- GÉNOVA OMEDES, F. J., *Anne Foerst: dimensión religiosa de la búsqueda de inteligencia artificial*. Col·lectània Sant Pacià: FTC-AUSP 113, Barcelona, 2016.
- , «Anne Foerst. El encuentro entre la teología e inteligencia artificial». *Salmanticensis*, 64, 2017, pp. 313-338.
- , «Intel·ligència artificial i teologia». *Temes d'avui: revista de teologia i qüestions actuals*, 2017, pp. 26-32.
- GERACI, R. M., «Spiritual Robots: Religion and Our Scientific View of the Natural World». *Theology and Science*, 4, 2006, pp. 229-246.
- , «Robots and the Sacred in Science and Science Fiction: Theological Implications of Artificial Intelligence». *Zygon*®, 42, 2007, pp. 961-980.
- , «Apocalyptic AI: Religion and the Promise of Artificial Intelligence». *Journal of The American Academy of Religion*, 76, 2008, pp. 138-166.
- , *Apocalyptic AI: visions of heaven in robotics, artificial intelligence, and virtual reality*. Nueva York / Oxford: OUP, 2010.
- , «The popular appeal of apocalyptic AI». *Zygon*®, 45, 2010, pp. 1003-1020.
- , «Cyborgs, Robots, and Eternal Avatars», en: HAAG, J. W. - PETERSON, G. R. - SPEZIO, M. L., (eds.), *The Routledge Companion to Religion and Science*. Londres / Nueva York: Routledge, 2012, pp. 578-590.
- GILSON, É., *L'Esprit de la philosophie médiévale*. París: Vrin, 1969^{II}.
- HARARI, Y. N., *De animales a dioses: breve historia de la humanidad*. Barcelona: Debate, 2014^{III}.
- HEFNER, P., «Technology and Human Becoming». *Zygon*®, 37, 2002, pp. 655-666.
- HERZFELD, N., *In our image: artificial intelligence and the human spirit*. Theology and the sciences, Minneapolis, MN: Fortress Press, 2002.
- , «A new member of the family? The continuum of being, artificial intelligence, and the image of God». *Theology and Science*, 5, 2007, pp. 235-247.
- , «Human-Directed Evolution. A Christian perspective», en: HAAG, J.W.; PETERSON, G.R.; SPEZIO, M.L. (eds.), *The Routledge Companion to Religion and Science*. Londres / Nueva York: Routledge, 2012, pp. 591-601.
- , «In Whose Image? Artificial Intelligence and the *Imago Dei*», en: STUMP, J.B.; PADGETT, A. (eds.), *The Blackwell Companion to Science and Christianity*. Chichester: Wiley-Blackwell 2012, pp. 500-509.

- KURZWEIL, R., *The Age of Spiritual Machines: when computers exceed human intelligence*. Nueva York: Viking, 1999.
- , *The singularity is near when humans transcend biology*. Nueva York: Penguin, 2005.
- LACOSTE, J. Y. (ed.), *Diccionario Akal crítico de Teología*. Madrid: Akal, 2007.
- LÓPEZ DE MÁNTARAS, R., «Present i futur de la intel·ligència artificial». *Temes d'avui: revista de teologia i qüestions actuals*, 2017, pp. 9-19.
- LÓPEZ DE MÁNTARAS, R.; MESEGUER, P., *Inteligencia artificial. ¿Qué sabemos de? 87*, Madrid: Los libros de la catarata – CSIC, 2017.
- LÓPEZ MORATALLA, N., *Inteligencia Artificial ¿Conciencia Artificial? una perspectiva desde las ciencias de la vida*. Madrid: Digital Reasons, 2017.
- LORRIMAR, V., «The scientific character of Philip Hefner's "created co-creator"». *Zygon®*, 52, 2017, pp. 726-746.
- MORAVEC, H., *Mind children: the future of robot and human intelligence*. Cambridge, Mass.: Harvard University Press, 1988.
- , *Robot: mere machine to transcendent mind*. Science/Computers, Oxford: OUP, 1999.
- NITROLA, A., *Trattato di escatologia*, I. Cinisello Balsamo (MI): San Paolo, 2001.
- NOBLE, D. F., *The religion of technology: the divinity of man and the spirit of invention*. Nueva York: Penguin, 1999.
- PETERS, T., «Where There's Life There's Intelligence», en: LOSCH, A. (ed.), *What is Life? On Earth and Beyond*. Cambridge: CUP, 2017, pp. 236-259.
- , «Intelligence? Not Artificial, but the Real Thing!». *Theology and Science*, 17, 2019, pp. 1-5.
- PIÉ I NINOT, S., *Teología fundamental*. Sapientia fidei 37, Madrid: BAC, 2016.
- PONTIFICIA COMISIÓN BÍBLICA, «*Qué es el hombre?»: un itinerario de antropología bíblica*. BAC documentos 75, Madrid: BAC, 2020.
- RAHNER, K., *Curso fundamental sobre la fe: introducción al concepto de cristianismo*. Barcelona: Herder, 1979.
- RUIZ DE LA PEÑA, J. L., *Imagen de Dios: antropología teológica fundamental*. Santander: Sal Terrae, 1988.
- , *Crisis y apología de la fe: Evangelio y nuevo milenio*. Santander: Sal Terrae, 1995.
- , «El desafío antropológico», en: ID., *Crisis y apología de la fe: Evangelio y nuevo milenio*. Santander: Sal Terrae, 1995, pp. 155-209.
- SÁNCHEZ CAÑIZARES, J., «La Inteligencia Artificial vista desde la filosofía y la teología», *Fronteras CTR* [On-line edition], 2018.
- TIROSH-SAMUELSON, H., «Transhumanism as a secularist faith». *Zygon®*, 47, 2012, pp. 710-734.
- TOMÁS Y GARRIDO, G. M.; POSTIGO SOLANA, E., *Bioética personalista: ciencia y controversias*. Madrid: Ediciones Internacionales Universitarias, 2007¹.
- TURING, A. M., «Computing Machinery and Intelligence». *Mind*, 59, 1950, pp. 433-460.
- , *¿Puede pensar una máquina?* Cuadernos Teorema 1, Valencia: Universidad de Valencia, 1974.
- ZARKADAKIS, G., *In our own image: will artificial intelligence save or destroy us?* Londres: Rider Books, 2015.
- ZIMMERMAN, M. E., «The Singularity: A crucial phase in divine self-actualization?». *Cosmos and History*, 4, 2008, pp. 347-370.

Juan Ramón LA PARRA
 Ateneu Universitari Sant Pacià (Barcelona)
 juanramonlaparra@gmail.com

Article rebut: 12 de novembre de 2020. Article acceptat: 1 de febrer de 2021

Jaume AGUSTÍ-CULLELL
Marco SCHORLEMMER

Abstract

We try to give an answer to the following question: What should we take account of when designing and using AI systems so they contribute to the development of the socially embodied human intelligence on which human fulfilment depends? First, we consider the current impact of AI on the development of human intelligence showing some of its ambiguities. The deterioration of human intelligence is the prospect that should frighten us, not the rise of ‘super-intelligent’ machines. It is our contention that computers do not understand anything: AI systems only compute. In order to avoid misunderstandings, we need to accurately distinguish between intelligence, knowledge, information, and data. We denounce the abuses of metaphors that attribute human qualities and powers such as experience, knowledge, and intelligence to machines by stretching their meaning excessively. We glance through the two main AI models of intelligence—the representational model based on ‘knowledge representation’ and the connectionist model based on ‘artificial neural networks’—and reflect on the necessity for combining the best of each. Finally, we consider the need to govern AI in ways that will be truly beneficial for all of humanity.

Key words: Intelligence, Knowledge, Information, Data Processing, Intelligent Machines.

1. Introduction

Rather than providing an abstract and neutral presentation of Artificial Intelligence (AI), we set out to make an evaluative reflection, with the purpose of clarifying AI’s cultural situation. Predominantly, the perspective on AI offered here is not technoscientific from inside the field, but rather from the outside, in order to locate it in relation to human intelligence. A particular aim is to understand the importance of AI, the values and counter-values it generates, in order to contribute to its effective governance. We offer a critical view, but not with the intention of belittling AI’s huge value. Our aim is to help it advance in what we believe is the right direction. The massive current hype around AI is not helping its healthy development.

1.1. Some key questions

Information, especially information-processing by computers, has become increasingly powerful and pervasive in all human activities. The power of computers dazzles us. The first author still remembers his amazement when, in the mid 1970s as part of his PhD in Physics, he was programming an Intel 8080 microprocessor in binary code, using a teletype and paper ribbon, to control the data acquisition in atomic collision experiments. Computer utility has fascinated us so much that we have tended to abuse it. The temptation is to use it like a magic wand that will tell us our individual and collective future. But we remain unaware of its limitations, the possible perversions of its usage, the hidden designs it has on us.

Empowered by this fascination, along with its undoubtedly useful capabilities, Artificial Intelligence sets out to capture the very heart of humanity, the intelligence that makes us human. This is a living intelligence, the foremost embodiment of universal intelligence, something that has been constantly evolving for millions of years. Seen from this perspective, AI might be said to have the most ambitious, even presumptuous goal of all the techno-sciences: to understand, describe, reproduce and even improve this amazing human intelligence by the use of computing machines. An ambition we can accurately describe as overweening.

Thanks to AI, it seems that everything will become ‘intelligent:’ phones, cars, homes, factories and cities are just the most obvious examples. We already rely on computing devices to navigate us through city streets, recommend movies to us, and provide us with answers to search queries. However, a question arises. What effect will all these AI products and services—mostly designed according to an economic rationale—have on humanity? From the perspective of intelligence, the primary question a wisdom-based AI needs to face is clear. Instead of asking itself how to emulate human intelligence, the wise question is this: how can AI systems be designed and deployed so that they contribute to the development of a socially embodied human intelligence—something on which all human fulfilment depends?

AI studies functional intelligent behaviour, in particular the accomplishing of goals by means of data processing. Given this fundamental understanding, it is clear that AI needs to be guided in ways that will be beneficial for all of humanity. AI only makes sense when integrated with the development of this socially embodied human intelligence.

1.2. Uses and misuses of AI

The current impact of AI on the development of human intelligence shows some ambiguities. We see very positive uses of those AI systems which are at the service of our creative intelligence, extending its power and reach. Think of the computer assis-

tant and other apps in your phone, which save you from endless tedious tasks and offer easy access to all kinds of information, freeing you up to concentrate on your creative work. Even more significantly, AI systems can be employed to deliver high-quality education to people wherever they are on Earth.¹ AI also makes possible the creation and renewal of human teams and *teams of teams*, thus facilitating the most powerful form in which intelligence can be exercised.²

However, there are also many clear misuses of AI which tend to work in exactly the opposite direction. The result is that each of us becomes a consumer of information, no longer fully exercising our own intelligence, now degraded into one which is merely information-programmed.³ Rather than directly observing and controlling our life through collective intelligence, we credulously look for second-hand information to solve our problems.⁴ The terrible effect on adolescents and young adults produced by the time spent online in social networks and the abuse of electronic devices is all too obvious.⁵ Unconscious of the power of monopolistic information-technology corporations such as Google, Facebook and many other hugely profitable data-extraction firms, most of us all too frequently rely uncritically on what they tell us, rather than using them with a cautious and critical attitude, guarding our privacy. By merely *consuming* information, we end up leading second-hand lives, at the mercy of the powers of domination, plutocracy and imperialism. This is the way that not only information, but also the centralised power of modern states, is controlled. We become puppets in the violent hands of dominating forces.⁶

The massive degradation of human intelligence is the possibility that should alarm us, not that of a supposed super-intelligence of machines. Getting humans to conform to machine behaviour, rather than the other way around, is much easier to achieve and

¹ UNESCO, Beijing Consensus on Artificial Intelligence and Education, 2019. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000368303>. Last accessed on February 8, 2021.

² E. ANDREJCZUK, F. BISTAFFA, C. BLUM, J.A. RODRÍGUEZ-AGUILAR and C. SIERRA, «Synergistic team composition: A computational approach to foster diversity in teams». *Knowledge-Based Systems*, 182:104799, 2019.

³ This is a consequence of *computationalism* and its emphasis on *effective computability*, i.e., «the quest for universal knowledge and perfect self-knowledge», to «make cultural practices not just computational but programmable—susceptible to centralized editing and revision»: E. FINN, *What Algorithms Want*. The MIT Press, 2017.

⁴ AI-governed social interaction in privately-owned social network platforms that follow their own economic rationale without any regard for truth is no longer a 'marketplace of ideas' in which freedom of speech can flourish. This contributes to political polarisation, breeding divisiveness and eroding social solidarity (see, e.g., A. E. WALDMAN, «The Marketplace of Fake News». *Journal of Constitutional Law*, 20(4), 2017, p. 845-870).

⁵ See, for instance, the study carried by the International Center for Media & the Public Agenda (ICMPA), University of Maryland, USA (<https://theworldunplugged.wordpress.com>); and also: THE WASHINGTON POST, *Generation Z. What it's like to grow up in the age of likes, lols and longing*. Diversion Books, 2016.

⁶ S. ZUBOFF, *The Age of Surveillance Capitalism: The fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.

much more dangerous.⁷ This is a new version of an understanding that human wisdom traditions have had throughout history.

To emphasise the central insight: the great power of computers to process information has bewitched us to the point where we hold the computer up as a mirror to human intelligence. We look at the computer as a creation with a privileged status. In it we seek to understand the intelligence that created it. From the perspective offered here on intelligence, the AI ideal of emulating or even superseding human intelligence is a distracting and misleading goal, simultaneously provoking optimistic fantasies and unjustified fears about the future.⁸

Underlying this belief in an artificial super-intelligence is an understanding (or *mis*understanding) of intelligence in general, and human intelligence in particular, as narrow, abstract, disembodied, individualistic and representational. The fundamental misconception is that it thinks of intelligence as a primal quality, with individuals creating representations of the world and interacting externally between each other and with the environment. Further, this is taken to be a quality which is reproducible by machines that interact between each other and with the environment, using myriads of sensors that differ profoundly from the senses associated with human intelligence.

1.3. Computers do not and need not understand anything

It is important to begin distinguishing the intelligence of AI researchers, engineers and users from the intelligence attributed to the information-processing AI systems that they themselves create and use. Human beings, by designing, creating and using AI systems, exercise—and so improve—all dimensions of human intelligence, even when not explicitly aware of them. (We define and discuss these dimensions—*functional*, *axiological* and *liberating*—later in this article.) Setting ourselves to become fully conscious cultivators of all dimensions of intelligence would have several important effects. It would enhance the intelligence and wellbeing of the researchers themselves as well as adding to the overall effectiveness of the research. It would also add to the quality and to the potential of AI systems to produce humanly beneficial outcomes. In the absence of characteristics that we would describe as *sense-creating and loving engagement*, *communication*, *cooperation* and *freedom*, there can be no authentically

⁷ «Through brain plasticity and changing social norms, we are adapting ourselves to become more knowable for algorithmic machines. In this way we are evolving ... in conjunction with our technical systems, slowly moving toward some consummation of the algorithmic love affair»: E. FINN, 2017 (as n. 3 above).

⁸ M. SCHORLEMMER, «La distracció d'una intel·ligència artificial sobrehumana». *Quadern de les idees, les arts i les lletres*, 217, 2019. Available at <https://www.quaderndelesidees.press/la-distraccio-duna-intelligencia-artificial-sobrehumana/>. Last accessed on February 8, 2021.

creative and beneficial research. AI systems only compute. They are devices programmed to perform or acquire a known skill in order to achieve a given goal.

Intelligence lies in the creators and users of AI systems, not in the systems displaying the skills that have been programmed into them. The expansion of human intelligence is one of the main positive contributions of AI. The danger, of course, is that it might degrade that intelligence. Generally speaking, 'intelligent' machines do not understand anything; they compute complex functions only understood by their programmers and users. However, it is possible to see how confusion arises. Simply through data processing, AI systems perform tasks such as language translation that at first sight seem irreducible to computation. Together with the fact that humans do these tasks through intelligence and understanding, this has led to the metaphoric attribution of intelligence and understanding to computers.⁹ The obvious fact that they do not operate through sense-making and understanding has become blurred by their seemingly magical powers, helped by many people's superficial understanding of their own intelligence. (This ignorance, it is worth noting *en passant*, is one of humanity's major problems. The development of our intelligence, through continuing education, is a political must.) Even scientists of great reputation have endorsed the prediction of future 'super-intelligent' machines, and issued dark warnings about them.¹⁰ Here, once again, are the effects of the split between the sciences and the humanities.

Given the limitations that this lack of a sense of *meaning* builds in to AI systems, we should be wary about leaving important and far-reaching decisions concerning, for instance, finance and employment, in the hands of such machines. (As of 2017, 70% of all financial transactions were performed by algorithms.¹¹) Although this automatic decision-making may empirically be very useful in routine matters, for more complex and challenging issues, where human values and the dynamic of shifting contexts apply, it is simply not up to the job. The inevitable conclusion is that automatic decision-making can only be harmful to social justice.

Another important consequence of AI's rapid development and increasing ability to perform complex tasks is that people are now obliged to develop and exert their crea-

⁹ G. LAKOFF and M. JOHNSON, *Philosophy in the Flesh. The Embodied Mind and its Challenge to Western Thought*. Basic Books, 1999.

¹⁰ See, for instance, S. HAWKING, M. TEGMARK, S. RUSSELL and F. WILCZEK, «Transcending Complacency on Superintelligent Machines». *Huffington Post*, 19 April 2014. This a consequence of adopting computational functionalism and its assumption that intelligence is mere information-processing, which is often taken as an unquestionable fact (see, e.g., the statement by Sam Harris: «Intelligence is a matter of information processing in physical systems. Actually, this is a little bit more than an assumption ... [W]e know that mere matter can give rise to what is called 'general intelligence'»: S. HARRIS, «Can we build AI without losing control over it?». *TEDSummit*, June 2016. Available at https://www.ted.com/talks/sam_harris_can_we_build_ai_without_losing_control_over_it. Last accessed on February 8, 2021).

¹¹ F. HELBIG et al., «Will Democracy Survive Big Data and Artificial Intelligence?». *Scientific American*, 25 February 2017.

tive intelligence as fully as possible, particularly in the workplace. Otherwise, at the level of mere skills, they will be displaced because of the increasing number of tasks now performed quite satisfactorily by ‘intelligent’ machines. (You would hope that writers and translators, in love with language, forever enquiring into meaning and creating new forms of expression, would never be replaced by machines.)

A warning is also called for in connection with the recently developed field of trans-humanism, in which individual brains are enhanced by AI systems implants (or vice versa), attempting to dump a human mind into a computer. These projects arise from the old (and misleading) *individualist* understanding of intelligence. As an idea and as a suggested direction of travel, it only distracts from humanity’s greatest need: to develop and manifest a mature and harmonious socially embodied intelligence, based on universal lifelong education.

2. The Role of Information

2.1. The need to dispel some confusions and ignorance of fundamental principles

Information abounds in modern societies, and an understanding of its central role is vital. In order to avoid its misuse, we need to distinguish between *intelligence*, *knowledge*, *information* and *data*. Unfortunately, these words are all too frequently confused and their meanings profoundly misunderstood.

This confusion is rife in most discussions of AI. Computers, properly speaking, process data in the form of bits, zeros and ones, rather than through knowledge or information, both of which have *meaning*. Meaning is alien to computers; their operations are meaningful only to us, not to computers themselves. So, when we use the widely accepted term *information processing*, we are actually referring to *data processing*.

Our understanding of intelligence can be framed as *cultural* or *humanist*. We consider human intelligence to be a collective phenomenon. In other words, an individual, embodied intelligence depends on the *socially embodied intelligence*. Intelligence belongs to the commons; it grows because it is cultivated and held in common, not owned privately. It is our primary common good, and the source of everything else in our lives. Its essence is *communication* and *symbiosis*—the quality of being able *to live together*. Its growth or decline is dependent mainly on cultural factors.

Our focus is on five distinctly identifiable aspects—or powers—of intelligence: *interest in reality*, *communication*, *subsidiary symbiosis or cooperation*, *research* and *freedom* (this last being the most important of all). How these powers of intelligence are exercised—the degree of intensity and the priority of each power over the others—will determine the different uses of any particular aspect of human intelligence.

For the moment, three specific uses of this intelligence demand our attention. The first two address our material needs as living organisms and our investment in meaning and survival. They constitute the intelligence of need, and they are conditioned by our biology. We call the first *functional intelligence*. This is mainly an instrumental, abstract form of intelligence, the intelligence proper of techno-science—the entanglement of science, technology, economy, and their products and services. By way of this intelligence we humans focus on what intelligence is for: its function for adaptation, for problem-solving, for attaining goals, for success, for survival. The second, we call *axiological intelligence*. In interdependence with functional intelligence, this is the intelligence by way of which we humans imbue meaning and value to what we do. It connects us to the aesthetical dimension of life. Artists, for instance, exert very much this sort of intelligence. Through its creation of values, axiological intelligence responds to our need for meaning and direction in life. Functional and axiological intelligence constitute the intelligence of need, which depends very much on our bodily interactions with other organisms and the environment. It creates lawful models of reality, a stable world relative to our human needs and interests.

Inseparable of the intelligence of need, there is also a third, contemplative dimension to intelligence. This is the dimension that provides us with the insight that functional and axiological models are ultimately relative to our contingent needs as human beings. This is a subtle but powerful form of intelligence that frees us from getting caught within any particular reality as we conceptualise and experience. Its highest manifestation is the silent contact with the origin, of what is unique and free of law: the intelligence of the whole, of the creative freedom and unity of reality, the source of human creative freedom. Because it is centred on freedom, we call it *liberating intelligence*. Through it, the other powers of intelligence can attain their greatest potential. Interest, the basic energy of intelligence, can reach its highest degree—unconditional compassion and love—as the hallmark of full intelligence; communication can become sincere and trusting to the point of silent communion; subsidiary symbiosis becomes a union of love and service; research reaches the highest degree of collective creativity and is undertaken for the good of all humanity. It is important to keep in mind that intelligence is, at its core, as untameable as our freedom and creativity. Human intelligence, we come to see, is primarily an intelligence to be lived and developed rather than defined.¹²

¹² J. AGUSTÍ-CULLELL, «Reflexions sobre intel·ligència humana i intel·ligència artificial: el repte d'una intel·ligència lliure i creativa». *Poblet*, 38, June 2019, p. 30-36. J. AGUSTÍ-CULLELL, «Intel·ligència alliberadora». *Ars Brevis*, 25, 2019, p. 29-66. Available at <https://www.raco.cat/index.php/ArsBrevis/article/view/371411>. Last accessed on February 8, 2021. J. AGUSTÍ-CULLELL, «Una intel·ligència que libere la condició humana». *FronterasCTR*, January 29, 2020. Available at <https://blogs.comillas.edu/FronterasCTR/?p=4819>. Last accessed on February 8, 2021.

2.2. Distinguishing between knowledge, information and data

Knowledge is an essential component of the models of reality created by the intelligence of need: the co-working of functional and axiological intelligence, informed by the creativity of liberating intelligence. Thus, knowledge is not only descriptive (its information content) but fully charged with sense, emotions and values. It originates in a creative act where freedom is operative. Because of the hidden workings of this freedom, in the known always lurks the unknown, the basic awareness and stimulus behind all research.

To distinguish it from knowledge, we characterise information as the form of knowledge which is descriptive, de-contextualised, free of emotion, value-neutral and frequently conceptually abstract. It is most liable to be formally represented and then used via computation. If someone tells you that a person you hardly know has died, that is received as part of the inevitable pattern of birth and death. This news is merely information. However, if this person is your sister, then it becomes knowledge, full of meaning and deeply personal emotion.

Knowledge and information both involve the use of data. The material and conventional signs are meaningless in themselves (every language uses different signs and sounds to convey similar meaning). So, simplifying the complex evolutionary history of data (i.e., the creation of pictures and writing), we can say that data or signs support and transport meaning for us. The articulated sounds of speech, or the pure syntax of texts or pixels in physical images, are the main ways in which data is held and knowledge and information transported between us. Syntax holds semantics.

Data is the object of computing after it has been represented by lists or matrices of zeros and ones, or bits. Once data is digitised, the bits are the objects processed by computers, the atoms of computation being the electronic logic gates—or artificial neurons, in the connectionist architecture of computation that we will look at later in this article. We feed the computer with data that has informational content for us, but not for the computer. The computer processes these data, we interpret the results as new information, and we say that «information has been processed».

Information is an abstract form of knowledge that is specifically characteristic of the techno-sciences. Information is essentially logical in character and deals with the regularities of the world. Its main goal and purpose are the prediction and control of phenomena. Techno-scientific information has proved to be indispensable in helping humanity to deal with its needs and to improve the quality of life. The fundamental mistake, though, is to confuse what is *abstract* (a tool for the satisfaction of life's needs and interests) with the *concrete*—with life itself. This mistake is universally made, to the point of converting the abstract into the concrete, as when matter is taken to be the concrete foundation of the universe, or frequency is made into a colour, or a complex algorithm into an intelligence. It is because of its power to reveal and resist this very

tendency that contemplative practices such as silence, meditation, beholding, or deep listening are so significant for the human understanding of intelligence. These practices help quieting the mind and thus offer a liberating possibility, putting us in direct contact with concrete reality, with its beauty and freedom. Techno-scientific inquiry thus needs to go hand in hand with the sort of inquiry nourished by contemplative practices, in order to sustain the freedom and creativity that enables human intelligence to go beyond the forms into which it shapes reality by way of abstraction and information.¹³

Because of its abstract and de-contextualised nature, information is considered to be universally applicable, an assumption that we should be very wary of making. The case is rather different with knowledge, which is clearly inseparable from the values and the overall context of the culture within which it originates. This misuse of information is particularly dangerous in medical settings where programs make diagnoses using much more information than a doctor is able to deal with. However, they do it without fully knowing the patient in the present moment and—among other things—in ignorance of the possible harmful effects of even a correct diagnosis when informed without care. To avoid doing harm, doctors need to contextualise the diagnoses produced by these programs. The importance of contextualisation can be illustrated with the following personal experience: The first author knew a very apprehensive woman whose wise doctor, with the agreement of the woman's family, decided not to tell her immediately or all at once that she had leukaemia, opting for a gradual, sensitive and respectful communication with her over the years. His much less experienced successor told her the diagnosis right away; from that moment she felt much worse and she died in a matter of months.¹⁴

For us to attain full knowledge of anything, our level of interest in it has to reach its maximum degree—in other words, *love*. In the case of information, its form of interest is best described as *curiosity*. Scientists and engineers are immersed in information, moved mainly by curiosity about how things work. As such, they are liable to forget the abstract character of information, ruling out the possibility of integrating it into full evaluative knowledge.

The untamed curiosity of scientists holds the danger of impelling them into indiscriminate research, reaching after information regardless of the consequences and the

¹³ M. SCHORLEMMER, *Indagació Contemplativa per a l'era de la intel·ligència artificial*. Quaderns 121, Fundació Joan Maragall, 2019.

¹⁴ This is a problem of so called 'shallow medicine,' when emotional connections between patients and doctors break down, combined with a systemic problem that focusses on excessive diagnosis. These issues can worsen if AI-based medicine is not developed and deployed with care, in the context of a rich multi-dimensional understanding of intelligence, in order to foster presence, attentive listening, empathy, and human-to-human bonding, which are difficult to quantify and digitise (see, e.g., E. TOPOL, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, 2019).

possible harmful uses to which it might be put. In AI, this results in the tendency to automate everything that can be automated, despite the terrible consequences of creating potentially monstrous machines.

Computational power has doubled roughly every year and a half, while the cost of computers has halved. If the automobile industry had followed the same trend, you would be able to drive to the sun on a few litres of fuel. Information supported by data and processed by computers has become the most pervasive and useful commodity throughout much of the world. Such is its importance that the time we live in has been designated as the Information Age. More precisely, we are at the beginning of the so-called digital transition, the age of technological hyper-connectivity (i.e., by means of the Internet of Things, the vision of adding sensors to all kind of everyday objects—buildings, domestic appliances, furniture, clothes, vehicles, etc.—and to have all these objects interconnected through the Internet¹⁵). This is a social process through which technologies of information and communication are so widely deployed that they are transforming our daily lives. It is a reflection of the exaggerated importance and role currently awarded to information, much to the detriment of full evaluative knowledge.¹⁶

Using information when the corresponding data has been processed and applied mechanically, without appropriate contextualisation and interpretation by human intelligence, is a source of many possible errors. In general, a clear understanding of the data, their provenance and their characteristics, must be captured, so that others using the data set can understand the potential flaws.

The general paradigm of this information revolution is to reduce everything to information and then to data. The culmination of this is the installing of information as a basic component of the Universe, processed by a supposedly universal algorithm. “It from bit” is the catchphrase that summarises this myth.¹⁷ However, the *bit* is in the machine and the *it* is in human intelligence. The flexible and wide use of the interaction *it-bit* is what brings about the enhancement of human intelligence. Taking information to be primordial is a misleading belief. The true role of information and data is that they are extremely useful *instruments* of intelligence, created by intelligence to extend its power. In truth, the *bit* comes from *it* and, once processed, returns to *it*.

The amount of data we produce doubles every year. Every minute, millions of Google searches and Tweets occur. This is testimony to the enormous usefulness of information technologies. When used with care by a creative intelligence they hugely increase its power. However, they bring with them new risks and threats. These search-

¹⁵ S. GREENGARD, *The Internet of Things*. MIT Press, 2015.

¹⁶ L. FLORIDI (ed.), *The Online Manifesto: Being Human in a Hyperconnected Era*. Springer, 2015.

¹⁷ J. A. WHEELER, «Information, Physics, Quantum: The Search for Links», in: *Proc. 3rd Int. Symp. Foundations of Quantum Mechanics*. Tokyo, 1989, p. 354-368.

es and posts contain information that reveals how we think and feel. Those in control of information technologies can easily control us. Specialised tech companies that you have probably never heard of are tapping vast troves of our personal data, using extraction technologies such as Big Data. These tech companies know far more about their users' future behaviour than they themselves know. These firms sell their scoring services to major businesses and make huge profits as a result. From Big Data comes big money. People are profiled for targeted advertising by online service providers and in political campaigns. This undermines their freedom and so their humanity. We are now exposed to behavioural scrutiny, prediction, control and—eventually—alteration.¹⁸ As a result, our tendency to become programmed intelligences, drowning in information, is progressively reinforced. And so we undergo an atrophying of the intelligence, a degradation of intelligence's truly creative powers.

To avoid being manipulated by AI systems or being overwhelmed by the current flood of information, we urgently need to develop and empower these creative powers of intelligence—above all, its power to undertake research informed by its essential quality of freedom.

3. Different Views on AI

3.1. A bit of history

The idea of thinking artificial beings arises in antiquity in the form of story-telling devices. There has long been a great fascination for imagined automatons. More significantly, the study of mechanical or formal reasoning began with early philosophers and mathematicians. In the 13th century the philosopher, theologian and mystic Ramon Llull attempted to automate reasoning in order to convert Muslims to Christianity. Much later his work was of a great influence on Leibniz's *Mathesis Universalis*, a computational view of the universe, giving rise to his celebrated dictum: «Let us calculate, without further ado, to see who is right». In 1834, Charles Babbage conceived the Analytical Engine, a universal computing device. The study of mathematical logic determined Alan Turing's theory of computation. By shuffling data or signs as simple as 0 and 1, a machine could simulate mathematical deduction. This insight, that digital computers can simulate any process of formal reasoning (or, more precisely, any computable function), is known as the Church-Turing thesis. Beyond this, discoveries in neurobiology, information theory and cybernetics led researchers to the idea of building an electronic brain.

Artificial Intelligence as an academic discipline was born in a workshop at Dartmouth College, New Hampshire in the summer of 1956. The term was coined by John

¹⁸ S. ZUBOFF, 2019 (as n. 6 above).

McCarthy, then a young assistant professor of mathematics. He wanted to distinguish the new field from cybernetics and the influence of one of its leading proponents, Norbert Wiener. Initially research concentrated on the definition of general methods applicable to most problems, whereas contemporary AI research has focused on more specific methods that will be most effective in each particular type of problem.

Although only 65 years old, AI has become a highly significant branch of the techno-sciences and also one of the great scientific objectives, arousing expectations around its enormous benefits, particularly economic ones. It also provokes much speculation about risks and dangers.¹⁹

We can best understand it as an outpost of computer science, but one that benefits from many other disciplines, in particular the cognitive sciences. It has an enormous social and economic impact. AI techniques and systems have become an essential part of the technology industry, helping to solve many challenging problems in computer science, software engineering and operations research. Exploiting concurrent advances in computer power, large amounts of data, and theoretical understanding, it has become one of the most revolutionary of the techno-sciences.²⁰

The initial hypothesis of AI is that human intelligence can in principle be described so precisely that a machine can be made to simulate it. Hence, the aim of AI is to create ever more autonomous and powerful computational systems—a true revolution in techno-science and engineering. However, as we've already established, AI has a narrow view of intelligence—one that is useful to create AI systems, but in no way comparable to human intelligence.

Within this general aim, there are two main classical approaches to AI. Firstly, the pragmatic approach called *Weak AI*. It aims at the automation of complex but well delimited tasks that are usually performed by human intelligence, such as playing chess or the recognition of objects in images. Secondly, the ambitious approach called *Strong AI*, which is oriented towards emulating or even surpassing human intelligence (i.e., machines capable of designing other machines). Within Strong AI, there has been a recent specific undertaking, with the aim of defining and measuring intelligence in general, and that of machines in particular. It proposes the testing of machine behaviour in respect to what is called Artificial General Intelligence (AGI). AGI is the foreseen ability of computational systems to learn or acquire skills and adapt to a wide range of environments, applying what is learned from one domain to another.

¹⁹ «El objetivo último de la inteligencia artificial —lograr que una máquina tenga una inteligencia de tipo general similar a la humana— es de los más ambiciosos que se ha planteado la ciencia. Por su dificultad, es comparable a otros grandes objetivos científicos como explicar el origen de la vida, el del universo o conocer la estructura de la materia». R. LÓPEZ DE MÁNTARAS and P. MESEGUER, *Inteligencia Artificial*. Colección ¿Qué sabemos de?, Editorial CSIC y Libros de la Catarata, 2017.

²⁰ K. SCHWAB, *The Fourth Industrial Revolution*. Crown Business, 2016.

3.2. Weak AI

Weak AI, an engineering and pragmatic view of AI, uses computers to automate complex tasks that, when done by humans, we describe as intelligent. Weak AI does not start from the supposition or hypothesis that all intelligence is computation, that is, information processing. It is essentially a more advanced informatics, extending the boundaries of automating more and increasingly complex tasks.

The Weak AI approach closely follows the techno-scientific method of abstraction. Not subject to the need to reduce reality to information, it leaves aside what is not solvable by information processing. As a result, it focuses research on the appropriate techniques to solve complex problems. Algorithms solve a fully described task in a fully described environment where all possible inputs can be explicitly enumerated or analytically defined. The goal is to accomplish the task efficiently, inspired by what humans do (and how), but not tied to it; the main intention is to make advances in computational techniques through exploiting the computational power and data available at each moment.

One of the most popular basic models is that of *autonomous agents* or *multi-agent systems*. Leading AI textbooks define the field of AI as the study of ‘intelligent autonomous agents.’ It refers to any device that perceives its environment, interacts with other agents and takes autonomous and rational actions that maximise its chance of successfully achieving its goals. Rational actions imply the ability to choose at every moment the action that will produce the best result following a pre-established performance measure.

However, Weak AI systems, although very efficient at doing the job for which they are designed, fail when confronted with activities that differ even slightly from those original tasks. This is one of the reasons why this approach to AI is described as ‘weak’ in terms of intelligence. IBM’s champion chess-playing system Deep Blue, for instance, is incapable of playing checkers. The researchers who created it realised that the program, which is based on the AI methods of *minimax* and *tree search*, tells us nothing about human intelligence. The ability to play chess does not demand any specifically human abilities; the problems involved can be solved by techniques independent of human cognition. The same is true of many other human skills, as AI proves every day. In the match between Kasparov and Deep Blue, we can say that only Kasparov was really playing chess, with all its cultural connotations. Deep Blue was just computing. Deep Blue did not experience victory; but Kasparov felt the defeat.

Weak AI embraces most of the existing applications of AI in daily life. It has been deployed in a range of contexts and social domains, with mixed outcomes: insurance, finance, education, employment, marketing, governance, security, policing, etc. However, since the very beginning, many AI systems have failed to be fully reliable. They make errors when they are applied to tasks that have not been foreseen in every detail.

It's for this reason that research on robust AI systems has become increasingly important. The main aim of Weak AI is to create systems that are sufficiently safe and robust for us to trust them in all their applications. In particular, robustness and flexibility are perceived as important requirements for certain broader sub-fields of AI, such as self-driving vehicles, domestic robotics, or personal assistants.

3.3. Strong AI

Strong AI views intelligence as a program that works with the hardware of a brain. It believes in building machines that can outperform not only our muscles, but our minds as well. In other words, the assumption is that the best way to understand the mind is to create one on a computer, using the knowledge we have of the human mind. To Strong AI, the essence of intelligence is the processing of information or data. From this premise, it makes sense to assume that human intelligence efficiency is not an upper limit. Hence, Strong AI goes for systems with a much broader scope of application than Weak AI—systems with domains of application beyond human intelligence. This, however, does not make them intelligent or super-intelligent. Intelligence and machines should not be compared. Each has its own domain of efficiency. The wise attitude is to aim for cooperation, rather than creating AI systems that are fully independent or self-governing.

Despite its failure up to this point, the ideal of Strong AI still seduces a great part of the research community. It seems that the idea of emulating human intelligence is a powerful incentive for the advancement of AI. Despite the failures, its proponents are determined to carry on.²¹ The danger remains that what is being promoted is a narrow view of human intelligence. One that deprives it of its central and most important feature: its freedom and full creative power. The outcome is a continued degrading of human intelligence, the opposite of what wisdom-based AI research should have as its goal.

The project of emulating human intelligence is frequently understood as mimicking the supposedly independent and self-governing intelligence of individuals. They are thought of as possessors of their own intelligence, relating and interacting externally with each other and with the environment; and reproducible by computational systems, processing and exchanging data captured from the environment through sensors. This, though, is a fundamental misconception. The embodied autonomy of an individual's intelligence is actually in a permanent process of being constituted, through sense-making interactions between individuals and with the environment. The human intelligence of individuals depends on a socially embodied intelligence. A collective or

²¹ V. C. MÜLLER and N. BOSTROM, «Future Progress in Artificial Intelligence: A Survey of Expert Opinion», in: V. C. MÜLLER (ed.), *Fundamental Issues of Artificial Intelligence*. Synthese Library, Vol. 376. Springer, 2016, p. 555-572.

cultural intelligence that embraces all the interactions of social life, through the participatory sense-making powers of the creative intelligence. From this perspective, it starts to become clear that the project of emulating socially embodied human intelligence is, at the very least, a far-away goal. And probably a misleading and dangerous one that threatens to degrade human intelligence.

3.4. Speculations on AI

So-called *Speculative AI* anticipates a future explosion of electronic or silicon-based super-intelligences with the potential to simulate the characteristics of a human AI developer, thus having the capacity to reproduce themselves. The far-fetched goal is the creation of an AGI which will colonise the universe. According to this fantasy, ‘intelligent’ machines will be our successors, the new inhabitants of Earth when the evolution of the sun has rendered life on this planet impossible.

There are countless debates and predictions concerning the future, the capabilities and social impact of AI and the myriad products and services it offers. Arguments also abound about the philosophy and the ethics of AI. Throughout history the issues provoked by the idea of *autonomous intelligent machines* have been addressed not only through myth and fiction but also by philosophers. This level of discussion and scrutiny is now intensifying. Speculations proliferate about the possibilities and impact of so-called *super-intelligence*—pictured by some as a promising challenge, and as a terrible threat by others. The latter consider AI to be a danger to humanity if it progresses uncontrolled. They also believe that the technological revolution brought about by AI will risk mass unemployment for the first time in history.

Based on the view of human intelligence that has been presented in this article, we suggest that what is urgently needed is a full awakening of our *liberating intelligence*—the silent, contemplative dimension, fundamental to human freedom. This is the only way that we can avoid submission to algorithmic guidance by supposedly ‘intelligent’ machines on matters ranging from love to real estate. Or submission in the form of alienated work, in which humans become mere processors handling simple tasks assigned by an algorithmic apparatus. The overriding imperative is for us to break away from the dominance of an economic rationale, the force that drives all current AI development.²²

3.5. Use and abuse of metaphors

Metaphors are constitutive of human language and so of human understanding. They have inspired much of human research—that of AI in particular. In general, AI

²² M. SCHORLEMMER, 2019 (as n. 8 above).

has quite rightly taken its inspiration from human skills when designing its systems, using metaphors to describe them, even attempting itself to simulate metaphorical language. However, there are also abuses of metaphor. Attributing human qualities and powers such as experience, knowledge and intelligence to machines, stretches the metaphor to an impossible extent. This is to talk about machines as if they were human, as though they were experiencing, thinking and responding intelligently to the environment just as humans do. The outcome is much confusion and many misleading projects, such as artificial general super-intelligence and its ambition of going far beyond the human.

AI specialists in machine learning have trained an artificial neural network to detect tumours after radiography. The computers are able to undertake the task more efficiently than humans. The AI researchers say—metaphorically—that the artificial neural network learned by itself to recognise tumours. Whereas what has actually happened is that they have exploited the power of artificial neural networks to identify statistical regularities or patterns in huge amounts of data. In other words, the network was structured to approximate a computational function, to detect a data pattern that humans judge to be a tumour. The machine does not recognise tumours, it detects patterns in data; it is only a *tool*, the use of which is the responsibility of the doctor.

Given that only humans can make sense of what the artefact is doing, it follows that we are also responsible for making sure the artefact is doing something that makes sense and is beneficial to humans. Machines cannot do it themselves. Humans exercise their functional intelligence with the aid of these useful machines, but nothing is gained by metaphorically ascribing human-like intelligence to them. Metaphors provide a basic way to make sense of the world, but we need to be aware of their limits.

4. Ignoring Liberating Intelligence

Amidst current attempts to understand and characterise intelligence, the humbling, foundational experiences that liberating intelligence can open up for us are widely ignored. The origin of this ignorance lies in the confusion between models and reality, between the abstract and the concrete. This is a fundamental confusion with far-reaching consequences. In particular, it gives rise to purely materialist models of intelligence. This is to misunderstand intelligence as if it were a derived phenomenon emerging from abstract concepts such as matter, which is mistakenly believed to be primary and concrete.²³

²³ M. TEGMARK, *LIFE 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf, 2017.

This ignorance with respect to liberating intelligence is at the root of human presumptuousness and of many misconceptions about models of intelligence such as AI. Presumptuousness combined with violence renders *Homo sapiens* a species without a future.²⁴ In order to have one, it must transform itself into the peaceful and humble *Homo quaerens*—the inquiring human.²⁵ Presumptuousness is implicit in the declaration made by the founders of Artificial Intelligence itself:

«We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College ... An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer».²⁶

This is evident ignorance of the primary and indefinable character of creative intelligence. In particular, they seem unaware that intelligence lies in sense-making interactions with nature, between humans, and with tools, instruments or devices. It is not—as is implicit in their rationale—a primal *possession* of individual entities, be they humans or machines.

This ignorance and presumptuousness go hand-in-hand with a world-view based on individualism and domination. It manifests in many different ways. Mirroring terminology used in software development, we can refer to Presumptuousness 1.0 as the primitive rudeness of the animal, imposing itself through sheer brute force. Presumptuousness 2.0 is that of a violent *Homo sapiens*, the one who *knows* as a way of dominating nature and other beings. Presumptuousness 3.0 would be the presumptuousness of ‘singularity,’ the fantasy that machine ‘super-sapiens’ will surpass human intelligence, start to reproduce and go on to colonise the Universe. A great deal of misunderstanding and hubris is folded into such speculations. Techno-sciences focus their attention on what is regular, predictable, quantifiable or computable. They rightly excuse themselves (or should do) from the realms of axiological and liberating intelligence, along with other complex phenomena of social life. The presumptuousness comes from the implausible pretence (and simplification), that technology—no matter how useful—can be used to explain the whole of reality and to artificially emulate its free creative agency: intelligence.

²⁴ One could certainly qualify this statement as presumptuous itself (from Latin *praesumptiosus*, i.e., full of boldness); but we hope the reader will indulge our confessed presumptuousness of this and the following paragraphs.

²⁵ See also <https://www.HomoQuaerens.info>.

²⁶ J. McCARTHY, M. MINSKY, N. ROCHESTER and C.E. SHANNON, «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence», 31 August 1955. Available at <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>. Last accessed on February 8, 2021.

5. Two Main AI Models of Intelligence

5.1. The representational model

In the representational view of cognition, cognitive models of reality belong to individual organisms. The process indicated by these models is understood to happen as follows: organisms receive perceptual information from the environment; based on this information they dynamically build and update internal cognitive models of the external world. Then, using these models, they reason, make decisions and act in response to the sort of entities there are in the external world, what their properties are, how those entities relate to one another and how they can be manipulated. Despite their incompleteness and possible inaccuracy, AI cognitive scientists consider these models to determine the way in which an organism views the world and acts in it. The representational cycle of an organism is conceived in this way: *perception - update models - make decisions - act*. In this view, an organism's viability depends on the efficiency of its representational cognitive models. This is a good and useful model of cognition for implementation in a machine using sensors to gather data from the environment.²⁷

However, representational cognitive models are obliged—through reduction or simulation—to turn every kind of cognition into representational form. Consequently, in their attempt to be comprehensive, they become hard work. This is due to the inherent complexity involved in representing certain types of contextual intra-active cognition through formal languages. The models contain inaccuracies and are always incomplete. Formal representations of reality cannot be otherwise.

Beyond this fundamental problem there is the practical question about the type of knowledge that we humans will actually be able to model, formalise and program so that it will be finally computable in a machine (bearing in mind that not even all mathematical knowledge is computable, as Gödel's theorem reminds us). A case in point is the difficulty AI has in simulating common-sense human behaviour. This difficulty surely results mainly from the fact that human intelligence deals with contextual meaning and computers do not. It is for this reason that behaviour based on common sense resists representation as knowledge in machine-interpretable form. Any attempt at such representation cannot match the continual increase generated by the sense-creating activity of human intelligence. No wonder that the problem of relating abstract knowledge-representations in the machine to specific real-world situations, at scale and in an efficient manner, has still not been solved. Nor has the challenge of transferring what has been modelled in one context over to another. A further indication that

²⁷ G. MARCUS, «The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence». *arXiv:2002.06177 [cs.AI]*, 14 February 2020. Available at <https://arxiv.org/abs/2002.06177>. Last accessed on February 8, 2021.

the role of AI is not to emulate human intelligence, but to free it from tasks which are eminently performable by computable functions.

Using these languages of logic, AI researchers and engineers create detailed representational and computational models of the external world. Through them computers perform functions corresponding to different types of human skills: handling logistics, theorem-proving, playing games such as chess or Go. Generally, these computable models are used to automate jobs which can be reduced to a complete, mathematical, formal and rational description. We also see them being applied in industrial robotics, the control of home appliances, investments in the stock market and so on.

This AI representational model is in sharp contrast with the *enactive* models of intelligence that cognitive science offers. In these, sense-creating cognition is understood to arise through a process of self-individuating interactions between a living organism and its environment. In other words, in the enactive model, organisms do not create internal representations of the world after passively receiving information from their environments. The intelligence of the life-form is not accessing its world in order to build accurate pictures of it. Rather, in each organism, intelligence is actively participating in the generation of meaning in connection with what matters to the organism, *which lives in the world it creates*.²⁸

In the case of human intelligence, language is a special kind of social agent. It emerges from the intra-dependence of creative intelligence's powers, most especially those of symbiosis and research. Human language makes us conscious of our freedom, of the liberating intelligence within and between us which is our means of immediate access to reality.²⁹

The contrast with representational models, which rely on mere sensors capturing environmental data, is clear. Passive sensors and their signals are not enough to make self-individuating, embodied intelligences out of machines. The same can be said about connectionist models in which training and learning is based on processing huge amounts of data. Both the symbolic and the connectionist models produce an intelligence with powers radically different from that of humanity.

5.2. The connectionist model

The connectionist model of cognition conceives of the mind as a *tabula rasa*. The term is borrowed from the philosopher John Locke who used the metaphor of the blank slate to support his view of the mind as a flexible, adaptable, highly general pro-

²⁸ E. DI PAOLO and E. THOMPSON, «The Enactive Approach», in: L. SHAPIRO (ed.), *The Routledge Handbook of Embodied Cognition*, chapter 7. Routledge, 2014.

²⁹ J. AGUSTÍ-CULLELL, 2019 and 2020 (as n. 12 above).

cess, adept at turning experience into behaviour, knowledge, and skills. Later, Alan Turing, one of the great precursors of AI, thought of the child's brain as being something like a notebook—not much mechanism and many blank sheets—thus, transferring his invention of a universal computing machine into the workings of the human brain.

The connectionist model tries to get away from model building through the sheer hard labour of knowledge representation. Because of the implied potential for learning, the metaphor of the 'tabula rasa' has been attractive to designers and proponents of this model. Simply put, connectionist computing systems are loosely inspired by biological neurology. An artificial neural network (ANN) simulates the physiological structure and functioning of animal brain structures. (In particular, the brain of the nematode *Caenorhabditis elegans* has been studied and simulated computationally.) An ANN is based on a collection of connected units or nodes called *artificial neurons*, loosely modelling the neurons in a biological brain. ANNs derive their skills from the huge amounts of appropriately labelled or classified training data used to program the network.

The connectionist model started to be successful when implemented by several levels of hierarchically interconnected ANNs. In such an arrangement—known as *convolutional*—each level performs what is known as an *object detection step*. At the lowest level, the straight lines of a chair might be detected; at the highest level, it is the full detection of the chair, with interaction at all levels between. This method has proved to be immensely powerful for machine learning—so called *deep learning*. From huge amounts of training emails, for instance, which have already been classified as spam or not, an ANN is programmed through a kind of algorithmic (and very long) trial-and-error process which successively upgrades its performance. The result is a computational function that can determine whether new emails are spam or not. An early outstanding success came when an ANN classified 150,000 digital images with an 85% success rate. Before long, this rate had moved up to 98%.

Deep learning has been applied mainly to image processing, games, speech recognition and natural language processing. Its success has made it the dominant cognitive framework in which AI research is taking place. This model is adequate to deal with those situations where there are vast correlative databases—more data than knowledge. However, these processes tend to be excessively data-hungry. They need ever-larger training sets and more and more computer power. 30 million training situations, for instance, is not enough for a deep learning model to learn to drive a car in a plain supervised setting. Moreover, they are brittle because of their lack of memory and reasoning power and so are limited in their ability to adapt to anything new. The main problem is that they are not fully reliable. A slight modification of the pixels in an image—imperceptible to human eyes—can result in the trained ANN making extensive errors of object detection. This is particularly so when real world

circumstances deviate from the training data, as is so often the case. When robustness and security are of the essence, they cannot be relied on. An image classifier trained only with pictures of brown horses and black cats might classify all black patches as cats, or brown ones as horses. Future research will no doubt push back these limitations. But the resulting systems will still not be comparable to human learning based on creative intelligence.³⁰

The limitations of both models—the representational and the connectionist—are well known. Many researchers have proposed ways of combining them, harnessing connectionist power to certain aspects of the representational model to avoid its inherent brittleness. Such a model would, for instance, combine perceptual knowledge of what dolphins look like with a verbal characterisation that they live in water along with other features, thus producing more reasoning power. The AlphaGo programme, designed to play the game Go, merges a dynamically constructed, symbolically represented search tree with a variety of connectionist modules for estimating the value of various positions.

As we tried to explain until now, AI studies functional intelligent behaviour, in particular the accomplishing of goals by means of data processing, leaving aside considerations of value. Given this fundamental understanding, it is clear that AI needs to be governed in ways that will be beneficial for all humanity. A particular aim is to understand the importance of AI, the values and counter-values it generates, and so to contribute to its effective governance. This governance of AI must be part of the democratisation of the techno-sciences. It demands democratic, interactive initiatives within civil society.

6. The Governance of AI

In society's current state of development, algorithms work at the intersection of computing, culture and human intelligence. Their increasing presence in every aspect of life emphasises the need for proper democratic and ethical governance. This governance will become more urgently necessary as the social impact of AI increases. There is already research on the ethical regulation of the design and use of AI systems. It is a complex but necessary task. The challenge is to devise the global norms, policies and institutions which will best ensure the beneficial development and use of advanced AI. Some examples of such proposals suggest the banning of AI armaments, of AI bots in financial decisions and of any invasions of privacy by AI.³¹

³⁰ R. LÓPEZ DE MÁNTARAS, «El traje nuevo de la inteligencia artificial». *Investigación y Ciencia*, July 2020.

³¹ A. DALY et al., «Artificial Intelligence Governance and Ethics: Global Perspectives». The Chinese University of Hong Kong Faculty of Law Research Paper No. 2019-15, 4 July 2019. Available at <https://doi.org/10.2139/ssrn.3414805>. Last accessed on February 8, 2021.

Independent review boards will be increasingly required to assess the ethical validity of AI applications. Although such boards currently assess academic research, AI applications by governments (in the military sector) or private corporations (undertaking subtle invasions of privacy) are unlikely to fall under their oversight. AI research, unlike other techno-sciences, is largely undertaken by those technological companies that currently rank top among the public corporations with greatest market capitalisation. Only ten years ago these top positions were dominated by the oil and gas industry.³²

There is a clear need to significantly expand the purview of independent review boards and government control. However, the ethical analysis of AI is complex because it involves countless interactions—between designers, developers, users, software, and hardware. This dispersed nature of AI implies a shared responsibility for its effects and actions. In addition, any ethical analysis must not be merely general; it needs to take on particular AI technologies and systems such as self-driving cars or automatic decision-making by AI systems which are not fit for purpose.

AI has to face specific governance challenges, some of them extreme. These include labour displacement; inequality; a global market dominated by a small group of large sellers; totalitarian regimes as digital dictatorships with autonomous weapons or AI sensor technology enabling cheap, extensive and effective surveillance; shifts and volatility in civil society; strategic instability; an AI race that sacrifices safety among many other values.

6.1. Proper wording

The ancient wisdom of Confucius about proper wording being the prerequisite of good governance is more apposite than ever. The world's current crisis of democracy is closely related to the misunderstanding of language. Such misunderstanding gives rise to its misuse—it goes from superficiality to fake news (facilitated by information technologies) followed by the corruption of all the other powers that make us human. Trust—another basic condition for good governance—is no longer possible. Trust requires comprehension and comprehension requires the right use of language's creative power.

Proper wording is the first requirement in the governance of AI. Why do we use the word *intelligent* for a machine that is a stranger to both meaning and creative freedom and does not understand anything? As a general principle, the terms *intelligence*, *autonomy*, and *ethics* should not be applied to machines. This simple recognition would clarify and solve many of the issues connected with the governance of AI.

³² List of public corporations by market capitalization. In *Wikipedia*. Retrieved February 8, 2021, from https://en.wikipedia.org/wiki/List_of_public_corporations_by_market_capitalization.

6.2. The inadequacy of rational ethics

There have already been many different attempts to guarantee the beneficial development of AI. Governments have designed strategic plans in anticipation of its growing economic and social impact. In 2019 the EU drew up *Ethics Guidelines for Trustworthy AI*. The Institute of Electronics and Electrical Engineers (IEEE) have produced its *Global Initiative on Ethics of Autonomous and Intelligent Systems* to advance public debate about the values and principles underpinning the uses of AI. There are several declarations of principles; they include the *Asilomar AI Principles* (2017), the *Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe* (2017), and the *Montreal Declaration for the Responsible Development of Artificial Intelligence* (2018). There is also an open letter on *Research Priorities for Robust and Beneficial Artificial Intelligence* that emerged from a 2015 conference.

These debates are mainly shaped by developed nations and dominated by economic rationales. This raises obvious concerns about the neglect of other countries and their contributions; local knowledge is dismissed; cultural pluralism and demands for global fairness are overlooked. Moreover, in these documents, values are treated mainly at a purely *rational* level—as concepts—rather than as collective feelings. They are addressed more to people's heads than to their hearts. But if we consider, for instance, the value of privacy, this can only be effectively addressed by being aware of its counter-value: the surveillance to which we are all constantly subjected. Approached in this way, we are aware of what it is that we find repellent about it, and so have a clear sense of the value of privacy.

Ethical principles, established conceptual values, regulations and codes of conduct are not in themselves enough. We need to develop our creative axiological intelligence which in turn must be rooted in an awakened liberating intelligence. Rational ethical considerations are a part of axiological intelligence, but by themselves they are incapable of establishing values which will motivate and guide human activity. This impotence of modern rational ethics is what lies behind the crisis of values in modern societies.

6.3. Beneficial AI

Much current AI research is underpinned by the old ideal of designing computational systems to be autonomous agents that will adapt and self-improve until they become essentially autarchic, i.e., independent and self-governing. These computational systems are thought of as individuals in a society which in turn is conceived as being merely a set of individuals. This view has given rise to questions about ways to ensure these systems will benefit humanity. Such questions even extend to the apportioning of benefits and costs between machines and humans with conflicting desires.

Declarations have emerged: «Machines are beneficial to the extent that their actions can be expected to achieve our objectives».³³

Working with the assumption that near-autarchic ‘intelligent’ machines are coming, different proposals have been made to ensure a positive outcome for humankind. The general thrust of these is that such machines should comply with the understanding of values as outlined above. Some essential conditions have been advanced: machines should be understandable; they should have transparent purposes; they should ask permission before carrying out any potentially dangerous actions; they should accept inspection and correction by humans, with the potential to be switched off when necessary. It has also been stressed that we should prevent machines from interfering with parts of the world the real value and interests of which they are ignorant. In addition, it is proposed that machines should be capable of learning more about our true, underlying preferences for how the future should unfold.³⁴

Based on the metaphoric attribution of reasoning power to AI systems, there has been an attempt to create beneficial AI systems by making the machines themselves ethical—empowered to take ethical decisions without human intervention.³⁵ They would be able, for instance, to operate in this way in the field of automatic hiring.³⁶ Even more ambitious is the attempt to ensure that autonomous vehicles, and systems in other safety-critical contexts, would make the ethically preferable choice.³⁷

However, if human rational ethics are often inadequate, this will be even more the case with those produced by ‘rational’ machines. Generally speaking, ‘ethical’ machines *could* functionally simulate some given ethical behaviours. However, they lack the sensibility of *life* that comes with axiological intelligence. ‘Ethical’ machines only address a fixed set of values; they cannot deal with the extensive background field of values that must be considered in making any ethical decision. ‘Ethical’ machines, making independent decisions, threaten to erode human autonomy.

The priority within AI should always be the design of beneficial machines rather than ones that are supposedly intelligent, independent and self-governing. We have a name for such machines: *tools*. Their essential characteristic is that they are aligned with human values. If we design our AI systems as tools or as services which automate specific tasks—always remaining open to our intervention in order to direct and im-

³³ S. RUSSELL, *Human Compatible: AI and the Problem of Control*. Penguin Books, 2019.

³⁴ *Ibid.*

³⁵ M. ANDERSON and S. L. ANDERSON, «Machine Ethics: Creating an Ethical Intelligent Agent». *AI Magazine* 28(4), 2007, p. 15-26.

³⁶ P. TAMBE, P. CAPPELLI and V. YAKUBOVICH, «Artificial Intelligence in Human Resources Management: Challenges and a Path Forward». *California Management Review*, First Published August 2, 2019.

³⁷ J. C. GERDES and S. M. THORNTON, «Implementable Ethics for Autonomous Vehicles», in: M. MAURER et al. (eds.), *Autonomous Driving. Technical, Legal and Social Aspects*. Springer, 2015.

prove their activity—there is no room, and no need, for autarchic machines and the problems they bring.³⁸

6.4. Autonomous AI systems

Within the framework of enactive cognitive science, the autonomy of an organism is understood to be based on a network of intra-dependent, recursively enabling processes through which the organism constitutes and sustains itself under precarious conditions. This precariousness requires it to interact with the world. These interactions have intrinsic value (positive, neutral, or negative) for the organism itself and for the continuation of its own autonomy, its freedom. This is a constitutive autonomy that living systems enjoy by virtue of their self-individuation. We can understand autonomy as a basic feature of life's sense-making intelligence.³⁹ For humans, precariousness demands not only interactions with the environment in order to maintain autonomy, but also continuous social-learning interactions, based on the exercise of creative intelligence. Human autonomy depends on participatory sense-creating, through language and the background field of values, out of which responsibility is central.⁴⁰

In contrast with the autonomy of an organism, there is a limited version, as seen in intelligent autonomous agents. Their type of interaction with the environment cannot be described as a sense-making activity that is constitutive; therefore, the machine cannot be said to have actual autonomy, much less any kind of responsibility. Humans should always be the only actors ultimately responsible for technological artefacts. It is far from clear how we could align our goals with those of 'autonomous' machines by getting them to learn, adopt and retain those goals. Their very design by itself puts many obstacles in the way of their potential to be of benefit to humanity. The prospect of 'intelligent' machines with their own goals of self-preservation, self-understanding and self-replication, threatens to create monsters which it would be impossible for humanity to live with.

³⁸ The *Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe*, for instance, calls for 'constrained autonomy' («to have clear rules constraining the behavior of autonomous AI systems») and for not neglecting the importance of the 'human role' («All AI systems critically depend on human intelligence. ... and often real benefit comes from the synergy between human and artificial intelligence»: L. STEELS and R. LÓPEZ DE MANTARAS, «The Barcelona declaration for the proper development and usage of artificial intelligence in Europe». *AI Communications*, 3(6), 2018, p. 485-494.

³⁹ M. VILLALOBOS and D. WARD, «Living Systems: Autonomy, Autopoiesis and Enaction». *Philosophy & Technology*, 28, 2015, p. 225-239.

⁴⁰ H. DE JAEGER and E. DI PAOLO, «Participatory sense-making». *Phenomenology and the Cognitive Sciences*, 6, 2007, p. 485-507.

7. Conclusion – Towards Cooperative AI systems

We should abandon the fixation with producing autonomous or autarchic computational systems that are made to emulate or even rival humans. Instead we should address the challenge of creating tools and services that will increase the extent of human intelligence. This is the kind of AI we need in order to handle the complexity of our current world.

In designing AI, attention should be focused on its interaction with human intelligence in its social embodiment. Research in human intelligence and in AI should not only collaborate but also be alongside each other, with interdisciplinary teams in the same institutions. In such a cooperative model each partner is doing what it does best, thus avoiding the metaphorical confusions that abound in any discussion of AI.

The great benefit of AI systems is that they can free us from tedious tasks involving what we already know, liberating us to be creative—the specific power of human intelligence. The question now is how AI will help to develop each of the powers of this creative intelligence, particularly in creating *subsidiary symbiosis*, the kind of non-hierarchical ways of cooperation that will be key in any future creative democracies, i.e., democracies that are organised following the principle of subsidiarity, preventing the concentration of power and control, and in which creativity and freedom is nourished in a generalised way and at all levels.⁴¹

Humans and AI can actively enhance each other. For humans, leadership, teamwork, creativity, social skills and a sense of humour all come naturally. For machines, speed, scalability and analysing terabytes of data are all straightforward undertakings. An AI system might handle thousands of designs matching the designers' specifications. They then choose what they like or dislike, leading to a new round of designs. Axiological intelligence—their creativity, professional judgment and aesthetic sensibility—can thus be deployed.⁴²

In the near future, many activities will be redesigned to support the partnership between human intelligence and AI. New jobs will appear to ensure that AI systems work properly and safely; data officers will guarantee that data feeding AI systems complies with consumer-protection regulations. Cooperative AI systems bring maximum transparency—the best way to safeguard the values described earlier. They reduce the prospect of a future AGI more powerful than human intelligence.

This cooperation requires a particular education of human intelligence, axiological as well as technical. It is the user rather than the designer who manages software coupling, who creates and communicates meaning. As is the case in human-computer

⁴¹ J. AGUSTÍ, «Democràcies creatives». *El Punt Avui*, January 27, 2018. Available at <http://www.elpuntavui.cat/opinio/article/8-articles/1328420-democracies-creatives.html>. Last accessed on February 8, 2021.

⁴² S. L. EPSTEIN, «Wanted: Collaborative intelligence». *Artificial Intelligence* 221, 2015, p. 36-45.

interface research, AI should put psychological and sociological insights at the heart of the design process in order to create a better fit with everyday human activity, understanding and interaction.

The governance of AI must be a part of the democratisation of the techno-sciences. It demands democratic, interactive initiatives within civil society along with increased attention paid to the distribution of whatever benefits (and they are predicted to be huge) that might arise. All of this is dependent on the development of a mature, harmonious and socially embodied intelligence within the majority of the population. This is the necessary foundation of a creative democracy, the only form of society that can ensure a future for humankind.

Jaume AGUSTÍ-CULLELL
Cofundador de l'Institut d'Investigació en Intel·ligència Artificial, IIIA-CSIC
jaumeagusti@gmail.com

Marco SCHORLEMMER
Institut d'Investigació en Intel·ligència Artificial, IIIA-CSIC
marco@iia.csic.es

Article rebut: 16 d'octubre de 2020. Article acceptat: 1 de febrer de 2021

Yuval Noah Harari,
21 lecciones para el siglo XXI.
Barcelona: Debate, 2019, 399 pàg.

El pensador Y. N. Harari és un *best-seller* del temps actual (segle XXI). A aquest segle hi dedica la seva darrera obra, titulada *21 lecciones para el siglo XXI* (Ed. Debate, 399 pàgines). El pensador israelià viu –i pensa– preocupat per les possibilitats del futur i consagra la seva missió intel·lectual a preparar, des de la perspectiva antropològica («la veritat de la humanitat», pàg. 235), els nous temps que ja s'albiren. Com a fill de la Il·lustració europea, i marcat profundament per l'experiència del sofriment, Harari vol projectar el futur amb el compromís del present i la crítica lúcida del passat. El nostre autor, que és alhora un bon coneixedor de la història medieval (sens dubte una època clau per entendre els inicis de la modernitat i el capgirament de la història), domina els *tempos*, de tal manera que és capaç, per mitjà de les presents lliçons, de motivar el ciutadà contemporani a lluitar per la realitat. Les armes pacífiques d'aquesta lluita intel·lectual i política rauen, com a bon il·lustrat, en la recerca apassionada de la veritat, i sobretot, d'una veritat fonamentada en les evidències.

La ment de Harari sembla trobar-se en un estat de neguit existencial i alhora de pau intel·lectual. Com si fóssim capaços d'aturar la barbàrie i preparar uns

nous temps a l'alçada de l'ésser humà. L'autor es mou entre la desesperació (profundes amenaces dels temps venidors) i l'esperança (confiança que l'ésser humà se'n sortirà, d'aquests desafiaments, tot trobant una via ètica comuna i global) (pàg. 179). L'autor parteix dels *big-data* que ofereix el panorama del món i de la història. Sorpren el seu mètode, molt enganxat a la realitat concreta del món present, aportant exemples i dades significatives per poder interpretar amb fidelitat els esdeveniments, i d'aquesta manera poder esdevenir un inspirador profètic del futur de la humanitat. La seva manera incisiva d'escriure, les seves constatacions provocadores i la seva facilitat per divulgar d'una manera clara i distinta grans idees de fons, fan molt interessant entretenir-se en la lectura d'aquest llibre. I en la seva aportació intel·lectual no menysprea ningú, tot i que sí que exerceix la raó crítica davant de l'injustificable o dels intents espuris de justificar la veritat.

Quina és la pretensió del nostre historiador i antropòleg? Cap a on es mou? Què aporta a través d'aquestes 21 lliçons tan actuals i incisives? Harari, davant dels desafiaments de caire tecnològic i polític del temps present, cerca una resposta –una gran recerca espiritual, pàg. 35– que ha d'estar guiada per aquelles actituds bàsiques de la humilitat (sobretot davant de Déu, no prenent el nom de Déu en va) i de la ignorància

(seguint el ja conegut principi socràtic de ser molt conscients del poc que sabem). Resulta decisiu esclarir qui és l'home («una idea clara sobre el sentit de la vida», pàg. 16), per així aprendre com hem de fer servir la revolució de la *biotecnologia* i la *infotecnologia*. D'aquesta manera, el futur que s'anticipa, marcat ineludiblement per les conquestes de la IA (intel·ligència artificial) i del projecte dels *super-humans* (simbiosi d'home i màquina), exigeix una idea de l'home que preservi el món de l'anihilament d'aquella consciència («el major misteri de l'univers», pàg. 343) que, segons l'autor, es diferencia de la ment (cervell) (pàg. 341). Justament aquesta consciència amb vocació ètica és la que ha de permetre a l'ésser humà generar un nou món on convisquin la màquina i l'esperit, sense «posar en perill el seu equilibri mental» (pàg. 55). Així, en el rerefons de les anàlisis socials, culturals i històriques, s'hi troba una significativa preocupació ètica de l'autor envers la humanitat. Doncs, què seria d'un món sense homes? L'home és justament el garant humanitari del món. I encara som en el temps de «la possibilitat d'escollir» (pàg. 346).

Harari divideix les 21 lliçons en cinc grans blocs temàtics, ben lligats l'un amb l'altre mitjançant un fil que els entrellaça. Els dos primers blocs obren el text a tall d'introducció tocant la qüestió dels desafiaments actuals, tant els tecnològics (decepció, treball, llibertat, igualtat) com els polítics (comunitat, civilització, nacionalisme, religió, immigració). Així com advoca pel sentit col·lectiu dels indivi-

us, també assenyalava les ambigüitats perniciososes del nacionalisme i de les religions. En canvi, la immigració suposa el gran repte del moviment global, que ha de trobar la seva mesura adequada i respectuosa. El tercer bloc té com a títol: desesperació i esperança (terrorisme, guerra, humilitat, Déu, laïcisme). L'autor aborda aquelles qüestions que considera delicades i importants per obrir-se pas entre obstacles i oportunitats. El penúltim bloc resultarà decisiu: veritat (ignorància, justícia, postveritat, ciència-ficció). L'aposta per la veritat exigeix persistència, creativitat i preparació per als nous escenaris. Finalment arribem al cinquè bloc: resiliència (educació, significat, meditació). Harari, en aquesta darrera part, tracta de mostrar un camí concret per recórrer l'aventura humana de la veritat. Al llarg de totes aquestes pàgines àgils i fecundes, destaquem una triple noció que resultarà clau en l'univers mental de l'autor, i que suposarà una provocació per a la filosofia antropològica i, en darrer terme, per a la raó metafísica. Com concebre la *llibertat*? Quina *ètica* es proposa? I, finalment, quina relació cal tenir envers la *veritat*?

L'autor assenyalava que la modernitat ha estat marcada pel signe eminent de la *llibertat*, per la lluita revolucionària per la llibertat individual, així com pel sorgiment del sistema polític de la democràcia liberal («l'epifania liberal», pàg. 323). Ara bé, què en restarà, de la llibertat, quan avanci la revolució tecnològica, que podria ser capaç de crear la corporalitat humana, el cervell humà i l'organisme animal (pàg. 69)? Quan la

creació estarà en mans de la llibertat mateixa, no suposarà el pas definitiu a un món on tot serà creat de tal manera que el poder sobre les vides serà absolutament imperiós? Quin lloc ocuparà la llibertat en aquest món sense ambigüitats, sense ombres ni contrastos, on tot serà llum i l'home s'haurà convertit en creador de tot i vencedor de qualsevol principi de mal? Llavors, en el nou panorama, la societat lliure de les llibertats individuals obrirà pas a l'ordre mundial d'una llibertat que es coneixerà a si mateixa amb tal lucidesa que deixarà que la *singularitat* de la gran intel·ligència decideixi per nosaltres. El *Sapiens* haurà resolt per sempre el misteri de la pròpia llibertat, que seguirà de manera dòcil i obediència la gran Intel·ligència agent –rememorant Aristòtil– de la humanitat. El *super-humà* serà l'humà que ha salvat la seva pròpia llibertat de la ignorància i de la possibilitat nefasta del mal, perquè suposadament haurà explorat l'origen dels sentiments –ja que «entrarem en contacte amb les profunditats de la nostra tristesa» (pàg. 45)–. Però heus aquí el confí inquietant: Com gestionar aquest nou paradigma antropològic i històric? Com mantenir la resistència de la llibertat enfront del regne llampant de la transparència?

La posició antropològica de Harari és radical. El pensador israelià té una concepció biològica i neuronal de la vida humana («producte de milers de milions de neurones», pàg. 325). Ja no podem continuar parlant d'una llibertat que té el dit del lliure albir, d'una ànima que subjecta l'obra de l'home, o d'un jo que

és el nucli últim de l'ésser humà i que constitueix la persona. «En cap moment trobem una ànima, un autèntic jo ni un lliure albir» (pàg. 277). L'escepticisme kantià en relació amb aquestes instàncies es converteix en negació (*de-construcció*) amb Harari. Per a l'autor, el món dins de la gran ficció i el món exterior a la ficció, són un mateix món –tot evocant les pel·lícules de ciència-ficció de *Matrix* i *El show de Truman*–. També el personatge de la novel·la d'Aldous Huxley *Un món feliç* (1931), John el Salvatge, nostàlgic de la noblesa i l'heroisme, no troba lloc en l'emergent societat del control absolut i originari. Perquè «l'autenticitat és un mite» (pàg. 273). L'autor pren una posició antime tafísica desconfiant d'aquells «mites personals» (pàg. 327, que, al seu parer, ens allunyen de la veritat més profunda sobre nosaltres. «Al final ens adonem que la nostra identitat fonamental és una il·lusió completament creada per les xarxes neurals» (pàg. 274). Per tant, «no existeix un jo autèntic a l'espera de ser alliberat de la carcassa manipuladora» (pàg. 275). El que cal és «escapar de la reduïda definició del jo, que ben bé podria convertir-se en una habilitat de supervivència necessària en el segle XXI» (pàg. 281). Així, Harari proposa una altra veritat 'real' sobre l'ésser humà i les seves capacitats pràctiques. Ara la llibertat ha d'aprendre a sobreviure amb aquesta nova visió de si mateixa i posar-se al servei de la creació d'un món millor. Impacta la visió dura i gèlida de l'autor, i que sembla tenir les arrels en Nietzsche, encara que no ho expliciti –sí

que reconeix el deute amb el budisme—. Aleshores, cal anar més enllà del «drama intern de la creació humana. L'univers no té sentit, i els sentiments humans tampoc no tenen cap sentit. No són part d'un gran relat còsmic: són tan sols vibracions efímeres que apareixen i desapareixen sense propòsit concret» (pàg. 329).

En aquesta línia *de-constructiva*, Harari veu la necessitat d'una *ètica* per tal d'afrontar el futur de la història. La funció de l'ètica és reguladora de la varietat dels sentiments, de la complexitat de la societat i de la profunditat de la consciència. Així, l'autor reivindica el paper d'una ètica global i secular, respectuosa amb les aportacions de les grans religions i savieses de la humanitat, però que es regeix pels principis suprems de la veritat i l'evidència (què és bo?, què és dolent?, què evita el dolor?, què provoca dolor?, què és just?, què és injust?), i alhora es desenvolupa amb la presa de consciència docta i lúcida de la pròpia humilitat i de la palesa ignorància de l'ésser humà. Ell mateix assegura que «jo personalment confiaria més en qui admet la seva ignorància que no pas en els qui proclamen la seva infal·libilitat» (pàg. 238). Així, defensa «que no l'atmoreix la seva ombra, i en principi està disposat a admetre els seus errors i els seus punts cecs» (pàg. 237). Fa elogi de la modèstia intel·lectual i d'un pensament allunyat dels centres de poder —perifèric— (pàg. 246). Segons el pensador israelià, l'ètica ha d'exercir la crítica amb els grans relats i la història de les religions. Així, planteja provocativament al

cristianisme que tant ha influït a Europa i en el món: «S'haurien de plantejar preguntes molt dures: de quina manera, exactament, la seva 'religió de l'amor' va permetre que fos distorsionada de tal manera, i no pas una vegada sinó moltes?» (pàg. 237). I dirà que «la idea monoteïsta va tendir a promoure el fanatisme» (pàg. 216). El codi ètic seglar (pàg. 228) consagra els valors de la veritat, la compassió, la igualtat, la llibertat, el valor i la responsabilitat. Segons l'autor, «la moral... significa 'reduir el sofriment'» (pàg. 224). Així, davant dels dilemes més difícils, val el principi suprem que caldrà buscar «una via mitjana que causi el mínim de mal que sigui possible» (pàg. 229). I la gran pregunta ètica és «com ens podem alliberar del sofriment?» (pàg. 333). Harari planteja una ètica sensible al dolor i que conrea unes actituds espirituals i nobles. Però, capaç de fundar el bé en el món i d'afrontar els temps cibernètics?

Segons l'autor, aquesta consciència ètica, que ha d'acompanyar els processos futurs i imbuir les decisions oportunes, necessita un compromís amb l'estranya *veritat*, davant de la reiterada era de la *postveritat* (pàg. 256), i preparant els esdeveniments amb la responsabilitat de frenar el mal i el sofriment humans. Aquí Harari sembla donar la mà a Michel Foucault, que advocava per una *parresia* envers la veritat. Cal tenir el valor de construir i pensar camins que no causin patiment a la humanitat, que transcorrin sota la llum de l'ètica, i que defensin la pervivència del *Sapiens* en la història que ell mateix ha estat capaç de

forjar. Així, l'autor també convida a viure l'actitud de la *resiliència* incorruptible i educadora, que és com una manera de romandre en el pols de la veritat. «Com viure en una època de desconcert on els relats antics s'han esmicolat i encara no ha sorgit un nou relat que els desbanqui?» (pàg. 283). Es tracta de tornar a «conèixer la veritat del món, el significat de la vida i de la seva pròpia identitat» (pàg. 335). L'obra fuig del relats mítics i grandiloqüents «—ja que la resposta no és un relat—» (pàg. 335), a favor de la tossuda realitat que sempre roman, tenint en compte que «la cosa més real en el món és el sofriment» (pàg. 333). Quines respostes alternatives més pragmàtiques ens proposa l'autor? (cal trobar-ho en el capítol sobre el *significat*). Però, és veraç una crítica tan demolidora dels mites i relats? En tot cas, Harari invoca grans qüestions que són un *challenge* per a la filosofia d'avui i de sempre, i especialment per a la metafísica: *Què és la llibertat? Quins són els orígens de la llibertat? No és l'ésser humà un misteri del qual brolla un corrent vital? Què és el que marca el futur de la història?* L'autor està convençut que la temptativa de resposta només pot començar *des del sofriment* (pàg. 335). Aquesta obra aconsegueix sacsejar la filosofia amb l'auguri que només podrà sobreviure si realment és un pensament que afronta els *orígens* del drama de la història. En tot cas, el debat decisiu sobre la *veritat global* (pàg. 255) està servit.

Ignasi FUSTER CAMP
Facultat de Filosofia (URL)

**Joan Vergés Gifra (ed.),
*Balmes, quin llegat, avui?***

Barcelona: Documenta Universitaria,
2018, 434 pàg.

Jo soc d'aquells xiquets que jugaven al carrer. Un d'aquells carrers s'anomenava Balmes i era al costat d'un barranc. Com que la retolació, en aquells temps, era en castellà, fèiem una inversió de l'ordre de les paraules de la plaça del carrer i cofoïament solíem exclamar: «Calle Balmes no és el mateix que val més que calles». Quant de *sentit comú* ja no mostraven aquells xiquets! El que jo no sabia en aquells temps era que Balmes, com a intel·lectual, era un dels cinc noms que més carrers de Catalunya batejava.

El llibre que ens disposem a ressenyar vol esbrinar si Balmes és una fita que encara serveix per orientar-nos en el nostre paisatge intel·lectual o si simplement és un senyal que ja ningú sap qui i per què va plantar en un lloc determinat, com els passava a aquells xiquets que jugaven a pilota al carrer. Com assenyala el seu editor, Joan Vergés, cal preguntar-se si el pensador que fou Balmes continua sent alguna cosa més que una simple retolació de carrer. El recull d'articles són ponències presentades en el Simposi Trias sobre Jaume Balmes que la Càtedra Ferrater Mora, conjuntament amb l'Institut de Ciències Religioses de Vic, va organitzar l'any 2016, i que tenia com a intenció avaluar la petjada del llegat de Balmes i la filosofia del segle XIX en la nostra cultura catalana actual.

Joan Vergés, fent un espòiler en el prefaci del llibre, ja ens anticipa que l'exa-

men no augura un resultat gaire satisfactori, conclusió que rebla amb unes paraules d'Ignasi Roviró: «Avui, Balmes no interessa, ni per anar-hi en contra» (pàg. 62). Si més no, les diferents col·laboracions del llibre presenten lectures de l'obra de Balmes que no permeten fer una valoració conclusiva, en un sentit taxatiu de pervivència o oblit, del seu llegat. Vergés apunta que aquest no era tampoc l'objectiu del simposi, «la intenció era simplement *revisitar* l'obra de Balmes per tal de comprovar quin llegat en queda, avui, si és que en queda cap» (pàg. 12). Així, doncs, cal restar molt atent a les variades arestes de la personalitat del capellà de Vic que es dibuixen en el llibre.

Les ponències aplegades en el llibre segueixen tres direccions: la primera acompleix exactament la intenció del títol del llibre, en intentar identificar la pervivència del pensament i l'obra de Balmes en el segle XX, o la utilització, segons propòsits diversos, que la posteritat n'ha fet; la segona estira intuïcions de l'obra balmesiana i les confronta amb autors posteriors o amb problemàtiques actuals; l'últim camí, al meu parer el que més s'allunya de la finalitat que s'exposa en el prefaci, se centra en l'explicitació d'intuïcions originals del pensament de Balmes.

El llibre comença amb dues ponències, la inaugural del simposi, a càrrec de Conrad Vilanou, i la d'Ignasi Roviró, les quals són les que s'orienten més clarament en la primera direcció. Vilanou («Glòria i dissort de Balmes en el segle XX») ressegueix la presència de Balmes al llarg del segle XX de la mà de les commemoracions dels centenaris del seu naixement

(1910) i de la seva mort (1948) i durant la Transició, en els anys setanta. El centenari del naixement de Balmes és aprofitat, segons Vilanou, per vigoritzar el que esdevindria un noucentisme catòlic, que pretenia una renovació en tots els àmbits d'una Església molt més porosa a l'embat laïcal. Representants d'aquest revifament eclesial foren el bisbe Torras i Bages, Miquel d'Esplugues, Ignasi Casanovas i els canonges Cardó i Manyà. Pel que fa a la commemoració del centenari de la mort (1948), el franquisme rescata la política d'unitat preconitzada per Balmes com un dels seus triomfs en aconseguir la unitat política i religiosa que tantes discòrdies va ocasionar durant el segle XIX, en una clara utilització pel règim franquista de Balmes per bastir el seu nacional-catolicisme. Finalment, la Transició va propiciar que els dirigents de l'Església del moment, els cardenals Tarancón i Jubany, recobressin l'escrit de Balmes *Pius IX* com un precedent d'allò que havia de permetre l'entesa de l'Església amb els nous temps polítics democràtics, caracteritzats per la tolerància, la defensa de la llibertat i el pluralisme.

Ignasi Roviró, en la seua ponència «*El Criterio*, vida i èxit d'una idea», exposa, d'una manera detallada i documentada, les lectures i interpretacions que l'obra més coneguda de Balmes ha provocat des de la seua publicació. Roviró distingeix tres paradigmes de lectura d'aquest llibre: lògica, pedagògica i psicològica. La lògica i la pedagògica esdevenen les primerenques, alhora que paral·leles i en algun cas no excloents. La tercera, la psicològica, té

com a protagonista el bisbe Torras i Bages, el qual esdevindrà el catalanitzador d'*El Criterio* en considerar-lo expressió del caràcter i la manera de ser catalana, el *Codech del seny català*, en paraules de Clascar. Així, *El Criterio* passa a ser l'obra balmesiana per excel·lència. Més endavant, després de l'oblit que representa el Congrés d'Apologètica de 1910, la traducció del canonge Colell al català d'*El Criterio* (1911) origina una acció convergent de diverses institucions i diaris que convertiran aquesta obra en un llibre de capçalera a les llars catalanes. L'objectiu és crear una Catalunya catalana i cristiana al voltant dels valors rurals i pairals, que encarnen el seny català. Roviró també ressegueix la sort d'*El Criterio* sota el franquisme i les noves edicions d'*El Criterio* durant aquest període.

En la línia de les dues ponències anteriors, Pompeu Casanova («La matriu catalana del pactisme català») aborda el pensament polític de Balmes i el fa continuador del pactisme medieval català. Aquest pactisme, condicionat per les circumstàncies històriques que visqué, esdevé pactisme polític, anticipant l'estratègia del catalanisme posterior, és a dir, una matriu que integra les dimensions religioses, les condicions materials històriques i el realisme polític. Es tractaria d'un programa possibilista que ja anticipava Balmes quan afirmava: «Negociar para transigir, alcanzar lo que pueda y resignarse con su mala suerte» (pàg. 137).

Nazzareno Fioraso («La veritat és la realitat de les coses. Breu història de la fortuna de Jaume Balmes») analitza la dissort actual del pensament de Balmes.

La tesi de Fioraso per explicar aquesta mala estrella és doble: a) el *Doctor Humanus* és massa liberal per als conservadors i massa conservador per als liberals, la qual cosa provoca una marginació en ambdós àmbits; b) la resistència de Balmes als filòsofs moderns, que en qualsevol cas ell mateix va donar a conèixer en l'erm de la filosofia espanyola del XIX, el condemna a un gradual oblit.

Antoni Mora, per acabar amb aquesta avaluació de la influència del pensament de Balmes en la posteritat, situa la figura de Balmes entre la plèiade de filòsofs polítics de principis del segle XIX, al costat de Tocqueville, Blanqui, Mill, Proudhon i Donoso Cortés, compartint una sèrie de trets comuns que converteixen Balmes en «filòsof i polític» (Batllori *dixit*).

La resta de ponències esporguen de manera més concreta traces del pensament de Balmes sobre autors posteriors o sobre camins de reflexió actuals en què la petja de Balmes es deixa sentir. Així, Lluís Quintana ressegueix la influència de Balmes en l'obra de Mañé i Flaquer; Casasús, en el periodisme català; Ivan Sánchez aprofundeix en l'arrel psicològica de l'art balmesiana en l'obra d'Ors i Pujols; Boldú assaja un acarament entre Balmes i Turró; Gensollen i Ramírez bussegen en les aportacions de Balmes a l'epistemologia del testimoni; Miró repensa els conceptes de consciència i tradició en Balmes i el bisbe Torras i Bages; Cuscó explora les idees estètiques de Balmes; Francesc Codina, les literàries; Alcoberro s'endinsa en el laberint del seny; Jaume Codina extreu lliçons sobre

l'ensenyament de la religió; i Max Pérez, sobre la immortalitat de l'ànima dels animals. Déu-n'hi-doret, quin tresor balmesià! Tota aquesta riquesa d'intuïcions semblaria que confirmen un gran llegat, si més no...

Després d'haver llegit les diverses comunicacions, una sensació de derrota ens envaïx, ja que el llegat de Balmes, sobre el qual s'interrogava el títol del llibre, s'ha perdut o, almenys, els seus rèdits semblen haver-se exhaurit a hores d'ara. Coincideixen amb aquesta opinió: Vilanova, quan afirma que «la presència de Jaume Balmes al llarg del segle XX, en un procés de gran recorregut que va portar-lo de la glòria a l'oblit» (pàg. 48); Roviró, assenyalant que «Balmes és l'oblidat més important dels nostres clàssics» (pàg. 62); i Fioraso, reblant que «el seu nom va ressonar molts anys després de la seva mort, i en part ressona fins i tot ara, però gairebé només com un “pare noble”» (pàg. 209). Si bé és cert que ens podem consolar afirmant que els clàssics ho són perquè ja no els llegeixen (Roviró), no ens veiem amb cor d'atrevir-nos a escriure un d'aquells epitafis rotunds, que tant li agradaven fer a Josep Pla com a *oratio finis* sobre els nostres homenots il·lustres, i acabar exclamant que «Balmes és Balmes», i amb això ja ha estat dit tot quan no s'ha dit res, i continuar la vetlla del mort amb paraules de simple compliment. El millor encomi que podríem fer a Balmes és recordar-lo llegint-lo.

Ramon ROSALES SEBASTIÀ
Facultat de Filosofia (URL)

Ramon M. Nogués,

La salut espiritual.

Barcelona: Fragmenta Editorial,
2016, 322 pàg.

El darrer dels llibres del reconegut biòleg i capellà Ramon M. Nogués és un tractat sobre l'espiritualitat entesa com una dimensió antropològicament essencial que configura i aconsegueix l'ésser humà. En el marc de la interrelació de les diverses competències conegudes que defineixen la persona humana, estudiades des de la perspectiva de la neurobiologia fonamental, aquest assaig s'endinsa en aspectes absolutament transcendents de l'experiència del viure des del coneixement més biològic del sistema nerviós central. Mitjançant la comprensió anatomofuncional del cervell vinculat a les necessitats més primàries i biofísiques, també analitza l'espai intel·lectual i del raonament, de la mà de la dimensió emocional de la persona. L'objectiu és analitzar el que l'autor anomena el cervell de luxe, el qual ens obre el viure a la plenitud de l'experiència estètica, ètica i espiritual.

El llibre de Nogués parteix d'una reflexió primera sobre la desorientació espiritual que pateix el món del segle XXI, en especial a Occident. Ho fa sota la premissa clara del reconeixement ja inexcusable de la capacitat de la ment humana de fer possible la consciència reflexiva, la concepció del propi jo, i l'opció d'un llenguatge simbòlic abstracte i recursiu, que han fet possible el seu raonament. Argumenta que l'espiritualitat neix obligadament d'una necessitat hu-

mana d'horitzons on simbolitzar respostes a qüestions essencials per al viure. La descriu com una necessitat no necessària que esdevé possible des de la llibertat de l'experiència humana en evolució.

Comença descrivint el cervell humà com una de les més perfectes realitats mai creades per l'evolució de les espècies. La relació del cervell amb el cos humà s'explica per diverses xarxes perfectament travades i estructurades anatomo-funcionalment. És gràcies a un sistema nerviós central que participa de les funcions superiors i més complexes de la persona i d'un sistema nerviós perifèric al qual es vincula. Aquest darrer s'estructura en un sistema nerviós esquelètic, responsable de tasques sensorials i motrius, i un de vegetatiu o autònom, que fa possibles les funcions vitals més essencials. Insisteix l'autor que aquest entramat està perfectament assistit i dibuixat en xarxa. Totes les parts hi participen, esdevenint un conjunt harmònic de plurals dimensions que ens fan viure com un tot, compartint experiències somàtiques, emocionals, intel·lectives i, finalment, com es defensarà, espirituals.

Entre d'altres, l'autor fonamenta l'antropològica dimensió luxosa de l'espiritualitat en la persona sobre la tesi que prové de reconèixer i demostrar la clara correspondència entre funcions cerebrals i fenòmens mentals. El sistema nerviós s'explica per algunes funcions determinants: la funció de coordinació i homeòstasi de l'organisme, pròpia del tronc cerebral, la part més arcaica del nostre sistema; les funcions de supervivència elementals i les bàsiques d'actua-

ció grupal, que, gràcies al que anomenem diencèfal, actua de subcervell de la vida vegetativa; les funcions relacionades amb reconèixer, integrar i coordinar el món emocional del viure, ubicades en l'àrea límbica cerebral, i el neocòrtex, que cobreix tot l'encèfal. El còrtex, aquesta estructura que s'ha desplegat extraordinàriament en les darreres evolucions dels homínids, fa les funcions d'abstracció de les realitats, del raonament, de l'ús del símbol i el llenguatge i de la consciència reflexiva que li permeten plantificar, sentir-se lliure, així com el capaciten per a un «cervell de luxe» que explicarà experiències transcendents com la de l'espiritualitat.

La consciència com a funció mental diferenciada en l'espècie humana pot explicar-se segons models neurobiològistes. Tot i reconèixer-ne la realitat, no han acabat d'esbrinar amb certesa un espai neural concret que la pugui explicar. Els autors als quals es refereix Nogués coincideixen en el fet que, tot i la segura participació del còrtex prefrontal i temporoparietal, l'experiència personal de la consciència reflexiva és fruit d'una xarxa activa que implica diferents zones cerebrals. Altrament, podem considerar la consciència des de la mirada psicològica i des dels marcs filosòfic i ontològic del pensament, que la descriuen vinculada a l'esperit. En qualsevol dels casos, explica Nogués, l'abordatge de la consciència és molt proper al de les espiritualitats.

El jo, que no es descriu en el text com la consciència i que ens permet elaborar el nostre avançat sistema nerviós central,

té unes bases neurològiques clares, encara que no disposem d'una evidència que localitzi una específica àrea anatòmica que l'expliqui del tot. Aquest jo dona unitat a les diverses realitats de les quals fem experiència, siguin conscients o inconscients. El jo neurològic requerirà un jo psicològic que ens diferencia del no-jo, permetent-nos esdevenir conscientment únics i diferents. Igualment, es diferenciarà d'un jo ètic que explicaria els comportaments que orienten les nostres accions envers nosaltres o envers l'altre, tant la persona individual com el grup col·lectiu, que és aquell sobre els quals actua de manera inconscient la ment extensa de la cultura i el context social, donant forma al que coneixem com a jo social que vehicula espais de creixement i qualitat de vida positius. Finalment, queda el jo metafísic, tal com el descriu l'autor, que intenta transcendir en el dia a dia i que cerca anhelant una experiència que les diferents cultures han explicat. Experiències del tot, del buit, o que s'han personalitzat en deïtats. En qualsevol cas, la consciència reflexiva i el jo que el nostre sistema nerviós ens ha permès viure, d'acord amb l'evolució de la qual coparticipem, demanen la llibertat. Efectivament, el text, sense negar el principi científic de la causalitat, admet que els sistemes hipercomplexos contempen conjuntament la teoria de la indeterminació, que genera espais de llibertat que permeten abocar l'ésser humà a les transcendències.

Aquest trajecte humà vers les transcendències que depassen les necessitats

fonamentals, les quals s'expliquen gràcies a sòlides xarxes neuronals més que no pas a estructures neurals definides, gratuïtes i sempre obertes al misteri i a l'indeterminat, s'expressen amb l'imaginari i amb el simbolisme metafòric, així com permeten equilibrar i harmonitzar moltes funcions superiors vitals, somàtiques, cognitives i emocionals en la persona que hi accedeix. Es manifesten en espais que l'autor anomena com a transcendències, les quals poden ser de diversos tipus. Les transcendències estètiques són capaces d'explicar valors arcaics com els de la bellesa i la proporcionalitat, o l'art com a expressió cognitiva i emocional alhora, cercadora de veritats primeres. En segon lloc, les transcendències ètiques ens apropen a la vivència interior de valors i principis relacionats amb fer el bé, en evitar el mal i actuar segons els paràmetres de les accions que potencien la llibertat individual, la dignitat i la justícia social. Finalment, ens queden les transcendències espirituals i religioses, capaces de referenciar les nostres accions i comportaments, adhesions i vòlers a veritats superiors, realitats no empíriques, que donen profunditat a la vida i consistència als interrogants més profunds que suscita un viure que es vol viure amb sentit, ja sigui expressat en marcs institucionalitzats i referents de les religions diverses o bé en els de les nombroses savieses de vida, centrades en l'aprofundiment interior desinstitucionalitzat. Altres conceptes que l'autor relaciona amb les transcendències vinculades amb l'espiritualitat són l'amor i la utopia.

Sobre les espiritualitats, el text s'endinsa a remarcar unes característiques que les fan espais de dimensió humana excepcionals. En destaca la funció guaridora de la persona, en el sentit que el seu conreu adult permet equilibrar la persona, segons que han propiciat diverses escoles terapèutiques al llarg de la història de la humanitat. Té un efecte hedonista i balsàmic alhora, i propicia estats de felicitat i benestar del viure. La descoberta de l'espiritualitat capacita la persona en el compromís per a l'alliberament individual, per a la llibertat col·lectiva i per a la cerca d'un món més just, mentre fa una crida a estimar la natura i la creació tota com a autèntic i únic hàbitat heretat per l'ésser humà. És capaç de desvetllar el sentit per la bellesa i per la recerca de la mirada holística que ens permet experiències místiques i que eixamplen horitzons, des de l'experiència de la llibertat i de la unió personal amb el Tot.

Aquesta espiritualitat que explica l'autor com a dimensió antropològica inherent que caldria desplegar avui més que mai, demanarà un marc intel·lectual suficient que la faci entenedora i plausible, que tingui un fonament raonable. Aquesta saviesa s'haurà de poder inserir fins i tot en cada realitat històrica, creadora de l'equilibri necessari per viure del real, on ens és real el viure de l'home ressituat per la ciència, la psicologia i la biologia dels darrers segles. Demanarà també un món neural on les dimensions racional i emocional, relligades per sentiments humanitzadors, juguin un rol important com a pont entre el món emocional i el racional. Una de les prin-

cipals funcions de l'espiritualitat és la d'harmonitzar els espais emocionals amb els racionals del nostre món interior. Els sentiments, fruit de les emocions treballades a la llum de la raó, desplegaran finalment actituds sanes en la persona, virtuts del viure virtuosament i sana, que les diverses tradicions expliquen a bastament, relacionant-se amb escoles i pràctiques iniciàtiques diverses. Esdevenen constants espirituals universals, tant la confiança transcendental i el coratge ètic com la confiança en l'ésser.

Aquestes característiques de les espiritualitats a les quals hem arribat pel camí de la cerca d'espais cerebrals de luxe a partir de les xarxes del món somàtic, intel·lectual i emocional ens permeten fer experiència de Déu. Una experiència de l'inefable per definició, al qual ens referim per mitjà del llenguatge simbòlic, que, sense context ni història, ha estat capaç de lliurar l'home a la seva cerca des de sempre. Podem seguir-ne les traces des del més proper neolític, de l'edat de bronze o la de ferro o de la mateixa era axial de Jasper als segles VIII-VI a. C. Un Déu que tradueix les espiritualitats però no les explica totes. Imprescindiblement requereix de la persona humana el lliure albir, l'anhel amorós que suposa l'*ek-sistència*, la via sentimental d'accés i el reconeixement del Misteri.

En uns darrers capítols, el text ens il·lumina respecte de la relació que les espiritualitats tenen amb el sexe femení, destacant-ne el greuge històric comès, i les implicacions de la realitat social i cultural de cada entorn en el desplegament i la vivència espiritual humana.

L'espiritualitat, dimensió antropològica innegable que relliga les capacitats de la ment humana, esdevé així un espai luxós que ens permet transcendir, aprofundir en la recerca personal de Déu i equilibrar i harmonitzar la vida de manera sana i integradora.

Jaume DURAN
Facultat de Filosofia (URL)

Antonio Bergamo,
Levinas e la curvatura dello spazio intersoggettivo.

Roma: Città Nuova, 2018, 142 pàg.

L'obra d'Emmanuel Levinas (1906-1995) ha estat sense cap mena de dubte un dels esdeveniments cabdals de la filosofia contemporània. Jueu lituà, es trobà ja des de la seva infantesa a la frontissa entre Orient i Occident. Tingué el rus com a llengua materna i fou la densitat moral i espiritual de literats com Dostoiévski la que encoratjà la seva vocació filosòfica. Aprengué l'hebreu de nen, amb els seus germans, per acostar-se a les Santes Escripures, font hebrea del pensament que marcaria profundament la seva filosofia, seguint les petjades de Rosenzweig. Amb la mirada posada a Europa, es formà a Estrasburg i, posteriorment, a Friburg, on conegué Husserl i Heidegger. I el seu itinerari vital el portaria finalment a França, on fent-se seva la llengua de la pàtria que l'acollí —com si assumís radicalment, fins i tot en això, la condició exodal de l'existència jueva— escriuria en francès la seva obra filosòfica.

Dues han estat com a mínim les aportacions crucials de Levinas a la filosofia francesa del segle XX, tal com recordava Jacques Derrida en l'al·locució pronunciada amb motiu del seu traspàs (*Adieu à Emmanuel Levinas*): d'una banda, la introducció de la fenomenologia a França, als anys 30; de l'altra, el gir profund que ell mateix aniria operant en aquesta fenomenologia, i que el conduiria a afirmar l'ètica com a filosofia primera, la superació de l'horitzó ontològic heideggerià o la reivindicació de la santedat del Sant i de la seva arrel bíblica, més enllà de la sacralitat del cosmos grec i neopagà. En aquests dos esdeveniments decisius, hi trobarà també les seves fonts el *gir teològic de la fenomenologia francesa contemporània*, tant pel seu arrelament i aprofundiment de la proposta de Husserl com per la recuperació, des de la filosofia, d'un diàleg fructífer amb la religió i la teologia. Així, Levinas ha pogut ser llegit amb admiració des d'àmbits molt diversos del pensament, molt més enllà del pensament jueu, i això el fa encara més interessant: des del pare de la desconstrucció o el crític Maurice Blanchot, fins a filòsofs catòlics com Jean-Luc Marion.

Foren, de fet, els ambients catòlics els primers a rebre'l amb entusiasme, fins i tot abans que els mateixos jueus. El pare Van Breda encoratjà la publicació de *Totalitat i Infinit*, el mateix papa sant Joan Pau II —també format de jove en la fenomenologia de la mà del deixeble polonès de Husserl Roman Ingarden— el convidà en diverses ocasions a Castelgandolfo, i fou un salesià, Roger Burggra-

eve, qui n'organitzà la primera bibliografia exhaustiva. No ens hauria de sorprendre gens, doncs, que l'obra que ara presentarem hagi estat escrita per un teòleg catòlic i que proposi obertament el diàleg entre el pensament de Levinas i la teologia trinitària cristiana.

Antonio Bergamo, autor del llibre, és professor d'Antropologia teològica i d'Escatologia a la Facoltà Teologica Pugliese (Bari) i ha estat també recentment nomenat Director de l'Istituto Superiore di Scienze Religiose Metropolitano «Don Tonino Bello», a Lecce. El 2016 va publicar a la mateixa editorial una obra dedicada precisament a l'antropologia trinitària: *Identità reciprocanti. Figure e ritmica di antropologia trinitaria* (2016).

En la línia de la seva recerca precedent, intenta respondre, amb aquest volum, a la urgència de cercar un nou paradigma de pensament que ens permeti trobar categories per dir filosòficament allò que se'ns escapa: l'obertura a l'altre, l'alteritat del món i l'horitzó de sentit que ens depassa i en el qual Déu es deixa trobar. Bergamo es proposa apropar-se a aquestes qüestions prenent com a clau de volta la categoria de relació. El llibre, per tant, no pretén oferir un estudi sobre el pensament complet de Levinas, sinó aprofundir en aquests interrogants que ressonen també en el seu apropament fenomenològic. L'obra del filòsof d'origen lituà propiciarà, sens dubte, el diàleg entre filosofia, teologia, literatura i ciències humanes. D'altra banda, Bergamo no es proposa pas «cristianitzar» Levinas, però sí descobrir-hi una comu-

na vocació de la qual participen tant el judaisme com el cristianisme.

L'objectiu darrer de l'obra serà reflexionar sobre l'esdeveniment de l'amor en la relació antropològica oberta a la transcendència i com la subjectivitat humana hi realitza la pròpia llibertat, sobre la reciprocitat com a dinamisme i categoria del pensament, i sobre la possibilitat de l'experiència de Déu per al subjecte humà en tant que es pugui donar en la relació mantinguda en l'espai intersubjectiu. Com subratlla Bergamo, d'aquesta manera, «l'actualitat de Levinas es troba, així, amb allò més específic del cristianisme, que veu en l'esdeveniment pasqual de la mort i la resurrecció de Crist la clau hermenèutica del real i la via d'accés a una vida *trinititzada*, conscient que "la fe trinitària té un abast real"» (pàg. 8).

El llibre, en els capítols inicials que segueixen la Introducció, proposa un breu apropament a l'«humus» cultural i filosòfic en el qual es troba i amb el qual ha de dialogar la reflexió de Levinas. El primer capítol, *Platone, fenomenologia, rivelazione*, introdueix la figura del pensador d'origen lituà. El segon, *Il senso: notte della fenomenologia?*, aborda les nocions d'alteritat i subjectivitat, traçant un breu recorregut de l'evolució del pensament de l'alteritat en la filosofia moderna, i proposa, finalment, un apropament als trets que caracteritzen la postmodernitat. El tercer, *Al di là del neutro*, oposa a les filosofies de Hegel o de Heidegger una superació del pensament del neutre, duta a terme per Levinas gràcies al *logos* bíblic, i a un pensa-

ment de l'alteritat i de l'asimetria. A continuació, en el quart capítol, titulat *Acceso al reale*, Bergamo explorarà tres vies d'accés al real en la nostra experiència finita: el cos, el llenguatge i la intersubjectivitat.

Tot seguit, tres capítols centrals –del V al VII– examinen respectivament les categories que constitueixen les claus de volta del llenguatge de la reciprocitat fonamentada en l'amor: l'asimetria, la diacronia i la curvatura de l'espai intersubjectiu (*Asimetria; Diacronia; Curvatura*). L'*asimetria*, d'una banda, «no és concebuda com a dissimetria sinó, per descomptat, com una diferència que diu la possibilitat mateixa de l'encontre. Dir asimetria per a Levinas vol dir afirmar que jo soc pròxim de l'altre, independentment de la reciprocitat de la seva resposta» (pàg. 51). Aquesta asimetria és reconeguda en la separació de la realitat creada respecte de Déu, en la relació interpersonal amb els altres, en la relació entre allò masculí i allò femení (que Levinas pensà especialment a *Le Temps et l'Autre*), en la relació filial i en una reciprocitat asimètrica entesa com a do. Aquest capítol es clou amb un preludi teològic, en què la imatge d'Abraham, contraposada a la d'Ulisses, dona testimoni d'aquesta asimetria.

Per la seva banda, en el capítol VI, la *diacronia* «fa emergir un ulterior vector estructural del real i de l'humà: la temporalitat. La diacronia està contraposada, en un cert sentit, a la linealitat del temps» (pàg. 65). La diacronia, sosté Bergamo a partir de Levinas, és «descompaginament del flux temporal i

obertura a una alteritat de tipus qualitativament divers, és suspensió de la linealitat del temps i acompliment de la proximitat com a ésser l'un-per-a-l'altre, implicat en un *nosaltres-a-priori* originari que tanmateix depassa el mer agregar-se de les subjectivitats individuals. És per aquest motiu que allò diacrònic no pot ser tematitzat, “la diacronia és el refús de la conjunció, allò no tematitzable, en aquest sentit precís, l'Infinit.”» (pàg. 69-70). Aquesta relació diacrònica es realitza en l'esdeveniment de l'amor, en una crida i un do de si, que troba la seva font «en una transcendència que no desresponsabilitza la immanència, sinó que il·lumina la consistència escatològica del real» (pàg. 74). També aquest capítol culmina amb un nou preludi teològic, en aquest cas, proposant una relectura de la perícopa evangèlica dels deixebles d'Emmaús (Lc 24, 13-53).

En tercer lloc, en el capítol VII, Antonio Bergamo aborda la noció de *curvatura*. Si l'*asimetria* reinterpreta, per dir-ho així, l'*espai* de la relació i la *diacronia*, el seu *temps*, la *curvatura* repensa la convergència d'espacialitat i temporalitat des de l'interior de l'experiència relacional de l'home. L'autor del llibre es recolza, de fet, en una expressió –“*curvatura de l'espai intersubjectiu*”– que el mateix Levinas introduïa en les conclusions de *Totalitat i Infinit*, malgrat que aquesta imatge no torni a aparèixer posteriorment en la seva obra. Segons Bergamo, «les condicions particulars en les quals aquesta curvatura emergeix són l'experiència d'un amor intersubjectiu que suscita en la relació l'encreuament dels

subjectes en l'espai creat. És en aquesta pluralitat dels camps de sentit subjectius que *pot* irrompre la revelació de Déu» (pàg. 84). Antonio Bergamo rellegirà especialment, en aquest capítol, *De Dieu qui vient à l'idée* (1982).

Finalment, i abans de la Conclusió de l'obra, el capítol VIII ens proposa un diàleg molt més explícitament teològic entre Levinas i la teologia trinitària – *Tornante teologico: con-vocati dall'Altro* –, i s'endinsa en el gir teològic que aquelles tres categories indicaven, explorant allò que Levinas pot suggerir al saber filosòfic i teològic, per tal d'aprofundir, en particular, l'esdeveniment de l'amor en la relació antropològica, dinàmica i oberta a la transcendència. El capítol s'emmarca en l'evocació de la trobada que el mateix Emmanuel Levinas mantingué el 1986 amb, entre d'altres, Hans Hermann Henrix, Bernhard Casper i el bisbe i teòleg Klaus Hemmerle, i aborda d'una manera encara més directa el possible diàleg entre la filosofia i la teologia que s'obrí ja en aquella trobada. La tesi que Antonio Bergamo es proposa sostenir és que, en el fons, les traces que podem filosòficament reconèixer en la relació interpersonal es fan intel·ligibles quan l'enigma de la transcendència es desxifra per l'esdeveniment de l'Amor, des de la teologia trinitària, i en la seva significació escatològica. Per tal de desenvolupar aquesta qüestió, rellegeix especialment l'*ontologia trinitària* proposada per Klaus Hemmerle. Segons Bergamo, «l'amor ontològicament altre que l'esdeveniment trinitari desplega, no tan sols *per nosaltres* sinó també *entre*

nosaltres, és travessat per una tensió dialèctica/dialògica en la qual determina asimetries que revelen la identitat dels subjectes involucrats i fa reverberar dia-crònicament la tonalitat del Dir en el Dit que es manifesta en aquest circuit relacional» (pàg. 121).

Tal com Bergamo sosté en la Conclusió de l'obra – *Una conclusione aperta: attraversare l'abisso* –, aquesta experiència relacional, i especialment l'experiència de Déu, és alhora una *no-experiència*, quelcom d'inexhaurible que des del pla fenomenològic potser només es deixa copsar com un excés que desborda i que atreu cap a si, o com la pobresa extrema i l'abisme que apunta més enllà de si mateix. No obstant això, la transcendència de Déu capgira l'espai viscut i ens obre a l'acompliment de la relació més enllà de la immanència. «En aquesta perspectiva la teologia trinitària té molt a dir i a explorar sobre aquesta reciprocitat que, en definitiva, sorgeix de la Trinitat» (pàg. 132).

Joan CABÓ RODRÍGUEZ
Facultat de Filosofia (URL)

Maciej Bała, Piotr Stanisław Mazur (eds.), *The Polish Christian Philosophy in the 20th Century (Mieczysław Kamiński; Kazimierz Kłoskowski; Kazimierz Kłósak; Feliks Koneczny; Mieczysław Albert Krąpiec; Tadeusz Ślipko; Józef Tischner; Karol Wojtyła; A companion to Polish Christian Philosophy of the 20th and 21st Centuries)*.

Cracòvia: Ignatianum University Press, 2019-2020, 9 vol.

Publicats per la Ignatianum University Press, de Cracòvia, es presenten nou volums més de la col·lecció *Filosofia Polonesa Cristiana en el segle XX*, que es proposa presentar el pensament i l'obra de representants destacats de la filosofia polonesa cristiana contemporània. Els nous volums publicats recullen representants de diversos àmbits de la filosofia (des de l'ètica fins a l'antropologia, la fenomenologia, la metafísica, la filosofia de la cultura o la filosofia de la natura). Aquesta col·lecció pretén esdevenir un focus de recerca i una font d'inspiració intel·lectual. Cadascun dels volums publicats fins al moment s'ha consagrat a un pensador, del qual se n'edita una selecció de textos traduïts a l'anglès, i precedits per una presentació biogràfica i filosòfica a càrrec de diversos especialistes en la seva obra. Per la seva banda, el volum titulat *A companion to Polish Christian Philosophy of the 20th and 21st Centuries* constitueix una presentació sintètica dels assoliments acadèmics d'aquests filòsofs cristians en les diverses branques del pensament.

Miguel Á. Granada, *Filosofía y religión en el Renacimiento*.

Sevilla: Cypress Cultura / Editorial Thémata (Colección Humanistas 2), 2021, 412 pàg.

Es reuneix en aquest volum una recopilació de treballs sobre la cultura filosòfica del Renaixement escrits i publicats al llarg dels darrers vint-i-cinc anys per Miguel Á. Granada, una de les autoritats mundials en la matèria. El conjunt il·lumina el pensament d'autors molt influents però poc coneguts, com és el cas de Gemist Pletó o Agostino Steuco, així com sobre qüestions de màxima importància per al posterior desenvolupament de la cultura europea: el nexa entre Antiguitat, Edat Mitjana i present; la relació entre Reforma religiosa i Revolució científica, de Copèrnic a Galileu; o el concepte de *libertas philosophandi*, entre d'altres. No hi manquen tampoc anàlisis sobre les aportacions al debat filosòfic per part de figures remarcables de l'època: Giovanni Pico de la Mirandola, Nicolau Maquiavel, Thomas More, Erasme de Rotterdam, Giordano Bruno, etc. En definitiva, aquest és un llibre d'un gran valor pels aspectes innovadors que planteja, així com per la

rigorosa fonamentació de les seves conclusions, i pot resultar útil tant als apassionats pel Renaixement com a aquelles persones que desitgin ampliar els seus coneixements sobre aquesta època sempre actual i seductora.

Giorgio Agamben, *Arqueologia de la política*.

Trad. Robert Garcia Orallo.

Barcelona: Arcàdia, 2019, 106 pàg.

L'abril del 2014, Giorgio Agamben va ser convidat per la Càtedra Ferrater Mora a pronunciar les lliçons que es recullen ara en aquest volum. Va obrir les sessions amb un apunt metodològic: fer de l'arqueologia la via d'accés al present. Aquesta disciplina, que li permet anar al moment d'emergència d'un determinat fenomen i dibuixar-ne la genealogia fins avui, traçarà el camí de les seves reflexions. Conceptes com «crisi», «identitat», «ordre», «voluntat», «a priori històric», «ús» o «política» troben en aquestes pàgines una nova potencialitat en allò que Agamben defineix com a antropogènesi, és a dir, l'esdevenir humà de l'home. I això mirant de dur sempre el llenguatge fins als seus límits. La filosofia, per a Agamben, és una intensitat que pot travessar i donar vida a qualsevol àmbit.

**Joan Vergés Gifra, Francesco Ardolino, Marta Nadal Brunès (eds.),
Maria Aurèlia Capmany. Escriptora i pensadora.**

Girona: Documenta Universitària / Publicacions de la Càtedra Ferrater Mora, 2019, 306 pàg.

Aquest llibre és la culminació i el resultat directe del Simposi Trias sobre Maria Aurèlia Capmany que va tenir lloc a Barcelona els dies 7, 8 i 9 de desembre de 2018 i que va ser un dels actes més destacats de l'any Capmany. El simposi va esdevenir una ocasió magnífica perquè alguns dels principals especialistes en l'obra i el pensament de l'autora fessin una revisió seriosa i fonamentada de la seva figura i la seva producció narrativa i assagística i, alhora, n'avaluessin el llegat. Així mateix, va representar una oportunitat preciosa perquè una colla de lectors avesats a la filosofia i les lletres, alguns dels quals veus novíssimes del panorama crític català, s'afegissin a l'estudi de l'autora. Després de les celebracions de l'efemèride i de la publicació d'aquest volum, es fa molt difícil negar que Maria Aurèlia Capmany ha esdevingut una veu imprescindible de la literatura catalana contemporània i una escriptora de referència tant en el terreny de la ficció com en el de la reflexió i l'activitat social, en què ressalta la seva funció pionera en les reivindicacions feministes de la segona meitat del segle XX.

Manuel Fraijó, *Semblanzas de grandes pensadores.*

Madrid: Trotta, 2020, 462 pàg.

Manuel Fraijó, catedràtic emèrit de Filosofia de la Religió de la UNED, recull en aquest llibre una selecció de conferències per mitjà de les quals s'apropa a una àmplia galeria de pensadors que han marcat la vida a Orient i Occident, des de Confuci o Luter, fins a Descartes, Pascal, Kant, Kierkegaard, Nietzsche, Bergson o Karl Rahner, entre d'altres. És possible que les recerques d'aquests pensadors ja no siguin les nostres, però el seu pensar i el seu sentir continuen marcant el nostre present, per molt que aparentment els haguem pogut oblidar.

Francisco Sánchez, *Que nada se sabe.* Traducción, introducción y notas de M^a Asunción Sánchez Manzano.

Madrid: Tecnos, 2020, 180 pàg.

Francisco Sánchez (1551-1623) s'havia format a la Universitat de Montpeller, i va exercir la medicina a Tolosa. Rere l'escepticisme, que sembla declarat en aquest títol *Quod nihil scitur*, sorgeix una proposta d'estudi de les condicions vigents per a l'ensenyament de la filosofia natural, i de la medicina en particular. Planteja un discurs sobre el subjecte del coneixement, les seves capacitats i dificultats, apostant per un nou mètode en la fonamentació de les ciències. Com a partícip de la cultura humanista, que guardava el record del *De rerum natura* de Lucreci, el seu pensament preludia els debats del *Novum Organum* de Francis Bacon i del *Discurs del Mètode* cartesià.

Giordano Bruno, *Cábala del caballo Pegaseo.* Traducción, introducción y notas de Miguel Á. Granada.

Madrid: Tecnos, 2020, 166 pàg.

Després de defensar l'heliocentrisme copernicà a *El sopar de les Cendres*, i d'haver-lo ampliat a un univers infinit tot exposant el substrat ontològic d'aquest univers infinit i substancialment u en *De la causa, el principi i l'u* i en *De l'infinit: l'univers i els mons* (publicats tots ells a Londres el 1584), Giordano Bruno extreu les implicacions morals, polítiques i religioses de la nova imatge de l'univers en els tres diàlegs morals, publicats també a Londres el 1584 i el 1585. El segon d'aquests diàlegs és la *Càbala del cavall Pegaseu* (1585), que continua la crítica del cristianisme, ja plantejada en l'*Expulsió de la bèstia triomfant* (1584), mitjançant una crítica, per la via de l'elogi irònic, de l'antiintel·lectualisme i de l'escepticisme. Al mateix temps s'hi defensa la doctrina de l'ànima universal amb els seus accidents particulars en constant mutació sobre la superfície de la matèria universal.

**Friedrich Nietzsche, *El Anticristo. Maldición sobre el cristianismo.*
Edición, traducción y notas de Joan B. Llinares.**

Madrid: Tecnos, 2019, 140 pàg.

L'Anticrist és una de les últimes obres de Nietzsche, escrita el 1888, però no publicada fins després del seu col·lapse mental. N'apareix ara aquesta edició i traducció de Joan B. Llinares. L'obra se centra en el significat i les repercussions del triomf del cristianisme al final del món antic, i com determina i marca la història posterior fins al present. Insisteix en la distinció entre el Jesús històric i la figura de Crist, elaborada per sant Pau i els Pares de l'Església. Per a Nietzsche, aquest Crist no és sinó una construcció teològica necessària per a la fonamentació de la doctrina oficial de l'Església com a religió política. El sentit profund d'aquest llibre, tanmateix, és el de personificar en aquest Crist paulí els valors ascètics que han originat el nihilisme, i fer una crida a la necessitat d'una transvaloració o canvi de valors, simbolitzada en l'expressió: «Dionís enfront del Crucificat».

Friedrich Nietzsche, *Así habló Zaratustra. Un libro para todos y para nadie.* Prólogo, traducción y notas de Alejandro Martín Navarro.

Madrid: Tecnos, 2020, 344 pàg.

Apareix publicada a Tecnos la traducció d'Alejandro Martín Navarro d'*Així parlà Zaratustra. Un llibre per a tots i per a ningú*. Aquest és el títol que escollí Nietzsche per a l'obra que ell mateix descriuria com la culminació i el zenit del seu llegat filosòfic, una obra que reclama per a si la consideració de cinquè Evangeli i de text sagrat, on els gèneres poètic, profètic i filosòfic es confonen sorprenentment. En aquesta obra, el vell profeta iranià baixa de la seva muntanya, després d'anys de soledat, per tal de compartir amb el món un missatge mai abans sentit: la notícia de la mort de Déu i la vinguda del superhome, la negació dels ídols creats pel ressentiment i l'experiència abismàtica de l'etern retorn. Tots els temes del pensament nietzschian hi conflueixen: la crítica a la cultura occidental, la denúncia d'una societat regida per les masses, el nihilisme, la transvaloració de tots els valors. Però el més peculiar d'aquesta obra és la manera com va més enllà de tots aquests tòpics i ens condueix al caos creador d'un llenguatge que ha exercit una influència inaborrable en la cultura contemporània.

José Díez (compilador), *Exploraciones pluralistas: las filosofías de C. Ulises Moulines.*

Madrid: Tecnos, 2019, 260 pàg.

Aquesta obra inclou els treballs de destacats filòsofs d'Espanya, l'Amèrica Llatina i els EUA sobre el pensament de C. Ulises Moulines, la major part dels quals originats en les

contribucions aportades pels autors a un simposi realitzat el juliol de 2014 a Barcelona sobre la trajectòria d'aquest pensador. El volum s'estructura en tres parts principals, precedides per una semblança intel·lectual a càrrec del compilador, i seguides d'una coda sobre Moulines i la literatura. Cada part agrupa els treballs relacionats amb un dels àmbits de la trajectòria filosòfica de Moulines, i conclou amb les seves respostes als autors. La part I se centra en aspectes metodològics i sistemàtics sobre el conjunt de la seva filosofia. Els treballs de la part II s'ocupen de l'obra de Moulines en filosofia de la ciència, l'àmbit filosòfic en què més ha destacat, especialment com a cofundador i màxim representant en les darreres dècades de l'estructuralisme metateòric. La tercera i última part agrupa col·laboracions que se centren en l'obra filosòfica de Moulines no directament vinculada a la filosofia de la ciència: filosofia del llenguatge, epistemologia general, metafísica i filosofia pràctica.

Mauricio Suárez, *Filosofía de la ciencia. Historia y Práctica.*

Madrid: Tecnos, 2019, 216 pàg.

Malgrat la seva brevetat, aquesta obra és ambiciosa, tant per la seva voluntat de síntesi com pel seu inusual caràcter híbrid. Combina objectius de divulgació, docents i de recerca. D'una banda, proporciona una primera aproximació a la pràctica de la filosofia de la ciència, accessible a qualsevol persona interessada. D'altra banda, constitueix un breu manual docent, en forma de bibliografia comentada, útil tant per a professors com per a alumnes, en cursos de grau i màster, en facultats de ciències i d'humanitats. Per últim, es tracta d'un monogràfic que defensa una tesi pròpia, al voltant del deute de la filosofia contemporània de la ciència respecte del positivisme lògic del Cercle de Viena. El resultat és el primer tractat amb una perspectiva històrica que arrenca en el Cercle de Viena i culmina en el debat actual, de marcat caràcter epistemològic, sobre el realisme científic. Aquesta particular manera d'abordar la matèria, en un moment en què la disciplina experimenta una revisió a fons de la seva historiografia, permet apreciar les arrels profundament europees de la nostra actual reflexió filosòfica sobre la ciència.

Juan Arana (dir.), *La cosmovisión de los grandes científicos del siglo xx. Convicciones éticas, políticas, filosóficas o religiosas de los protagonistas de las revoluciones científicas contemporáneas.*

Madrid: Tecnos, 2020, 524 pàg.

L'objectiu habitual de la filosofia consisteix a respondre preguntes relatives a les fronteres del coneixement, al sentit de l'existència, als principis de l'acció humana o a la convivència social. No obstant això, en el segle XX els filòsofs van adoptar una actitud més previnguda, van cercar formes d'expressió més críptiques, van pretendre assolir fites de

rigor més elevades a canvi de retallar les seves ambicions. Pel que fa a les qüestions últimes, els científics han compensat molt sovint aquest retraïment. Molts d'ells s'han atrevit a especular sobre assumptes tan controvertits com l'origen de l'univers i el destí de l'ésser humà. Ho han fet com a colofó d'investigacions més concretes i empíriques, al fil d'escrits d'alta divulgació, textos autobiogràfics, assaigs, allocucions o conferències. Aquest llibre recopila, exposa i critica la cosmovisió de 39 autors escollits d'entre els científics contemporanis més rellevants en tots els camps de la recerca.

José Ruiz Fernández, *Comprensión, significado y lenguaje.*

Madrid: Tecnos, 2020, 368 pàg.

Aquest llibre analitza alguns conceptes que ocupen un lloc central en moltes discussions filosòfiques: significat, llenguatge, intencionalitat, comprensió, interpretació, normativitat, ús, seguir regles, necessitat lògica, etc. L'anàlisi s'ha dut a terme sense donar per descomptat cap marc teòric i assoleix uns resultats originals, és a dir, exposa vincles conceptuals potser encara impensats. Aquests vincles es posen en relleu, en qualsevol dels casos, no pas per la seva possible originalitat, sinó a fi que un lector amb inquietuds teòriques pugui superar certes incongruències i confusions força comunes. El pensament madur de Wittgenstein ha estat un gran recolzament i font d'inspiració en la realització d'aquest llibre. La seva obra, no obstant això, no se segueix, comenta o cita en aquest treball, i fins i tot és esmenada en alguns punts. Aquestes divergències no poden posar en qüestió un deute, proximitat i comunitat de propòsit molt més bàsics.

Moisés González García, Hugo Castignani (coords.),

Filosofías del Barroco.

Madrid: Tecnos, 2020, 616 pàg.

El Barroc, des de fa algun temps, ha adquirit un especial relleu com a categoria filosòfica. Alliberat dels límits de les disciplines teòriques o de la cronologia, transformat en un espai de reflexió singularment productiu, el concepte de barroc ha passat a aplicar-se a diferents períodes de l'esdevenir històric, no tan sols de les arts sinó també del pensament i de la cultura, i amb una particular intensitat en el nostre present. Pel camí, les contribucions a l'estudi del barroc i el neobarroc han assolit un estatut de disciplina en l'àmbit de l'estètica i la teoria crítica. Així doncs, el propòsit d'aquest volum és oferir un quadre general, des de múltiples perspectives però sota un prisma filosòfic comú, d'aquest moment fonamental en la constitució de l'ésser modern, i en totes les seves arestes: des de Hobbes a Suárez, passant per Pascal, Gracián o Spinoza, de l'estètica a la teoria política. Però pensar el barroc és inevitablement repensar-lo, i la seva crítica implica sempre una reflexió sobre les eines que aquesta utilitza. És per això que en aquest volum es pretén també

mostrar quina i de quina manera ha estat la seva recepció en el pensament contemporani, per tal de poder preguntar-nos, d'una manera més general, sobre la pertinència d'aquest concepte en la filosofia i en el nostre present particular: Què constitueix una filosofia barroca? Com podem parlar de barroc en l'actualitat? I en quin sentit som en un nou moment neobarroc?

**Adela Cortina, *Ètica mínima. Introducció a la filosofia pràctica.*
Decimoctava edició.**

Madrid: Tecnos, 2020, 320 pàg.

Es publica a Tecnos la divuitena edició d'aquesta obra de la filòsofa espanyola Adela Cortina, apareguda per primera vegada l'any 1986. El nostre temps és temps de modèstia. La reflexió defuig els grans sistemes i l'acció de les grans empreses. No obstant això, les preguntes per la rectitud i la justícia, per la legitimitat del poder i l'esperança de la salvació continuen demanant resposta a una cultura que necessita respondre-les per recobrar el seu sentit. Abordar aquestes qüestions és el propòsit d'aquest llibre. Amb aquesta finalitat, s'endinsa en aquest àmbit del saber anomenat «pràctic», des de les quatre dimensions que el configuren: moral, política, dret i religió. Des d'aquestes s'il·lumina avui, no tant una ètica de màxims, sinó una ètica de mínims. Però una ètica que es resisteix a renunciar al millor que hem après després de segles d'història: el valor de l'autonomia humana i la necessitat d'un consens –entès com a concòrdia, no com a estratègia– per a l'organització de la vida jurídica i política. *Ètica mínima*, per la seva claredat i fonamentalitat, pot servir de manual a tots aquells que desitgin introduir-se en un saber pràctic però irrenunciable.

COMPRENDRE

revista catalana de filosofia

Vol. 22/2

Any 2020

Editorial 5

Oriol Quintana

Articles / Articles

- La Mística como refugio para el *Espíritu de la Verdad*.**
Simone Weil en relación a otros filósofos y místicos 7
The Mystics as a refugee for the Spirit of the Truth.
Simone Weil linked to other philosophers and mystics
M^a Ángeles García-Carpintero Sánchez-Miguel
- Simone Weil. Una respuesta a la lejanía de Dios** 29
Simone Weil. An answer to the remoteness of God
Carmen Herrando
- Simone Weil, discípula de Alain** 49
Simone Weil, disciple of Alain
Carmen Revilla
- Simone Weil: història i nació** 69
Simone Weil: History and Nation
Josep Otón
- À la racine de l'enracinement weilien – l'arbre comme symbole réel** 79
On the roots of the Weilian's rootedness – the tree as a real symbol
Christine Zyka

Nota crítica / Critical reviews

- L'identité. Une supposition ou une réalité ?** 95
The identity. A supposition or a reality?
Robert Chenavier

Antonio Diéguez, *Transhumanismo. La búsqueda tecnológica del mejoramiento humano.* 105

Juan Ramón La Parra Martínez

Ignasi Fuster i Camp, *Meditación sobre el hombre. Una propuesta de synthesis antropológica.* 107

Ferran Blasi Birbe

Grancisco Quesada Rodríguez, *La bioética de la responsabilidad según Hans Jonas.* 109

Ricardo Mejía Fernández

G. K. Chesterton, *Vegetarianos, imperialistas y otras plagas. Artículos 1907.*
Traducció de Montserrat Gutiérrez Carreras. Edició i introducció
de Pablo Gutiérrez Carreras i María Isabel Abradelo de Usera. 113

Sílvia Coll-Vinent

Llibres rebuts / Books received 117

Índex anterior / Previous Index 121

Normes de publicació / Guideline for contributors 123

NORMES DE PUBLICACIÓ

- *Comprendre* és una revista de caràcter científic i de recerca que es publica dues vegades a l'any i que està oberta a treballs que tractin els àmbits clàssics de la filosofia: metafísica, epistemologia, lògica, ètica, filosofia de la ciència i de la natura, antropologia, història de la filosofia, filosofia de la religió, estètica, etc. Està dirigida a un públic universitari interessat pel debat filosòfic i humanístic actual.
- *Comprendre* accepta tres tipus de treballs: articles, notes crítiques i recensions. Els articles i les notes crítiques han de ser originals i inèdits i han d'estar escrits en català o en les principals llengües europees. Només s'admetran recensions en català.
- Cal enviar a l'adreça electrònica de la revista (comprendre@filosofia.url.edu) un fitxer preferentment en format Word. Els articles no sobrepassaran les 10.000 paraules (notes i bibliografia incloses), mentre que l'extensió màxima de les notes i de les recensions serà de 6.000 i 2.500 paraules, respectivament.

Tipus de lletra: *Times New Roman*, cos 12, interlineat 1,5.

- Cal incloure-hi un resum en la llengua del treball i en anglès (*abstract*), d'unes deu línies, destacant-hi, a més, les paraules clau en ambdós idiomes.
- Les anotacions a peu de pàgina es numeraran correlativament. Les referències bibliogràfiques es poden presentar també al final del text, sempre per ordre alfabètic d'autors.

Cal seguir les següents normes d'estil en totes les citacions.

Tipus de lletra: *Times New Roman*, cos 10, interlineat senzill.

a) per als llibres:

INICIAL DE L'AUTOR. COGNOM (en versaleta), *Títol* (en cursiva). Lloc d'edició: Editorial, any.

Exemple:

R. DESCARTES, *El discurs del mètode*. Traducció i edició a cura de P. L. Font. Barcelona: Edicions 62, 1999.

b) per als articles de revista:

INICIAL DE L'AUTOR. COGNOM (en versaleta) «Títol de l'article». *Nom de la publicació periòdica* (en cursiva) [Lloc d'edició], 000 (número), 0000 (any), p. 00-00.

Exemple:

C. LLINÁS, «Gerhard Krüger: Einsicht und Leidenschaft (Intel·ligència i passió). Una entrada «platònica» en el pensament del segle XX». *Comprendre. Revista catalana de filosofia* [Barcelona], IX/1-2, 2007, p. 159-189.

- Cal afegir al final de l'article les referències bibliogràfiques utilitzades.
- Les dades personals i acadèmiques de l'autor s'han de presentar en un fitxer a part. Han de constar-hi nom i cognoms, la institució de procedència i una adreça electrònica vigent.
- *Comprendre* segueix els criteris de conducta ètica per a la publicació dels articles i les notes crítiques. Per això requereix que els/les autors/autores adjuntin un compromís signat de compliment de bones pràctiques juntament amb els seus manuscrits. N'està disponible un model a la web de la revista.
- Els originals rebuts, siguin articles siguin notes, se sotmetran anònimament a l'informe de dos especialistes externs designats pel Consell de Redacció, el qual es reserva el dret de publicació. Es comunicarà raonadament als autors l'acceptació o el rebuig del seus treballs en el termini màxim de sis mesos.
- *Instruccions als avaluadors*: s'avaluaran l'originalitat, el rigor acadèmic i la metodologia, la bibliografia i l'estil de l'article, abans de procedir a recomanar-ne o no la publicació o a sol·licitar-ne modificacions.

GUIDELINE FOR CONTRIBUTORS

- *Comprendre* is a scientific review which publishes two issues a year. It is opened to contributions on the classic fields of philosophy: metaphysics, epistemology, logical, ethics, philosophy of science, anthropology, history of philosophy, philosophy of religions, aesthetics, etc. It is addressed to an academic audience interested in the current philosophical and humanistic debates.
- *Comprendre* accepts three types of contributions: articles, critical notes and short reviews. Only original manuscripts not published previously and written in Catalan or in the main European languages (English, Spanish, French, German, Portuguese and Italian) will be considered for publication. Book reviews will be written only in Catalan.
- Contributions will be submitted electronically (comprendre@filosofia.url.edu) in a Word format file. Articles should not exceed 10000 words (including notes and bibliography). Critical notes should not exceed 6000 words. Short reviews should not exceed 2500 words.

Type of letter: *Times New Roman*, body 12, space 1,5.

- The title, an abstract, over ten lines, and the key words in both the original language and English must be added at the beginning of the contributions
- Citations in footnotes will be numbered continuously. Bibliographical references can be placed in a final Work Citations section, always in alphabetical order by authors.

The next guidelines should be followed in all the citations.

Type of letter: *Times New Roman*, body 10, space 1.

a) For Books:

INITIAL OF AUTHOR'S NAME. SURNAME (small capital), *Title* (italics). Place of edition: Publisher, year.

Example:

I. KANT, *Critique of pure reason*. Translated by Paul Guyer & Allan Wood. Cambridge: Cambridge University Press, 1998.

b) For Articles:

INITIAL OF AUTHOR'S NAME. SURNAME (small capital), «Title of the article». *Name of the periodical publication* (italics) [Place of edition], 000 (number), 0000 (year), pp. 00-00.

Example:

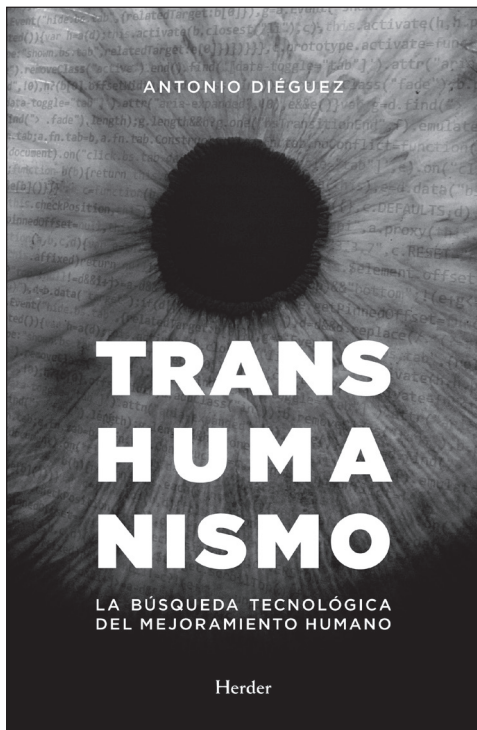
C. LLINÀS, «Gerhard Krüger: Einsicht und Leidenschaft (Intel·ligència i passió). Una entrada «platònica» en el pensament del segle XX». *Comprendre. Revista catalana de filosofia* [Barcelona], IX/1-2, 2007, pp. 159-189.

- The bibliographic references used should be added at the end of the article.
- Personal and Academic affiliation should be included in a cover sheet, containing an operative electronic address.
- *Comprendre* follows the Code of Conduct for Publication Ethics in the case of articles and critical reviews. Authors are required to attach a contributor's form with the manuscript. A model is available in its web.
- Contributions will be submitted to an external blind review process. The right of publication is reserved to the Editorial Board. The author will receive a response in six months. The acceptance or the refusal will be reasoned.
- *Guidelines for evaluators*: before being recommended to be published or not, or even to be modified, the articles will be evaluated according to the following items: originality, academic rigueur and methodology, bibliography and correct style.

Herder Editorial

TRANSHUMANISMO La búsqueda tecnológica del mejoramiento humano

Antonio Diéguez



248 págs.

14,1 x 21,6 cm

Rústica con solapas

ISBN: 978-84-254-3962-9

Precio c/ IVA: 19,80 €

El transhumanismo es uno de los movimientos filosóficos y culturales que más atención ha atraído en los últimos años. Preconiza el uso libre de la tecnología para el mejoramiento del ser humano, tanto en sus capacidades físicas, como en las mentales, emocionales y morales, trascendiendo todos sus límites actuales. Las tecnologías a las acude son la ingeniería genética y el desarrollo de máquinas inteligentes. Según los defensores del transhumanismo, con la ayuda de estas tecnologías podremos acabar con el sufrimiento, con las limitaciones biológicas que lo producen, e incluso podremos vencer al envejecimiento y la muerte.

Aunque muchos transhumanistas no ven deseable llevar estas mejoras hasta un punto en que el individuo mejorado ya no perteneciera a la especie humana, otros, designados como «posthumanistas», consideran que este es precisamente el objetivo final: la creación de una o varias especies nuevas a partir de la nuestra.

Las promesas que realizan los defensores del transhumanismo son muy ambiciosas, y no todas están justificadas. Pero por otro lado, la crítica de que modificar la naturaleza humana pone en peligro las bases de la vida moral, la dignidad y los derechos humanos, encierra supuestos filosóficos discutibles y sus consecuencias son excesivamente radicales. Hay otros enfoques que permiten hacer una evaluación más equilibrada, sobre la que podría edificarse en el futuro un mayor acuerdo social.



Herder Editorial

Herder Editorial S.L.
Provenza, 388
08025 Barcelona, España
Telf.: +34 934762626
www.herdereditorial.com

La societat pal·liativa

Byung-Chul Han



En l'actualitat vivim en una societat que ha desenvolupat una fòbia al dolor, a la qual ja no hi ha lloc per al sofriment. Aquesta por generalitzada es reflexa tant en l'àmbit personal com en el social i, fins i tot, a la política. L'imperatiu neoliberal «sigues feliç», que amaga una exigència de rendiment, intenta evitar qualsevol estat dolorós i ens empeny a un estat d'anestèsia permanent.

Com a *La societat del cansament*, Byung-Chul Han parteix del supòsit que en Occident s'ha produït un canvi radical de paradigma. Les societats premodernes tenien una relació molt íntima amb el dolor i la mort, que enfrontaven amb dignitat i resignació. Tanmateix, en l'actualitat, la positivitat de la felicitat desbanca a la negativitat del dolor, i s'estén a l'àmbit social. A l'expulsar de la vida pública els conflictes i les controvèrsies, s'instaura una post democràcia, que és en el fons una democràcia *pal·liativa*.

88 págs.

12,2 x 19,8 cm

Rústica con solapas

ISBN: 978-84-254-4634-4

Precio c/ IVA: 12,00 €



Herder Editorial

Herder Editorial S.L.

Provenza, 388

08025 Barcelona, España

Tel.: +34 934762626

www.herdereditorial.com