

# CARLOS ALCHOURRÓN Y LA INTELIGENCIA ARTIFICIAL

RAÚL CARNOTA

Maestría de Epistemología e Historia de la Ciencia de la Universidad Nacional de Tres de Febrero, Buenos Aires, Argentina

RICARDO RODRÍGUEZ

Departamento de Computación de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, Argentina

## Resumen

Las investigaciones que Carlos Alchourrón desarrolló en la Filosofía del Derecho se vincularon, desde inicios de la década de 1980, con problemáticas críticas de la Inteligencia Artificial (IA). Su contribución a la construcción de una lógica de las normas se conectó rápidamente con la deducción automática y los Sistemas Expertos Jurídicos. Su preocupación por la cuestión de los conflictos de obligaciones que pueden plantearse en un sistema normativo cuando el juez se enfrenta a la necesidad de emitir un veredicto considerando ciertos hechos, se vinculó, en sus trabajos, con la noción más general de “condicionales derrotables” y con la formalización del Razonamiento No Monótono. Pero su contribución de mayor impacto fue aquella derivada de la búsqueda de una solución al problema de los resultados múltiples que surgen de la derogación de una norma en un sistema jurídico, problema que puede ser visto, en un contexto epistémico, como el de la contracción de un conjunto de creencias. Su enfoque constructivo, que desarrolló con David Makinson, convergió con el axiomático elaborado por Peter Gärdenfors, dando lugar al surgimiento de la teoría de los Cambios Racionales de Creencias, más conocida como AGM, teoría que fue adoptada en IA como un modelo standard para la especificación de las actualizaciones de Bases de Conocimiento, operaciones que, habitualmente, involucran la introducción de una nueva creencia que puede ser inconsistente con las anteriores.

¿Por qué AGM adquirió rápidamente carta de ciudadanía en IA, al punto de que su influencia, medida según la cantidad de menciones bibliográficas del artículo fundacional de 1985, se ha duplicado en la última década? Mostraremos en este trabajo que la teoría apareció en un momento de crisis de la IA y que el alto nivel de abstracción del modelo de AGM era lo que muchos investigadores de la disciplina estaban buscando. El presente artículo, luego de describir la génesis de AGM, realiza un análisis histórico de las circunstancias en las cuales la teoría se introdujo en la IA y una evaluación cuantitativa y cualitativa de su impacto en dicha disciplina.

**PALABRAS CLAVE:** AGM - condicional derrotable - razonamiento no monótono - inteligencia artificial.

## Abstract

The main lines of research that Carlos Alchourrón developed in the Philosophy of Law, were connected, from the beginnings of the 80's, to some critical problems in the Arti-

ficial Intelligence (AI). His contributions to the construction of a logic of norms were rapidly linked to automatic deduction and the Legal Expert Systems. His concerns for the conflict of obligations that could arise in a normative system when a judge has to issue a ruling taking some particular facts into account were associated in his work with the more general notion of “defeasible conditionals” and with the formalization of Non Monotonic Reasoning. But his most relevant contribution was the one derived from the search of a solution to the problem of multiple outcomes resulting from the derogation of a norm in a legal corpus, which may be considered –within an epistemic framework– analogous to the contraction of a set of beliefs. The constructive approach he worked out with David Makinson converged with Peter Gärdenfors’ axiomatic, giving rise to the formulation of the theory of Rational Changes of Belief (AGM), which was adopted in AI as a model for the specification of the updates of Knowledge Bases as it usually involves a new belief that may be inconsistent with the old ones.

KEY WORDS: AGM - defeasible conditional - Non Monotonic Reasoning - Artificial Intelligence.

## 1- Introducción

Podría decirse que la Lógica Deóntica tuvo su origen como disciplina en el trabajo de Georg H. von Wright, *Deontic Logic*, de 1951.<sup>1</sup> O, en todo caso, esto es lo que hubiera dicho Carlos Alchourrón en sus apasionantes charlas. Von Wright propuso que a los conceptos deónticos de obligación, permisividad, prohibición e indiferencia les correspondía la misma interacción que existía entre los operadores modales de necesidad, posibilidad, imposibilidad y contingencia, con la particularidad de que en aquella se verificaba el principio de que obligación implica permisión (axioma deóntico). Así surgió la lógica SDL (*Standard Deontic Logic*) donde los operadores monádicos usuales de necesidad y posibilidad ( $\diamond$ ) de una lógica modal clásica son reemplazados por una ‘O’ y una ‘P’, por obligatorio y permitido, respectivamente. Eso dio lugar a una gran variedad de sistemas que, a fines de los años cincuenta y comienzos de los sesenta del siglo pasado, fueron beneficiados con la semántica de Kripke, que les proporcionó una nueva perspectiva. Paralelamente se introdujo la noción de obligación condicional como un operador diádico y surgieron los sistemas que von Wright recopiló en su trabajo de 1968.

Carlos Alchourrón conoció esta nueva rama de la lógica siendo todavía estudiante de Derecho, en la primera mitad de la década de 1950, a través del profesor Carlos Cossio, filósofo del Derecho que, en una fecha sorprendentemente temprana, había intuido su importancia para la teo-

<sup>1</sup> “Deontic Logic”, Georg H. von Wright, *Mind* (60), 1951. Reproducido en “Logical Studies”, Routledge and Keagan Paul, London, 1957.

ría jurídica. La fascinación de Alchourrón por este enfoque formal lo condujo hacia los estudios de lógica, disciplina en la que fue, esencialmente, un autodidacta, aun cuando, en su afán de lograr una buena formación, también tomó contacto con investigadores ligados a la Facultad de Ciencias Exactas de la Universidad de Buenos Aires (UBA).<sup>2</sup> De hecho sus primeros cargos docentes universitarios fueron en la materia Lógica y Metodología de las Ciencias, de los cursos de ingreso de Arquitectura y de Ciencias Exactas de la UBA en 1958 y 1959 y luego como profesor en Ciencias Exactas para el dictado de un curso de Lógica Modal en 1959. Desde aquellos primeros años se involucró en una de las principales controversias de la época en la Filosofía del Derecho: la cuestión acerca de si todos los conceptos deónticos debían estar basados en una noción de verdad. En este debate Alchourrón profundizó en la distinción entre las normas mismas, entendidas como actos de obligación, permisión, prohibición, etc., y lo que dio en llamar proposiciones normativas, que eran sentencias acerca de dichas normas en las que, por ejemplo, son declaradas como parte de cierto código legal o sistema moral. Las primeras tendrían una significación prescriptiva (y por lo tanto admiten condiciones de aceptabilidad) y las segundas una descriptiva (con lo cual son interpretables en términos de verdad o falsedad). Si bien la distinción ya existía, era tomada en forma vaga y sin extraer de ella consecuencias significativas. Alchourrón y su colega y amigo Eugenio Bulygin la tomaron en serio y esta línea de trabajo dio origen a diversas publicaciones propias de cada uno y en coautoría, algunas de las cuales se volverían referencias obligadas de cualquier investigación posterior acerca de sistemas normativos.<sup>3</sup> Von Wright caracterizó al libro *Normative Systems* de Alchourrón y Bulygin, afirmando que “*Con este libro, los dos autores pasaron, por así decirlo, al escenario internacional*”.<sup>4</sup>

La cuestión a dilucidar se planteaba en estos términos: si las normas no tienen valores de verdad, ¿cómo pueden darse entre ellas relaciones tales como contradicción o consecuencia lógica?, ¿es entonces posible una lógica de las normas, una lógica deóntica? Para von Wright, “*Alchourrón y Bulygin se cuentan entre los relativamente pocos que han desarro-*

<sup>2</sup> Eugenio Bulygin, en una comunicación personal a los autores.

<sup>3</sup> En particular: “Logique of Norms and Logique of Normative Propositions”, *Logique et Analyse*, 9/1969, París; “Normative Systems”, Springer Verlag, 1971 (en colaboración con E. Bulygin) y “The intuitive background of Normative Legal Discourse and its formalizations” editado por Mario Bunge en *Exact Philosophy: Problems, Tools and Goals*, 1973.

<sup>4</sup> G.H.von Wright, prólogo a *Análisis Lógico y Derecho* (recopilación de trabajos de Alchourrón y Bulygin), Centro de Estudios Constitucionales, Madrid, 1991.

llado una lógica deóntica tomando en cuenta la naturaleza problemática de sus fundamentos".<sup>5</sup> Años más tarde Alchourrón elaboraría una salida a este problema, expuesta en un trabajo conjunto con A. A. Martino de 1988, denominado "Lógica sin verdad".<sup>6</sup> En el mismo se sostiene que las normas no tienen valores de verdad pero que, a los efectos de la lógica, pueden ser tratadas como si los tuvieran y, en consecuencia, estar sujetas a relaciones por medio de las conectivas usuales dentro de una lógica que incluya la lógica clásica.

Ya a inicios de la década de 1970, Alchourrón y Bulygin comenzaron a ocuparse por comprender los efectos de la derogación de una norma en un sistema. Von Wright definió el problema en los siguientes términos: "Habitualmente los actos de derogación habían sido considerados como consistiendo simplemente en actos de eliminación o 'invalidación' de alguna norma dictada previamente. Pero no habían sido estudiadas sistemáticamente las implicaciones lógicas que tales actos pueden tener en el resto del sistema normativo. Interés particular tiene el caso en el que la derogación requiere el rechazo alternativo de alguna otra norma, a más de la derogada, y por lo tanto, conduce a una indeterminación en el sistema".<sup>7</sup> La cuestión había sido advertida poco antes por Cornides, pero sin que se le asignara mayor importancia.<sup>8</sup> Alchourrón y Bulygin presentaron el problema en su real dimensión en un trabajo expuesto en un Congreso de Lógica Deóntica y Semántica en 1975<sup>9</sup> y en otros artículos de esos años. Sin embargo el tratamiento formal de la cuestión de la derogación se desplegó, a partir de 1980, con los trabajos en colaboración entre Alchourrón y David Makinson, que, en 1985, confluyeron con Peter Gärdenfors en la formulación de la Lógica del Cambio de Teorías (AGM).

También fueron objeto de su curiosidad intelectual "Las dificultades que plantean las distintas formas de condicionales... (que) ... aumentan considerablemente cuando se trata de dar cuenta de las lógicas de las normas condicionales"<sup>10</sup> ya que "suelen omitirse aquellas condiciones nega-

<sup>5</sup> G. H. von Wright, *idem* anterior.

<sup>6</sup> "Lógica sin verdad", C. Alchourrón y A. A. Marino, *Theoria - Segunda época*, Año III, España, 1988; "Logic without Truth" *Ratio Juris*, Vol. 3, Nº 1, Bologna, 1990.

<sup>7</sup> G. H. von Wright, *idem* anterior.

<sup>8</sup> "Der Widerruf von Befehlen", Th. Cornides, *Studium Generale* 22, 1969.

<sup>9</sup> "Unvollständigkeit, Widersprüchlichkeit und Unbestimmtheit der Normenordnung", *Deontische Logik und Semantik*, Athenaiion, Wiesbaden, 1975. La versión italiana es "Incompletezza, contraddittorietà e indeterminazione degli ordinamenti normativi", *Logica, Deontica e Semantica*, Bologna, 1977.

<sup>10</sup> Carlos Alchourrón, *Síntesis de aportes originales*, presentación al concurso de Profesor de Lógica en la Facultad de Filosofía y Letras, UBA, 1985.

*tivas, cuya presencia impide el nacimiento de un derecho o de una obligación*".<sup>11</sup> Este interés se manifestó abiertamente en un trabajo publicado en 1986, "*Conditionality and the Representation of Legal Norms*"<sup>12</sup> y es el que lo ocupó más intensamente en los últimos años de su vida.

Todas estas actividades de indagación lógico-filosófica, en las que nunca abandonó como dominio pragmático el campo del derecho, junto a diversas circunstancias históricas, académicas y personales, se fueron entretejiendo progresivamente, desde inicios de la década de 1980, con las problemáticas que estaban a la orden del día en el campo de la Inteligencia Artificial (IA), disciplina en auge en esos años. Este fenómeno respondió, en gran medida, a la concurrencia de tres elementos: la pertinencia y generalidad de sus ideas, sus encuentros con otros investigadores que fueron sus socios intelectuales en distintos momentos y en distintas líneas de pensamiento y la existencia de una crisis de desarrollo de la IA, crisis que llevó a muchos investigadores a la búsqueda de modelos de un mayor rigor formal en otras disciplinas. En el resto de esta sección revisaremos más de cerca estas tres conexiones, para concentrarnos en el resto del artículo en el análisis del impacto en la IA de la teoría AGM.

En primer lugar, la cuestión de la distinción entre normas y proposiciones normativas y la construcción de una lógica de las normas, derivó a la posibilidad de la deducción en el terreno legal y de allí a la automatización de dicha deducción, en la línea de los Sistemas Expertos Legales. La conexión de Alchourrón con estos sistemas vino de la mano de Antonio Martino, profesor asociado en su cátedra en la UBA y luego exiliado en Italia a partir de 1976, donde llegó a ser director del *Istituto per la Documentazione Giuridica* del Consejo Nacional de Investigaciones de Italia. Martino estaba convencido de que "*la teoría normativa, creada por Von Wright y desarrollada por Alchourrón, Bulygin y otros autores, estaba suficientemente madura como para realizar un modelo computacional que permitiese la automatización del razonamiento jurídico*",<sup>13</sup> y lo convocó, desde inicios de los ochenta del siglo pasado, para ser consultor lógico de los proyectos, italianos primero y europeos después, orientados a la construcción de Sistemas Expertos Legales. En esos años, Alchourrón conoció el lenguaje PROLOG y a uno

<sup>11</sup> Carlos Alchourrón y Eugenio Bulygin, Introducción a *Análisis Lógico y Derecho*, Centro de Estudios Constitucionales, Madrid, 1991.

<sup>12</sup> Presentado en el Segundo Congreso Internacional "Logica, Informatica, Diritto", Firenze, Italia, 1985, publicada en *Automated analysis of Legal Texts*, North Holland, 1986.

<sup>13</sup> Antonio Martino, en una comunicación personal a los autores.

de sus primeros difusores en la IA, Robert Kowalski, quien había desarrollado, como aplicación modelo, un caso jurídico: la formalización del Acta de Nacionalidad Británica; con él discute sobre las limitaciones de sus inferencias lógicas. Entusiasmado por esta línea de trabajo y frente a la evidencia de la debilidad de la formación lógica de los profesionales con los que interactuaba en Italia y de la falta de un lenguaje común con ellos, decidió llevar adelante su propia experiencia en Buenos Aires, para lo cual presentó un proyecto de investigación en la Universidad de Buenos Aires (UBA), que se desarrolló entre 1987 y 1990, titulado *Investigación de Sistemas Expertos de Inteligencia Artificial*. Luego de algunas vicisitudes iniciales, este proyecto desembocó en la constitución de un auténtico grupo interdisciplinario, compuesto por lógicos, informáticos y abogados, y en el desarrollo de un razonador experto jurídico, que fue presentado en 1990 en un Congreso Internacional de Informática y Derecho realizado en Buenos Aires.<sup>14</sup> El proyecto de Sistemas Expertos fue continuado en los años siguientes, al tiempo que la impronta interdisciplinaria, de la cual Alchourrón era animador principal, se extendió también a proyectos en Revisión de Creencias y Razonamiento No Monótono.

Por otro lado, su inquietud por elucidar el sentido de la noción de derogación de una norma en un sistema se asoció a una dificultad más general, que en filosofía de la ciencia se presenta como el problema del cambio racional de una teoría científica, mientras que “*el mismo problema reaparece en los contextos epistémicos como el de la teoría del cambio racional de los estados de creencias*”.<sup>15</sup> Precisamente en esos años, los debates en la IA giraban en torno a la cuestión de la “*representación del conocimiento*” en un sistema inteligente (agente) y a la dinámica del cambio de dicho “*conocimiento*” cuando el agente interactúa con su entorno. Es en este contexto en el que se produce el fenómeno, que analizaremos en detalle en el resto del trabajo, por el cual AGM adquirió rápidamente carta de ciudadanía en la comunidad de IA.

Finalmente, los intentos de formalización de las obligaciones condicionales que lo llevaron, en un primer momento, a inclinarse, siguiendo a David Ross, por su interpretación como obligaciones *prima facie*, se conectaron con el llamado Razonamiento No Monótono (RNM) en IA. En

<sup>14</sup> El responsable directo del desarrollo del razonador fue Adolfo Kvitca y en su momento tuvo cierta repercusión en las Secciones de Ciencia y Técnica de los diarios de Buenos Aires.

<sup>15</sup> Carlos Alchourrón, *Síntesis de aportes originales*, presentación al concurso de Profesor de Lógica en la Facultad de Filosofía y Letras, UBA, 1985.

su artículo *Conditionality and the Representation of Legal Norms*, citado más arriba, se plantea la problemática de obligaciones en conflicto y es fácil observar que se trata del mismo patrón formal que en esa época motivaba los desarrollos en el campo del RNM, desarrollos que Alchourrón desconocía en ese momento.<sup>16</sup> En ese trabajo propuso utilizar una noción lógica clásica complementada por una operación de contracción, en la línea de sus trabajos con Makinson. Posteriormente, y en conocimiento de la existencia de las Lógicas No Monótonas en IA, mantuvo una posición poco afecta a dichos formalismos y sostuvo que muchos de los razonamientos que se deseaban reproducir y que eran inválidos desde el punto de vista de la lógica deductiva, podían ser manejados con esta lógica mediante la explicitación de premisas tácitas.<sup>17</sup> En los años siguientes investigó intensamente sobre el tema de la representación de los denominados '*condicionales derrotables*', trabajo que se reflejó en una serie de artículos<sup>18</sup> donde, en palabras de dos de sus discípulos, "*no sólo proveyó una elucidación filosófica de la noción de derrotabilidad, sino que mostró cómo aplicarla en la explicación de conceptos deónticos, tales como la noción de obligación 'prima facie'. De este modo llegó a una nueva semántica para los condicionales derrotables en términos de revisión*".<sup>19</sup> Como ya mencionamos, Alchourrón desconfiaba de la idea misma de una lógica no monótona y por eso no se involucró en las investigaciones acerca de las relaciones entre la lógica del cambio de teorías y las inferencias no monótonas, que Gärdenfors y Makinson desarrollaron en el nivel de

<sup>16</sup> La cuestión estaba ejemplificada por Alchourrón en el siguiente conjunto A de normas:

- (A1) Los jueces deben castigar a los que han cometido homicidio.
- (A2) Los jueces no deben castigar a los menores de edad.

La dificultad de la formalización de A reside en que si (A1) se representase con un condicional material, frente al caso de un asesino menor de edad se llegaría a resultados conflictivos. Es la misma situación del ya clásico "Los pájaros vuelan"; "Los pingüinos no vuelan"; "Pi-Pio es pingüino y pájaro".

<sup>17</sup> "Limits of Logic and Legal Reasoning", C. A. Alchourrón y E. Bulygin, *Preproceedings of the III Internacional Conferencie on Logica, Informatica, Diritto*, Vol. II, Firenze, 1989.

<sup>18</sup> Entre otros: "Philosophical Foundations of Deontic Logic and the Logic of Defeasible Conditionals", en J. J. Meyer and R. J. Wieringa (eds.): *Deontic Logic in Computer Science. Normative Systems Specifications*, Wiley and Sons, 1993; "Defeasible Conditionals as General Conditionals plus Revision Theory", *Actas del Workshop on Conditionals in Knowledge Representation de la IJCAI'93*; "Defeasible Logics: Demarcation and Affinities", G. Crocco, L. Fariñas del Cerro y A. Herzig (eds.) *Conditionals and Artificial Intelligence*, Oxford University Press, 1994.

<sup>19</sup> "DFT and Belief Revision", E. Fermé and R. Rodriguez, 2006, ver en esta misma edición.



la propia noción de consecuencia (en el metalenguaje).<sup>20</sup> La principal contribución de Alchourrón a la teoría de los condicionales derrotables fue la construcción del sistema modal DFT, que fue presentado en la comunidad de IA a través del *Workshop on Conditionals in Knowledge Representation* de la *International Joint Conference on Artificial Intelligence* desarrollada en Francia en 1993 (IJCAI'93), en el cual conjugó los conceptos de revisión y de condicional derrotable, usando una función de revisión en el nivel del lenguaje.<sup>21</sup>

En síntesis, en el terreno de la IA, las líneas de trabajo intelectual de Alchourrón que hemos descripto se vincularon a aspectos decisivos de la caracterización de un agente: las formas de representación y explotación de su Base de Conocimientos, la actualización de dicha Base en la medida en que el agente se conecta con su entorno, la necesidad del mantenimiento de la consistencia de lo que el agente "conoce" y la característica provisional de, al menos, una parte sustancial del contenido de esa Base de Conocimientos. De todas estas conexiones la que, sin duda alguna, alcanzó una repercusión más profunda y duradera es la que logró la teoría de cambio racional de creencias o AGM. El impacto de esta teoría, medido en términos de las citas de su artículo fundacional de 1985, se ha acrecentado notablemente desde mediados de la década de 1990, más que duplicando el logrado en la primera década desde su aparición, y este crecimiento se refleja, particularmente, en trabajos del área de Inteligencia Artificial.

El objetivo del presente trabajo es realizar un análisis de las circunstancias en las que se produjo el impacto de la teoría AGM en la IA, así como una evaluación cuantitativa y cualitativa de tal impacto. En la próxima sección se desarrolla un relato histórico general del surgimiento de la teoría. En la tercera sección se reconstruye el estado de crisis de la IA en los años 80 del siglo pasado, mostrando cómo una de las líneas de salida de dicha crisis encontró respuesta en la teoría de cambio racional de creencias y sus derivaciones. La cuarta sección profundiza en el impacto inicial, analizando en detalle los primeros trabajos que en la IA se refirieron y/o inspiraron en AGM y el modo en que cada autor conoció dicha teoría. La sección quinta considera el impacto posterior de AGM en

<sup>20</sup> "Relations between the logic of theory change and nonmonotonic logic", D. Makinson and P. Gardenfors en: A. Fuhrmann, M. Morreau (eds.), *The Logic of Theory Change*, LNAI 465, Springer Verlag, 1990 págs. 185-205. "Non Monotonic Inference Based on Expectation". P. Gardenfors and D. Makinson, *Artificial Intelligence* 65, 1994.

<sup>21</sup> "Defeasible Conditionals as General Conditionals plus Revision Theory", *Actas del Workshop on Conditionals in Knowledge Representation de la IJCAI'93*.



términos cuantitativos y cualitativos. Finalmente la última sección realiza una síntesis y extrae algunas conclusiones.

## 2- La génesis de la Teoría del Cambio Racional de Creencias

A finales de la década de 1960, David Makinson, que pasaba sus vacaciones en Argentina visitando a los parientes de su esposa, fue invitado a dar una conferencia acerca de lógica deóntica. Al finalizar la charla conoció a Alchourrón. Makinson se sorprendió al encontrar un interlocutor con ideas muy claras y comentarios penetrantes acerca de los problemas de las lógicas deónticas.<sup>22</sup> Desde sus primeros contactos, Alchourrón le mencionó sus inquietudes acerca del problema de cómo analizar la estructura lógica del proceso de derogación de una norma contenida en un código. La primera reacción de Makinson fue responder que, frente a la cuestión de la pluralidad de resultados, no había lugar para el análisis lógico.<sup>23</sup> Sin embargo, poco a poco, el tema se fue instalando y se abrió paso la convicción de que, aunque realmente la lógica no podía resolver entre los diferentes resultados de una contracción, podrían encontrarse algunos principios generales que todas las derogaciones deberían satisfacer y, en consecuencia, podrían existir maneras interesantes de generar formalmente la familia completa de posibles derogaciones. El primer intento apareció publicado en un artículo conjunto de 1981, denominado *Hierarchies of Regulations and their Logic*.<sup>24</sup> Si bien hoy se lo puede ver, en opinión de D. Makinson, como una dificultosa marcha hacia la formulación de ideas que luego se conocerían como ‘*maxi-choice*’ y ‘*safe contraction*’,<sup>25</sup> resulta interesante reseñarlo para advertir cómo se fueron generalizando las ideas en juego, a partir de una pura motivación jurídica y algunas intuiciones básicas. El escenario planteado en *Hierarchies...* era el de un juez o funcionario tratando de aplicar un cuerpo de normas o regulacio-

<sup>22</sup> “Carlos made some very penetrating remarks that made me feel quite ashamed: there was I, talking about deontic logic to people who I had assumed knew nothing about it, and this guy in front of me evidently had a clearer picture of what is involved than I did”. D. Makinson, en una comunicación personal a los autores.

<sup>23</sup> “I still remember my initial off-the-cuff response: the plurality is just an unfortunate fact of life, and logic cannot adjudicate between the different possibilities.” D. Makinson, en una comunicación personal a los autores.

<sup>24</sup> “Hierarchies of Regulations and their Logic”, pp 125-148 of the collection *New Studies in Deontic Logic*, ed. R. Hilpinen, Dordrecht, Reidel, 1981.

<sup>25</sup> “Groping painfully towards ideas that would later become formulated as maxi-choice contraction and safe contraction”, David Makinson, “In memoriam. Carlos Eduardo Alchourrón”, *Nordic Journal of Philosophical Logic*, Vol.1, N° 1, 1996, pág. 7.

nes, sobre el cual se había realizado una derogación de resultado múltiple, con el fin de llegar a un veredicto.<sup>26</sup> Dado un código A, el tratamiento formal suponía un orden parcial entre las normas de A, al que se consideraba un supuesto natural en el campo legal, y este orden se usaba para inducir un orden en  $Pot(A)$  (el conjunto formado por todos los subconjuntos de A). Los subconjuntos maximales de A que no implicaban la norma que se quería derogar eran denominados '*remainders*' y se demostraban condiciones sobre la jerarquía  $(A, \leq)$  con el fin de obtener un único '*remainder*' y, de este modo, un resultado único para la operación de derogación. Tan importante como el tratamiento de la derogación era el abordaje del caso en el que un código tenía contradicciones en vistas de ciertos hechos empíricos considerados verdaderos.<sup>27</sup> Los autores proponían una solución de este problema (al que denominaron '*delivery*') usando nuevamente una estructura de orden parcial entre las normas.<sup>28</sup> En este punto reconocían inspirarse en las discusiones de David Ross acerca de las obligaciones condicionales potencialmente conflictivas, que Ross interpretaba como obligaciones '*prima facie*'. Este tratamiento constituyó un antecedente de las inquietudes que Alchourrón desarrolló más adelante. El final del artículo sugería aplicaciones extra jurídicas para derogación y '*delivery*'.<sup>29</sup>

Durante el proceso de publicación de *Hierarchies...*, Alchourrón y Makinson tomaron mayor conciencia de que, tanto el problema como la línea de ataque del mismo no estaban en absoluto limitados al caso de las normas. El conjunto de partida A podía ser un conjunto arbitrario de sentencias y el problema se convertía, en ese caso, en el de eliminar un elemento del conjunto o una consecuencia indeseada de la teoría generada. Ya existían trabajos en el campo de la filosofía relativos a contracciones y revisiones de creencias y, entre el trabajo de 1981 y el artículo "On the

<sup>26</sup> El término "contracción" no aparece en el texto, sólo se usa "derogación".

<sup>27</sup> Sea A el conjunto de normas; sean B,  $C \subseteq A$  y sean F1 y F2 conjuntos de hechos. La situación planteada es que  $B \cup F1 \Rightarrow x$ ;  $C \cup F2 \Rightarrow \neg x$ .

<sup>28</sup> En el caso general de un código inconsistente A, la propuesta pasaba por realizar una derogación de A por  $x \wedge \neg x$ .

<sup>29</sup> Uno de estos ejemplos de posibles aplicaciones del delivery era el caso de un sistema de información computacional que, por algún "accidente", hubiera incorporado información inconsistente durante un proceso y se necesitase continuar operando en forma segura mientras se reparaba la causa del error. En esos años se desarrollaban los "TMS" en IA, pero los autores, lejos de sospechar sus futuros involucramientos, se estaban refiriendo a sistemas convencionales, lo que es evidente hasta por la idea de que al sistema del ejemplo se lo concebía como adquiriendo información inconsistente por un error que luego un equipo de soporte debería reparar.

Logic of Theory Change: Contraction functions and their associated Revision functions”, publicado en *Theoria* en 1982,<sup>30</sup> se produjo un cambio hacia un enfoque más general, que se reflejó, incluso, en el lenguaje utilizado: la derogación se generalizó como contracción. Alchourrón caracterizaría los aportes de este nuevo artículo del siguiente modo: “*En el mismo se analizan las consecuencias de varias formas de caracterizar la contracción. En la primera la contracción se identifica con la intersección de todas las subteorías maximales que no implican la proposición que se quiere eliminar. Luego se prueba un resultado paradójico que la descalfica. El segundo enfoque identifica la contracción a través de una función de selección que elige una de las subteorías maximales que no implican la proposición a eliminar. Pese a su apoyo intuitivo, la construcción tiene consecuencia paradójicas, a menos que la teoría con que se opera sea ella misma maximal*”.<sup>31</sup> Así es como hicieron su presentación pública la ‘*full meet contraction*’ y la ‘*maxichoice contraction*’ y también sus problemas.<sup>32</sup>

Cuando estaban elaborando “On the Logic...”, sus autores entraron en contacto con los trabajos de Gärdenfors y sus postulados de racionalidad para la revisión y la contracción y comprendieron que estaban trabajando en los mismos problemas formales.<sup>33</sup> Así fue como la elección de la revista *Theoria* como tribunal donde someter “On the Logic...” estuvo motivada por el hecho de que Gärdenfors era, en esa época, el editor de la misma. En la Introducción señalaban que en la lógica del cambio de teorías hay dos procesos principales, la contracción que, “*en el contexto deóntico de eliminar una norma de un código es conocida como derogación*”,<sup>34</sup> y la

<sup>30</sup> “On the Logic of Theory Change: Contraction functions and their associated Revision functions”, C. Alchourrón and D. Makinson, *Theoria, A Swedish Journal of Philosophy*, Vol. XLVIII, 1982, Part I, págs. 14-37.

<sup>31</sup> Carlos Alchourrón, presentación de sus temas de investigación contenida en la solicitud de inscripción al concurso para profesor titular de Lógica en la Facultad de Filosofía y Letras de la UBA, en 1985. La solicitud nos fue facilitada por Gladys Palau.

<sup>32</sup> En el caso de la *full meet*, la revisión que se deriva de la misma por la identidad de Levi produce como resultado sólo las consecuencias de la “nueva creencia”, perdiendo todo el background previo. En el caso de la *maxichoice*, si se contrae una teoría que no es completa, la revisión que se genera por Levi produce una teoría completa: el agente se torna omnisciente. Este resultado es antiintuitivo a menos que el conjunto de creencias de partida ya sea completo. Ambos problemas se solucionan con la ‘*partial meet contraction*’.

<sup>33</sup> Entre ellos “Conditionals and changes of belief”, *Acta Philosophica Fennica*, 30: 381-404, 1978. “Rules for rational changes of belief”; en *Philosophical essays dedicated to Lennart Aqvist on his fiftieth birthday*, Pauli, T., ed. Departamento de Filosofía, Universidad de Uppsala, 88-101, 1982.

<sup>34</sup> C. Alchourrón and D. Makinson, “On the Logic...”.

revisión (o enmienda en el caso deóntico) y que las investigaciones de estos procesos habían seguido hasta allí dos caminos principales. “*Uno, explorado por Gärdenfors en una serie de publicaciones, ... opera, esencialmente, a través de la formulación de un cierto número de postulados o condiciones... Otro enfoque, que subyace en el trabajo previo de los presentes autores, es buscar definiciones explícitas de funciones de contracción y revisión... y luego ver hasta dónde las funciones así definidas... resultan poseer las propiedades postuladas*”.<sup>35</sup> Consecuentes con este segundo camino, luego de la construcción de la ‘full meet’ y de la ‘maxi-choice’, el artículo incluía la demostración de que la ‘maxi-choice contraction’ cumplía con la mayoría de los postulados de Gärdenfors.

Gärdenfors quedó muy impresionado por los resultados que Alchourrón y Makinson habían alcanzado. Él había incursionado en la temática de la Revisión de Creencias en busca de una fundamentación de la lógica de condicionales que no tuviera compromisos ontológicos con los mundos posibles. Sus postulados o condiciones de racionalidad sintácticas para las operaciones de revisión fueron luego, con pequeñas modificaciones, los que formarían el núcleo de los postulados AGM. Pero hasta ese momento no había mostrado especial interés por la operación de contracción, ya que consideraba como operación primitiva a la revisión, ni tampoco poseía ninguna semántica de revisión. A partir de la interacción como trío, Gärdenfors adoptó la noción de contracción y el modelo para ambas en términos de subconjuntos maximales del conjunto inicial de creencias que no implican la sentencia que se busca abandonar. Luego de numerosos intercambios epistolares entre Buenos Aires, Beirut, Lundt y París (el e-mail no había irrumpido todavía en la vida cotidiana), se llegó al artículo fundacional de lo que luego se denominaría AGM.<sup>36</sup> Recurramos nuevamente a Alchourrón para ponderar los avances que implicó este trabajo: “*Las dificultades indicadas... (en referencia a las que aparecían en el trabajo previo)... llevaron a una generalización del marco conceptual. La contracción de una teoría se identifica allí con la intersección de una selección no vacía de subteorías maximales que no implican la proposición que se quiere eliminar. Para esta construcción se demuestra un teorema de representación para los postulados*

<sup>35</sup> C. Alchourrón and D. Makinson, “On the Logic...”. La mayor parte de las referencias bibliográficas de este artículo son trabajos de Gärdenfors, al que agradecen haberles facilitado el acceso a los mismos, incluyendo aquellos que aún estaban en proceso de tipeo.

<sup>36</sup> “On the Logic of Theory Change: Partial Meet Contractions and Revision Functions”, *Journal of Symbolic Logic*, Vol. 50, N° 2, June 1985, págs. 118-139.

*básicos de la teoría del cambio racional de creencias de Gärdenfors. Se investigan las consecuencias de introducir relaciones entre las subteorías de una teoría dada y se prueba, entre otras cosas, el teorema de representación para la totalidad de los axiomas de la teoría del cambio racional de creencias de Gärdenfors. De este modo coinciden dos enfoques que tienen justificaciones intuitivas independientes*".<sup>37</sup>

Alchourrón y Makinson consideraban como básica la operación de contracción (obteniendo revisión a través de la identidad de Levi) y Gärdenfors consideraba a la revisión como operación primitiva (definiendo contracción por la identidad de Harper). Finalmente, en el artículo de 1985, la discrepancia se resolvió por "dos votos contra uno".<sup>38</sup>

Posteriormente se iniciaron dos líneas de trabajo paralelas. Por un lado, Alchourrón y Makinson definieron la contracción llamada 'safe', por la que el primero tenía predilección, y demostraron diversas conexiones entre ella y la 'partial meet' sobre la que se había basado el trabajo del trío.<sup>39</sup> Alchourrón decía de esta nueva operación que "Resulta más intuitivo y, en cierto modo, más realista, pensar que los elementos que se preservan de una teoría cuando se pretende eliminar alguna de sus consecuencias se seleccionan comparando los elementos de la teoría, más que sus distintas subteorías. Este enfoque se expone en 'On the logic of theory change: safe contraction', donde se presenta una definición de contracción de una teoría a partir de una relación entre sus elementos. Se prueba que, bajo condiciones muy intuitivas de las propiedades de la relación ordenadora, la contracción satisface las condiciones del enfoque axiomático de Gärdenfors. En ese mismo trabajo y sobre la base de esa contracción se aborda el tema de la reiteración de cambios racionales... Por otra parte, no todas las contracciones definidas según el modelo de una relación entre subteorías es una contracción definida en base a una relación entre los elementos, aunque la conversa siempre es cierta. Sin embargo, si la teoría tiene un número finito de proposiciones no equivalentes, los dos enfoques se corresponden. Esto se muestra en 'Maps between some different kinds of contraction function: the finite case'".<sup>40</sup>

<sup>37</sup> Carlos Alchourrón, presentación de sus temas de investigación..., *idem* anterior.

<sup>38</sup> "Finally, more or less by a vote of two to one, we ended up by taking contraction as basic and spending a lot of time on its properties.", D. Makinson, en una comunicación personal a los autores.

<sup>39</sup> Ambos temas se tratan, respectivamente, en "On the logic of theory change: safe contraction", C. Alchourrón and D. Makinson, *Studia Logica* 44 (1985) 405-422, y "Maps between some different kinds of contraction function: the finite case", C. Alchourrón y D. Makinson, *Studia Logica* 45 (1986) 187-198.

<sup>40</sup> Carlos Alchourrón, presentación de sus temas de investigación... *idem* anterior.

Curiosamente, de las cinco caracterizaciones que posee AGM, la ‘safe contraction’ es la que ha tenido menos repercusiones en el área de IA, aun cuando sus motivaciones pretendían dar un modelo implementable. Una posible explicación puede residir en el hecho de que fue publicada en *Studia Logica*, un medio de escaso impacto en la comunidad de IA<sup>41</sup> y, además, en una fecha en la que todavía no era conocida la teoría AGM en dicha comunidad.

La otra línea de trabajo fue desplegada por Gärdenfors y Makinson y consistió en definir funciones de revisión en términos de un orden epistémico de las sentencias de una Base de Conocimientos, al que dieron en llamar “*Epistemic Entrenchment*”. Esta última es la línea que, como veremos en las próximas secciones, comenzó a hacer popular a la teoría AGM en el campo de la IA, a partir de su presentación en la conferencia TARK’88.

### 3. La situación de la Inteligencia Artificial en la década de 1980

En esta sección trataremos de reflejar el estado en el que se encontraba la investigación en IA al momento del surgimiento de AGM. Como lo mostraremos a través de los discursos de importantes referentes del área, la IA, a fines de la década de 1970 y comienzos de la siguiente, estaba pasando por una crisis y la modelización alcanzada en AGM estaba en sintonía con una de las líneas de pensamiento que se iban consolidando con vistas a la superación de tal crisis. Trataremos de abonar la hipótesis de que la rápida acogida que recibió la Lógica del Cambio de Teorías en un sector importante de la comunidad de IA se debió, en buena parte, a que ésta estaba madura y ávida para aceptarla.

Desde su nacimiento la IA despertó muchas expectativas sobre lo revolucionario de sus resultados. Sin embargo, tras veinte años de paciente espera y grandes inversiones los logros alcanzados eran escasos y muy lejos estaban de colmar las fantasías iniciales. Este desencuentro entre logros y promesas llevó a diversos replanteos y debates internos en la disciplina.

Alan Newell, uno de los fundadores de la IA,<sup>42</sup> en su *Presidencial Address* pronunciado en la AAAI ‘80,<sup>43</sup> denominado *The Knowledge*

<sup>41</sup> “This approach has never had much echo among computer scientists –perhaps because of the place where it was published– but I have always had a particular affection for it. I think that Carlos quite liked it too, although I would describe myself as the father and he as an uncle”, D. Makinson, en una comunicación personal a los autores.

<sup>42</sup> Hay consenso en ubicar la fundación de la IA en la Conferencia de Dartmouth en 1956.

<sup>43</sup> Conferencia de la American Association for Artificial Intelligence, (Stanford, 19/8/80).

*Level*,<sup>44</sup> se propuso abordar lo que consideraba una situación de crisis que, a su juicio, se manifestaba por medio de tres indicadores.

Un primer indicador era la continua asignación de un rol casi mágico a la representación del conocimiento. Era un cliché de la AI, afirmaba Newell, considerar a la representación como el problema real a enfrentar, el *locus* de la verdadera inteligencia del sistema.

Un segundo indicador era el residuo de la gran controversia sobre la prueba de teoremas y el rol de la lógica que se había desarrollado entre finales de los años sesenta e inicios de los setenta del siglo pasado. Los primeros trabajos en prueba de teoremas para lógicas con cuantificación culminaron en 1965 con el desarrollo por parte de Alan Robinson de una formulación de la lógica de primer orden orientada al procesamiento mecánico, denominada resolución. Siguió un período intenso de exploración en sistemas probadores basados en resolución. La idea básica era que se tenía un motor de razonamiento de propósito general a mano y que hacer lógica, y hacerla bien, era la piedra fundacional de la acción inteligente. En unos cuantos años se vio que esto no era así: ese motor no era suficientemente poderoso ni para probar teoremas que fueran dificultosos a escala del cerebro humano, ni para resolver tareas como el planeamiento de robots. Entonces se produjo una reacción cuyo slogan fue “los procedimientos uniformes no funcionan” de la cual surgieron como resultado positivo toda una nueva generación de lenguajes de programación para IA. El residuo negativo de esta reacción fue la súbita “mala prensa” de la lógica como herramienta de la IA: la lógica era estática, no admitía mecanismos de control de las inferencias, el fracaso de la técnica de prueba de teoremas por resolución involucraba a la lógica en general, etc.

El tercer indicador fueron los resultados de una encuesta promovida en 1979/80 por Brachman & Smith entre investigadores involucrados en distintas áreas y proyectos de IA, así como con distintas posiciones o actitudes respecto de las cuestiones críticas de la representación del conocimiento.<sup>45</sup> El resultado principal del análisis de las respuestas fue una verdadera jungla de opiniones, sin consenso en ninguna cuestión sustancial. Como dijo uno de los encuestados: “*The standard practice of representation of knowledge is the scandal of AI*”.

El propósito declarado de la disertación de Newell era aportar algo de claridad al área, habida cuenta de que las investigaciones en Represen-

<sup>44</sup> “The Knowledge Level”, Alan Newell, *The AI Magazine*, 2(2), págs. 1-20, 1981.

<sup>45</sup> “Special Issue on Knowledge Representation”, *SIGART Newsletter*, Feb. 1980 (70).



tación del Conocimiento y Razonamiento (KRR según su sigla en inglés) deberían ser una prioridad en la agenda de la disciplina.

El foco de su inquietud inicial era la cuestión de qué es conocimiento, cómo se relaciona con la representación y qué es lo que un sistema tiene cuando se dice de él que está basado en conocimiento.<sup>46</sup>

Newell postuló, en esa oportunidad, la existencia de algo llamado “Nivel del Conocimiento” (KL), dentro del cual el conocimiento debía ser declarado.<sup>47</sup>

En el nuevo nivel propuesto, el nivel del conocimiento, hay un agente (el sistema) que procesa su conocimiento para determinar las acciones a tomar, dentro de un repertorio de acciones posibles, con el objeto de cumplir sus objetivos. Su ley de conducta está definida por el principio de racionalidad, por el cual las acciones seleccionadas deben ser las que mejor aproximen al agente al cumplimiento de sus objetivos, dado el conocimiento existente. Por la autonomía definicional de cada nivel, la especificación del nivel de conocimiento reside completamente en el contenido del conocimiento y en los objetivos del agente y no en la forma en que dichos conocimientos se estructuran en el nivel inferior. Newell decía que las representaciones existen en el nivel simbólico (SL), que es el inferior inmediato y consisten en estructuras de datos y procesos que concretan o realizan el cuerpo de conocimientos que el agente posee en el KL. En definitiva, todas las formas de representación (lógicas, figuras, planos, modelos, escenas, textos) podrían ser adecuadas, en la medida en que existiesen procesos simbólicos eficientes

<sup>46</sup> En los trabajos de la época, conocimiento y creencia son términos que se usan en forma indiferenciada, incluso cuando los autores reconocen su diferencia filosófica. En este contexto de IA, “conocimiento” sería todo lo que se supone que está representado en la estructura de datos del sistema.

<sup>47</sup> Este “nivel” refiere a una noción previa que es la de niveles o capas en un sistema de computación. En el nivel más bajo aparece el nivel de los dispositivos, luego el de los circuitos y así sucesivamente hasta el nivel de programas o nivel simbólico (SL). Cada nivel se define en dos formas: en primer lugar en forma autónoma, sin referencia a otro nivel. Así, por ejemplo, un programador de una aplicación comercial no necesita saber de circuitos lógicos o circuitos eléctricos y un gerente que define formalmente sus requerimientos para dicha aplicación, no necesita conocer lenguajes de programación. En segundo lugar cada nivel puede ser reducido al nivel inferior. Ambas definiciones son fundamentales y deben existir y ser, en algún modo, consistentes. Cada nivel no es una “abstracción” de los inferiores ni un simple “punto de vista”. Tiene existencia real incluso independiente de las muchas posibles formas en que está realizado o soportado en el nivel inferior. Cada nivel o capa es una especialización de la clase de sistemas capaces de ser descriptos en el nivel siguiente. En consecuencia, está a priori abierto cuál es la realización física de un nivel en el inferior.

que extraigan conocimiento de ellas. Estas realizaciones físicas no necesitarían ser conocidas para predecir la conducta del agente.<sup>48</sup> El nivel del conocimiento, afirmaba Newell, “*permite predecir y comprender la conducta del agente sin poseer un modelo operacional del proceso que realmente es llevado a cabo por él*”.<sup>49</sup> Su utilidad se derivaba de la posibilidad, para un observador, de predecir la conducta de un agente y, para un diseñador, de diseñar un sistema de IA sin que los mecanismos internos estén aún especificados.

Algunas de las principales conclusiones que extraía Newell eran las siguientes.

- Que el concepto de representación existe en el SL y que, en ese sentido, la caracterización abstracta del conocimiento que debería poseer el agente (en el KL) debería funcionar como especificación de la estructura simbólica.

- Que, en ese contexto, la lógica puede verse de dos modos. Por un lado, considerada como una clase de representación entre otras, tiene tanto elementos positivos (sus estructuras de representación y procesos de extracción son bien conocidos), como problemas (el de ineficiencia en primer lugar, referido en este caso al costo del proceso simbólico para extraer conocimiento de esa particular estructura). Por otro lado, la lógica es la herramienta apropiada para el análisis, en el nivel del conocimiento, de qué es lo que sabe un agente. De hecho, dada una representación en el SL –red semántica, tablero de ajedrez, grafo o cualquier otra– si se quiere determinar exactamente qué conocimiento existe en esa representación y caracterizarlo, se requiere del uso de la lógica.

- Que, en consecuencia, las limitaciones al uso de la lógica en tanto representación no afectan el rol de la lógica como una herramienta fundamental de análisis en el KL.

- Que, al distinguir fuertemente el KL del SL, estamos haciendo una igualmente fuerte separación entre conocimiento necesario para resolver un problema y el procesamiento requerido para extraer y explotar ese conocimiento en espacio y tiempo reales.

<sup>48</sup> Si un sistema tiene una estructura de datos de la que puede decirse que representa algo (objeto, procedimiento o lo que sea) y puede usarla, por medio de ciertos componentes que interpretan la estructura, entonces del sistema mismo se dice que tiene conocimiento, el conocimiento embutido en esa representación de esa cosa. Cuando se dice que “el programa conoce K” lo que queremos significar es que hay cierta estructura en el programa a la que “vemos” como soportando K y que, además, selecciona acciones exactamente como esperaríamos que haga un agente que “conoce K”, de acuerdo al principio de racionalidad, es decir, las más adecuadas para alcanzar sus objetivos.

<sup>49</sup> A. Newell, “The Knowledge...”.

- Que, desde la perspectiva del KL, cualquiera sea la estructura S que soporte el conocimiento del agente, un observador externo atribuirá al agente todo lo que el observador pueda conocer a partir de S.<sup>50</sup>

Esta preocupación por establecer bases firmes para los desarrollos de IA estuvo presente durante los años que siguieron a la presentación de Newell, en especial en los trabajos de Brachman, Levesque, Moore, Halpern, Moses, Vardi, Fagin y otros.

En "Logic and the complexity of Reasoning",<sup>51</sup> Levesque se pregunta "¿Qué tiene que ver la lógica con los enfoques computacionales del estudio de la cognición?" y, reconociendo que la prueba de teoremas parece ser computacionalmente intratable, propone no clausurar la cuestión en este punto sino virar hacia una visión menos idealizada de la lógica y analizar la complejidad computacional de la lógica aplicada a la cognición, sobre la base de que hay tareas computacionales más fáciles y otras más dificultosas. Frente a las objeciones del tipo "una actividad cognitiva realista es mucho más compleja que cualquier tipo de prolijo análisis matemático a priori", Levesque sostiene que un modelo sólo tiene interés si sirve para explicar la conducta que se quiere modelar y que incluso si dicha conducta es desordenada o confusa, el modelo no debe serlo. El modelo puede ser más o menos idealizado o realista, pero esto no altera "el duro hecho de que un modelo que tiene la conducta errónea o la conducta correcta por razones misteriosas, no tiene poder explicativo".<sup>52</sup>

Cabe destacar que durante los primeros años de la década de 1980 el debate fue profundo, generalizado y controvertido.<sup>53</sup> El mismo se reflejaba y replicaba en ediciones especiales de las principales publicaciones y en sesiones especiales de cada congreso o reunión científica del área. Por otra parte, la problemática de la IA atraía a muchos practicantes de otras disciplinas, que encontraban un amplio campo para la experimentación de sus propios conceptos e ideas.<sup>54</sup>

<sup>50</sup> A priori podría ser cualquier consecuencia de S.

<sup>51</sup> "Logic and the complexity of Reasoning", H. J. Levesque, *Journal of Philosophical Logic* 17 (1988) 355-389.

<sup>52</sup> H. J. Levesque, "Logic and the...".

<sup>53</sup> Algunos de estos debates dividieron a simbolistas y a quienes buscaban simular los mecanismos neuronales. Ver por ejemplo "El nuevo debate sobre la inteligencia artificial. Sistemas simbólicos y redes neuronales", Stephen R. Graubard (comp.), Gedisa, Barcelona, 1993; versión original: "The Artificial Intelligence debate: False Starts and Real Foundations", MIT Press, 1988.

<sup>54</sup> Probablemente esta atracción se veía reforzada por el despliegue, durante esos años, del denominado 'paradigma informacional'. Ver Manuel Castels, *La Sociedad Red*.

En 1984, lo que pretendía ser un pequeño seminario de investigación interdisciplinaria sobre aspectos teóricos del razonamiento sobre el conocimiento, organizado en el IBM San Jose Research Laboratory, se vio desbordado con reuniones de un promedio de cuarenta asistentes y una lista de correo de unos 250 nombres. Moshe Vardi, uno de los organizadores, señalaba que los concurrentes “*incluyeron científicos de la computación, matemáticos, filósofos y lingüistas*” y que “*dado el evidente interés en el área por parte de grupos tan diversos parecía apropiada una conferencia, particularmente una que pudiese incrementar el conocimiento por parte de los practicantes de un campo respecto al trabajo desarrollado en los otros campos*”.<sup>55</sup> La Primera Conferencia on *Theoretical Aspects of Reasoning about Knowledge* (TARK) tuvo lugar en marzo de 1986, con una concurrencia restringida y la intención de estimular la continua interacción entre los participantes. Vardi afirmaba que “*el sentimiento general al final de la reunión fue que el formato interdisciplinario de la conferencia había demostrado ser muy exitoso*”.<sup>56</sup>

El artículo introductorio de Halpern a los Proceedings de esta primera TARK fue un panorama sobre el área de *Reasoning about Knowledge*.<sup>57</sup>

Su propósito fue “trazar un panorama de las cuestiones centrales en la investigación del razonar sobre conocimiento y que son comunes a la filosofía, economía, lingüística así como a la inteligencia artificial y a la teoría de la computación”.<sup>58</sup> Presentó allí el que denominó “modelo clásico” de conocimiento y creencias, al que se aproximaron los investigadores de IA desde el inicio de la disciplina: el modelo de mundos posibles (el agente “conoce” aquello que es verdadero en todos los mundos que él “piensa como posibles”).<sup>59</sup> Habitualmente se trataba de propuestas cuyo soporte eran las lógicas modales (A3, A4 y A5, según el caso) y de allí las abundantes referencias en los trabajos de la época a Hintikka y a Kripke y, en menor grado a D.Lewis, Anderson y Belnap.

<sup>55</sup> Prefacio a la edición de los *Proceedings of the Second Conference TARK (Theoretical Aspects of Reasoning about Knowledge)*, M. Vardi, 1988.

<sup>56</sup> M. Vardi, *idem* anterior.

<sup>57</sup> “Reasoning about Knowledge: an Overview”, J. Halpern, *Proceedings of the First Conference TARK* (1986).

<sup>58</sup> J. Halpern, “Reasoning about...”

<sup>59</sup> Por ejemplo, en el caso de un sistema distribuido se le podía atribuir externamente conocimiento así: un procesador X “conoce” A si en todos los estados globales en los que X puede encontrarse, A es verdadero.

Halpern destacaba algunas limitaciones de estos modelos, como las dificultades de cálculo y la omnisciencia lógica. En particular se centraba en lo que consideraba como “*la aplicación más interesante que motiva el estudio del conocimiento en IA: cómo comprender cuál es el conocimiento necesario para una acción y cómo dicho conocimiento puede ser adquirido mediante procesos de comunicación*”. Y luego afirmaba que “*La mayor parte de los trabajos ... suponen, implícita o explícitamente que los mensajes recibidos (por el agente) son consistentes. La situación se hace mucho más complicada si los mensajes pueden ser inconsistentes. Esto nos lleva rápidamente a todo el conjunto complejo de cuestiones involucradas en revisión de creencias y razonamiento en presencia de inconsistencia. **Pese a que no intentaré abrir esa lata de gusanos aquí,** éstas son cuestiones que en algún momento deben ser consideradas al diseñar una base de conocimientos, por ejemplo, pues siempre está presente la posibilidad de adquirir del usuario información inconsistente*”.<sup>60</sup>

El éxito de la primera TARK llevó a la convocatoria de una segunda, la TARK'88, a la que se presentaron 108 trabajos, de los cuales se seleccionaron 22, tanto por su aporte original como por su interés para una audiencia interdisciplinaria. Uno de los trabajos seleccionados fue *Revision of Knowledge Systems using Epistemic Entrenchment* de Gardenfors y Makinson.<sup>61</sup> Esta fue la ‘*presentación oficial*’ de AGM en la comunidad de IA.<sup>62</sup> Luego de la TARK'88 un número creciente de investigadores, muchos de ellos de gran ascendiente en el área de IA, comenzaron a interesarse por la Lógica del Cambio Racional de Teorías.

Lo expuesto hasta aquí pretende mostrar el estado “convulsivo” en el que se encontraba el área de IA durante los inicios de la década de 1980. La toma de conciencia de la fragilidad de las propuestas existentes y de sus eventuales implementaciones influyó en la adopción de una actitud especialmente receptiva a la incorporación de aportes de otras disciplinas cognitivas. En este contexto arribaron a la IA las ideas de AGM y produjeron un impacto que se prolonga hasta nuestros días.

<sup>60</sup> J. Halpern, “Reasoning about...”, el destacado es nuestro.

<sup>61</sup> “Revision of Knowledge Systems using Epistemic Entrenchment”, Gardenfors y Makinson, *Proc. of the Second Conference TARK*, 1988.

<sup>62</sup> De las restantes ponencias de la conferencia, sólo una citaba la teoría AGM: la de Cristina Bicchieri, investigadora en temas económicos y de teoría de los juegos, quien en una comunicación personal a los autores, comentó que la pertinencia de AGM para su trabajo le había sido señalada por I. Levi. Su trabajo era “Common Knowledge and Backward Induction: A Solution to the Paradox”, C. Bicchieri, *Proc. of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, 1988.

#### 4. Primeras repercusiones de la Teoría de Cambio en la Inteligencia Artificial

La primera referencia a AGM en autores del campo de la IA fue publicada en 1986. El artículo<sup>63</sup> apareció en una revista de mucha influencia y tuvo como objetivo divulgar un nuevo enfoque para el problema del “*Knowledge Representation and Reasoning*”, enfoque que se venía consolidando en respuesta a la crisis puesta en evidencia por Newell. Su autor era H. Levesque, un influyente investigador a quien ya hemos mencionado en la sección previa. En el contexto de un recorrido por las cuestiones claves para la IA, en un breve párrafo dedicado al tema denominado por entonces “mantenimiento de la verdad”,<sup>64</sup> apareció una mención genérica a los trabajos en revisión de creencias dentro de la filosofía, ilustrada por dos referencias bibliográficas finales, sin comentario alguno. Una de ellas era el artículo fundacional de AGM de 1985. Si bien el *paper* de Levesque tuvo mucha trascendencia, la aparición de AGM como una mera cita restó impacto a su mención, apareciendo así, en la perspectiva histórica, como un hecho aislado respecto al florecimiento de trabajos que comenzó en 1988.

Varios de los investigadores encuestados para este trabajo coinciden en señalar la TARK'88 como el punto de entrada de la teoría AGM en la IA, a partir de la presentación del trabajo de Gardenfors y Mankinson y de la presencia en la Conferencia de figuras relevantes de la IA, como Ray Reiter, Jon Doyle, J. Halpern, H. Levesque, R. Moore, M. Vardi y R. Fagin, entre otros. De hecho hay un punto de inflexión en 1988, ya que en ese año y el siguiente aparecieron los trabajos pioneros que usaban o referenciaban en forma activa a la teoría. Sin embargo este evento por sí sólo no explica la adopción del modelo de formalización AGM por numerosos investigadores de la IA. Hay que agregar que, como parte de la crisis ya reseñada y del interés que suscitaban los temas abordados por la IA, se produjo una importante intervención de investigadores de la filosofía y la lógica en el área y la apertura de revistas de aquellos campos a investigadores de esta última y viceversa. Una consecuencia de esta interacción fue la propia existencia de las TARK. Por otra parte, como surge de los trabajos comentados en la sección anterior, la cuestión de la revisión de creencias y de la actualización de Bases de Conocimiento con

<sup>63</sup> “Knowledge Representation and Reasoning”, H. J. Levesque, *Annual Review of Computer Science*, 1986.

<sup>64</sup> Que se identificaba en IA con sistemas del tipo del “Truth Maintenance Systems” (TMS) de Jon Doyle y una familia de sistemas derivados.

mantenimiento de la verdad (o de la consistencia para ser más precisos) estaba a la orden del día en IA. La dificultad estribaba en combinar una semántica precisa y clara (en el “nivel del conocimiento” parafraseando a Newell) con un esquema “realista” del punto de vista computacional (en el “nivel simbólico”).

El primer trabajo que consideró con cierto detalle al formalismo AGM fue “Investigations into a Theory of Knowledge Base Revision. Preliminary Report” de Mukesh Dalal presentado en la AAAI '88.<sup>65</sup> Decía Dalal: “*En el corazón de muchas aplicaciones de AI de los últimos años hay una KB... que necesita ser revisada cuando se obtiene nueva información. En muchos casos esta revisión contradice conocimiento previo por lo que ciertas creencias deben ser abandonadas para mantener consistencia*”.<sup>66</sup> Haciendo referencia a un trabajo de Levesque<sup>67</sup>, que, en la línea de Newell, definía operaciones sobre la KB en el KL, señalaba Dalal que la función de Levesque

Tell:  $KB \times L \rightarrow KB$ ,

donde L es algún lenguaje formal, función que agregaba conocimiento a la KB, sólo estaba definida si la nueva información era consistente con la KB.

Con el fin de superar esta limitación, Dalal definía

Revise:  $KB \times L \rightarrow KB$

Para garantizar que la caracterización de Revise estuviera en el KL, Dalal definía la revisión puramente en términos de los modelos de la KB. A su vez proporcionaba “*una descripción equivalente en el nivel simbólico, presentando un método sintáctico para revisar bases de conocimiento*”.<sup>68</sup> Al final de la Introducción agregaba: “*Mostraremos la relación de este trabajo con la investigación en filosofía sobre los aspectos formales del cambio de teorías (...) que ha atraído recientemente la atención de la comunidad de AI. Para cualquier esquema de revisión, es deseable que éste preserve el máximo posible de las creencias sostenidas antes*

<sup>65</sup> “Investigations into a Theory of Knowledge Base Revision. Preliminary Report”, Mukesh Dalal, *Proceedings del Congreso de la American Association for Artificial Intelligence*, 1988.

<sup>66</sup> M. Dalal, “Investigations into...”.

<sup>67</sup> “Foundations of a functional approach to Knowledge Representation”, H. J. Levesque, *Artificial Intelligence* 23 (1984), 155-212.

<sup>68</sup> M. Dalal, “Investigations...”.



*de la revisión*".<sup>69</sup> Dalal, que desarrolló su trabajo sin conocer previamente el modelo AGM, propuso una serie de principios, no formalizados, que tomó como guías intuitivas para definir una revisión. Ellos eran: adecuación de la representación (el conocimiento revisado debe tener la misma representación que el viejo para poder iterar); irrelevancia de la sintaxis (que el resultado de la revisión no dependa de la representación sintáctica del viejo o del nuevo conocimiento), principio fundamental en el caso computacional y que era un punto débil de diversos métodos de actualización propuestos previamente dentro del campo de IA; mantenimiento de consistencia; primacía de la nueva información; persistencia del conocimiento previo (asimilable a mínima pérdida) y *fairness* (no arbitrariedad de la elección entre los múltiples resultados de la revisión). La representación abstracta del contenido de una KB era un conjunto finito de fórmulas en un lenguaje proposicional. Dalal generó una medida cuantitativa de los cambios sobre un conjunto de interpretaciones y la aplicó a definir mínimo cambio sobre los modelos de la KB. Construyó luego un algoritmo de cálculo de la fórmula del lenguaje correspondiente a cada conjunto de interpretaciones, de modo de poder definir sintácticamente, a partir de la fórmula representativa de la KB original, la fórmula que surgiera como resultado de la revisión. Así resultaba un método que no requería pasar explícitamente por una construcción de modelos. Dalal reconocía que cada paso del algoritmo propuesto requería un chequeo de consistencia, lo que lo hacía, en el caso general, NP-completo, y por eso era que consideraba a su investigación como preliminar. En una sección del trabajo en que realizaba la comparación de su método con otras propuestas, el primer apartado estaba dedicado a AGM. Con el objeto de demostrar los postulados AGM para su esquema de revisión, Dalal consideró un conjunto de creencias constituido por la clausura lógica de las fórmulas que componían la KB ("*como lo sugiere el enfoque del KL*", afirmaba).<sup>70</sup> Bajo ese supuesto la revisión de Dalal cumplía todos los postulados de AGM. También se discutían en dicha sección las diversas propuestas previas presentadas en IA para la actualización de Bases de Datos Lógicas por Ullman, Fagin, Vardi, Winslett, Weber y Borgida. De su análisis surgía que, en general, estas propues-

<sup>69</sup> Dalal, "Investigations...".

<sup>70</sup> Dalal, "Investigations...". El sentido de esta expresión de Dalal ya está sugerido en la sección 3, en la última de las conclusiones que le asignamos al trabajo de Newell. Para un observador externo, las creencias del agente son, en principio, cualquiera de las consecuencias de su Base. Aunque no sea realista considerar que las 'conoce' todas, sin embargo es esperable que alcanzara cualquiera de ellas, a través de una búsqueda orientada por sus 'goals' y de acuerdo al principio de racionalidad.

tas no respetaban la irrelevancia de la sintaxis y tampoco el principio de mínimo cambio.

El trabajo de Dalal, que estaba motivado directamente por las necesidades de la IA e intentaba encuadrarse en el replanteo formal de la disciplina, que pretendía forjar modelos abstractos que pudieran analizarse en el nivel del conocimiento, es el primero que estableció un vínculo claro con la teoría AGM.

¿Cómo se introdujo esta teoría en su artículo? De acuerdo al testimonio actual de Dalal, fue un referee anónimo que evaluó su *paper* el que le indicó la pertinencia de AGM y Alex Borgida, su orientador, lo estimuló a explorar la conexión.

*"One of the anonymous reviewers (assigned by AAI program committee) suggested the connection with AGM's work. Prof. Alex Borgida, who was initially a co-author of my paper, advised me to probe this connection deeper. We both liked the formal foundations (especially of the principles) provided by AGM and wanted to bring that into the AI work on belief revision".<sup>71</sup>*

Por su parte Alex Borgida afirma haber conocido AGM a través de su amigo, David Israel, un filósofo volcado al mundo de la IA.

*"I seem to recall that the connection to AGM was pointed out to me by a friend at SRI (Stanford Research Institute), Dr. David Israel –a famous philosopher turned AI researcher–. He may have been one of the AAI reviewers, too. Mukesh's paper was actually about a model theoretic description of propositional belief updates (essentially, minimal model mutilations), which followed my ideas for minimal mutilations for exceptions to integrity constraints. In this context, AGM made perfect sense, but we did not know the philosophy literature, and Israel did... As usual, it was a matter of being in the right place with the right knowledge".<sup>72</sup>*

Por su parte D. Israel parece haber conocido AGM a través de la literatura filosófica, y lo vinculó de inmediato a las búsquedas en IA alrededor de las Lógicas No Monótonas.<sup>73</sup>

<sup>71</sup> Mukesh Dalal, en una comunicación personal a los autores.

<sup>72</sup> Alex Borgida, en una comunicación personal a los autores.

<sup>73</sup> "I certainly did, quite early on –though never in print– notice a connection between the AGM work on belief revision and the work in AI on nonmonotonic reasoning and simply assumed, a little glibly, that the latter could be subsumed within the former –roughly the case dealt with via entrenchment, where some parts of a theory are protected against revision–. In terms of abstract consequence relations, that wasn't a bad guess; but notice that it simply leaves unaddressed the issues of finding the relevant non-monotonic (default) fixed points and that, in the context of various models of extended logic programming, is where much of the interesting action of late has been. So, I'd give myself a B-/C+ for prophecy on this one.", David Israel, en una comunicación personal a los autores.

También apareció en 1988 el trabajo de Ken Satoh “Non Monotonic Reasoning by Minimal Belief Revision”.<sup>74</sup> Satoh buscaba diferenciarse de los formalismos conocidos de NMR (Default Logic de Reiter y Circumscription de Mc Carthy) y para ello se propuso “*investigar una estrategia especial de revisión de creencias, llamada ‘revisión de creencias minimal’*”<sup>75</sup> y mostrar que esa estrategia realizaba algunas clases de razonamiento no monótono. Satoh distinguía entre conocimiento y creencias. El conocimiento era un subconjunto de las creencias que era indiscutible (si el agente conoce  $p$ , entonces cree en  $p$ ) y las restantes creencias eran derrotables. La nueva información tenía status de conocimiento, por lo que el corpus de creencias debía acomodarse a ella.

En el caso clásico de (1)  $\forall x (\text{bird}(x) \supset \text{fly}(x))$   
(2)  $\text{bird}(A)$

cuando se adquiría la información de que  $\neg \text{fly}(A)$ , la estrategia de revisión de creencias minimal de Satoh cambiaba la creencia (1) en la siguiente:

(3)  $\forall x (x \ll A \equiv \text{bird}(x) \supset \text{fly}(x))$

La idea central de Satoh para la revisión de creencias era que el orden a minimizar dependía tanto de los modelos de las creencias previas como de los modelos de las nuevas, a diferencia del modelo AGM que sólo dependía de las originales. Satoh formalizó esta idea planteando formulaciones de segundo orden en una forma muy similar a Circumscription. La sección Related Research del trabajo de Satoh comparaba brevemente su propuesta con los formalismos conocidos de NMR, con el TMS de Doyle, con los trabajos en actualización de Bases de Datos, a los que cuestiona por no cumplir con lo que Dalal llama “irrelevancia de la sintaxis” y dedica los apartados finales y más extensos a la ‘*Logic of Theory Change*’ (AGM) y a la reciente propuesta de Dalal, a la que consideraba similar a la suya, aunque señalando algunas diferencias. Como AGM no distingue entre conocimiento y creencias, consideraba la situación en la que el conocimiento era el conjunto de las tautologías. En ese caso ambas definiciones eran comparables. Luego probaba que su propuesta cumplía los postulados AGM, salvo vacuidad

<sup>74</sup> “Non Monotonic Reasoning by Minimal Belief Revision”, Ken Satoh, *Proceedings of the International Conference on Fifth Generation Computer Systems*, (1988), 455-462.

<sup>75</sup> Ken Satoh, “Non Monotonic...”.

y los postulados adicionales,<sup>76</sup> aunque si el lenguaje era proposicional, también se cumplía el primero.

Al igual que Dalal, Satoh recuerda que fue un evaluador el que le mencionó AGM. *"When I wrote the FGCS 88 paper, I had a feedback from an anonymous referee which mentioned the following paper: Alchourrón, C., Gardenfors, P., Makinson, D.; "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions", The Journal of Symbolic Logic, Vol. 50, No.2, (1985), pp 510-530."* reforzando la idea de que la formalización AGM apareció en un momento clave de cuestionamientos y búsquedas en AI que potenciaron su impacto.<sup>77</sup>

Es interesante resaltar que los dos autores analizados hasta ahora desconocían la teoría AGM al momento de concebir sus trabajos y, sin embargo, tenían una sintonía casi perfecta con ella.

En 1989 se publicó "Minimal Change and Maximal Coherence: A Basis for Belief Revision and Reasoning about Actions".<sup>78</sup> Sus autores, Annad Rao y Norman Foo, investigadores en la Universidad de Sydney, Australia, se proponían abordar una cuestión candente en la IA: el razonamiento sobre los resultados de las acciones que va ejecutando un agente a lo largo del tiempo. Estas acciones afectaban el estado del mundo exterior, incluyendo efectos laterales o ramificaciones de la o las acciones.<sup>79</sup> Este razonamiento involucraba determinar el nuevo estado del mundo luego de realizada una acción y el consecuente cambio en las creencias del agente al pasar de un instante del tiempo al siguiente. En este punto, razonar sobre acciones aparecía, en la visión de los autores, asociado a revisión de creencias, entendida como un proceso por el cual un agente revisa

<sup>76</sup> El postulado de vacuidad afirma que si la nueva información no es contradictoria con la vieja entonces la revisión es simplemente incorporarla directamente (sin remoción alguna). Los postulados adicionales afirman que revisar por una disyunción debe ser igual a la revisión por uno de los disyuntos o a la revisión por el otro o a la intersección de ambas.

<sup>77</sup> Ken Satoh, en una comunicación personal a los autores.

<sup>78</sup> "Minimal Change and Maximal Coherence: A Basis for Belief Revision and Reasoning about Actions", Rao, A. S. and N. Foo, *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pp. 966-71, Morgan Kaufmann, San Mateo, Aug. 1989.

<sup>79</sup> Habitualmente se hablaba del "frame problem" y del "ramification problem" y estos temas acarreaban razonamiento presuntivo. Dado que a priori se ignoraban todas las posibles consecuencias de las acciones sobre un escenario dado, el agente debía suponer que se había producido el menor cambio dentro de lo explicitado por el conocimiento previo sobre las consecuencias de dichas acciones y pasar a un nuevo estado de creencias sobre el escenario en el que se movía, que no era "seguro". Aquí ingresaba en este problema el razonamiento no monótono.

su conjunto de creencias en el tiempo actual, a partir de algún input proveniente del mundo externo. En ambos casos la cuestión central era determinar qué creencias deberían cambiar y cuáles no deberían cambiar.<sup>80</sup> En otras palabras, tanto revisión de creencias como razonamiento sobre acciones involucraban razonamiento acerca del cambio. Los autores señalaban que hasta ese momento ambos temas habían sido encarados por separado en la IA, y en el caso de revisión de creencias, presentaban como el camino standard en IA el de la construcción de sistemas eficientes del estilo del TMS de Doyle. El trabajo se proponía dar un enfoque unificado de ambos fenómenos, a través de formalizar sus principios subyacentes, sobre la base de dos principios intuitivos: el principio de minimización del cambio, entendido como *“minimización de la adquisición de creencias o minimización de la pérdida de creencias cuando un agente se mueve de un estado al otro”*,<sup>81</sup> y el principio de maximización de coherencia, entendido como *“la retención de tantas creencias coherentes como sea posible durante un cambio de estado”*.<sup>82</sup> Para esta tarea reconocían inspiración en la axiomatización de la teoría de revisión de creencias debida a AGM. El desarrollo del trabajo utilizaba el enfoque y la nomenclatura AGM. Definía tres posibles estados de creencias de un agente respecto a una fórmula A, en un tiempo t y en un mundo w, a saber, creencia en A, creencia en  $\neg A$  e indiferencia respecto de A y consideraba que una dinámica de creencias consistía en el proceso por el cual el agente pasaba de un estado a alguno de los otros dos, dando lugar a la expansión, contracción o revisión de sus creencias. Formalmente se construía un sistema modal denominado CS –por Coherence Modal System–, con operadores modales EXP, CON y REV para representar las tres operaciones de la dinámica de creencias. Este sistema se caracterizaba por conjuntos de axiomas para cada operador. Parece indudable que el artículo, que formaba parte de la tesis doctoral de A. Rao dirigida por N. Foo, recibió el impacto del conocimiento del trabajo de AGM cuando ya estaba en curso y asimiló parte de sus conceptos dentro de un esquema lógico ya constituido. Esta hipótesis puede considerarse convalidada con el testimonio de Norman Foo. Este nos cuenta que ya desde 1986, se enfrentaban, durante el desarrollo de la tesis doctoral de A. Rao, a la cuestión de cómo resolver el cambio de creencias de un agente cuando éste descubría que algunas de las mismas eran erró-

<sup>80</sup> Posteriormente Katsuno y Mendelzon desarrollaron la idea de “updating” como distinta a “revision” para distinguir entre las consecuencias sobre una Base de Conocimientos de un cambio en el mundo o de un cambio en las creencias sobre el mundo.

<sup>81</sup> Rao y Foo, “Minimal Change and...”.

<sup>82</sup> Rao y Foo, “Minimal Change and...”.

neas. En un viaje a Auckland (Nueva Zelanda) para visitar al Director del Departamento de Ingeniería Eléctrica de la Universidad, Jack Woodward, éste les presentó a Krister Segerberg, profesor de filosofía en dicha Universidad, con el cual concertaron una entrevista. Durante esa reunión, Segerberg les habló del libro sobre Revisión de Creencias que estaba escribiendo Peter Gärdenfors<sup>83</sup> y, al retornar a Sydney, contactaron a Gärdenfors, quien les mandó capítulos de su libro, en los cuales descubrieron la teoría AGM. A partir de este momento se constituyó en Sydney una fuerte tradición en revisión de creencias basada en AGM. Posteriormente las investigaciones abarcaron diversas áreas no reconocibles de forma inmediata como derivadas de la teoría AGM, pero inspiradas en y por ella. En palabras de Norman Foo: *“These chapters were eye-openers for us, and we became aware of the AGM seminal work then. We never looked backed, and I focussed the research of my group on belief revision for the next decade. My students soon produced cutting-edge work in the area. We were also fortunate in being able to attract international post-docs in the area to work with us. Anand Rao and I produced the first Prolog implementation of a version of AGM revision. My students, besides Anand Rao (who went on to co-invent BDI logics with Mike Georgeff), who did PhDs in the AGM style –particularly with finite base revision or applications to reasoning about actions– were Mary-Anne Williams, Simon Dixon, Pavlos Peppas, Yan Zhang, Maurice Pagnucco and Boon Toh Low... Our research has since moved on to using ideas inspired by the AGM paradigm to many areas which may not be recognized as such by people not familiar with the history, but we can confidently say that had it not been for the pioneering paper by Alchourrón, Gardenfors and Makinson we would not have progressed so quickly and so far.”*<sup>84</sup>

Para dimensionar el peso de esta tradición, nacida del cruce, en el momento adecuado, entre una necesidad genuina de la IA con una teoría que –sin proponérselo– venía al encuentro de la misma, podemos decir que la producción de Foo, Rao y sus discípulos representa un 7% del total de las referencias al artículo de AGM que hemos recopilado y el porcentaje sube al 11% si lo tomamos respecto a las citas provenientes de la IA.<sup>85</sup>

Los otros trabajos pioneros, a diferencia de los anteriores, tomaron en cuenta AGM en el punto de partida.

<sup>83</sup> “Knowledge in Flux”, P. Gardenfors, 1988.

<sup>84</sup> Norman Foo, en una comunicación personal a los autores.

<sup>85</sup> Los cuadros de citaciones se encuentran en la Sección 5.

Bernhard Nebel en “A Knowledge Level Analysis of Belief Revision”<sup>86</sup> planteaba que revisar creencias era una tarea que todo agente inteligente debía realizar y que por eso “en AI se han desarrollado cierto número de los llamados sistemas de mantenimiento de la verdad (TMS) que soportan revisión de creencias”.<sup>87</sup> Más adelante señalaba: “Sin embargo subsiste la cuestión de cómo Belief Revision (BR) puede ser descrito en un nivel abstracto, independientemente de cómo las creencias están representadas y manipuladas dentro de la máquina... En particular no es claro cómo describir BR en el KL tal como dicho nivel fue introducido por Newell y otros. Levesque y Brachman reclaman que todo sistema de información pueda ser descrito en el KL sin referencia alguna a cómo la información esté representada y manipulada en el sistema. Sin embargo esto parece ser dificultoso en revisión de creencias... parecería que la representación en el SL es crucial... sin embargo, la teoría del cambio epistémico y la lógica del cambio de teorías, desarrolladas por Alchourrón, Gardenfors y Makinson... muestran que es posible un análisis parcial en el nivel del conocimiento de la revisión de creencias”.<sup>88</sup> Partiendo de un análisis en el KL, provisto en este caso por los postulados de racionalidad de AGM, Nebel se propuso lograr la reconstrucción de funciones de revisión en el nivel simbólico. De la lógica del cambio de teorías decía que “recientemente recibió mucho interés en la comunidad de AI”<sup>89</sup> y proporcionaba como referencias bibliográficas la presentación de Gardenfors y Makinson en la TARK’88 y el trabajo de Dalal que hemos comentado antes. Esta afirmación, no respaldada aún por suficientes publicaciones de investigadores de IA, viene en apoyo de la tesis de que existía un impacto importante, que recién empezaba a manifestarse.

Nebel reconocía que la lógica del cambio de teorías tenía un enfoque idealizado que ignoraba importantes características requeridas por una revisión en el contexto computacional. En particular, parafraseando al propio Gardenfors, “los conjuntos de creencias no pueden ser usados para expresar que algunas creencias pueden ser razones para otras creencias”.<sup>90</sup> Esta característica parecía reducir su utilidad en IA, frente a los TMS (que por otra parte eran construcciones en el nivel simbó-

<sup>86</sup> “A Knowledge Level Analysis of Belief Revision”, Bernhard Nebel, en *Principles of Knowledge Representation and Reasoning*, Brachman, Levesque y Reiter eds., Morgan-Kaufmann, 1989, pp. 301-311.

<sup>87</sup> Bernhard Nebel, “A Knowledge Level...”.

<sup>88</sup> Bernhard Nebel, “A Knowledge Level...”.

<sup>89</sup> Bernhard Nebel, “A Knowledge Level...”.

<sup>90</sup> “Knowledge in Flux”, P. Gardenfors, 1988, p. 67.



lico, sin una caracterización independiente de las estructuras de representación). Ante esto Nebel planteaba que el mantenimiento de las razones o justificaciones podía obtenerse como efecto lateral si se elegía “correctamente” la operación de contracción sobre bases finitas.

En la primera parte del trabajo se presentaban las distintas variantes de funciones de contracción (*‘full meet’*, *‘partial meet’* y *‘maxichoice’*).<sup>91</sup> Luego se pasaba a definir operaciones de contracción sobre bases finitas o “bases de creencias”. Para poder demostrar que una operación de contracción  $\sim$  sobre bases finitas cumplía, en algún sentido, los postulados de AGM, Nebel estableció que, si KB es la base de creencias finita, podían definirse una teoría A y una operación de contracción sobre teorías  $\downarrow$ , del modo siguiente:

$A = \text{def } Cn(KB)$

$A \downarrow x = \text{def } Cn(KB \sim x)$

Considerando a las proposiciones de la base como “más importantes” que las derivadas construía una función de selección, con la cual probaba que la contracción sobre bases podía verse como una *partial meet contraction*. Luego probaba que satisfacía los postulados de contracción (salvo *inclusión conjuntiva*).<sup>92</sup> La revisión la definía, de acuerdo con la identidad de Levi, como

$KB^\circ x = \text{def } (KB \sim \neg x) \wedge x$

Esto le permitía afirmar que el caso finito de contracción (sobre bases de creencias), no era cualitativamente distinto de la operación de cambio epistémico sobre conjuntos de creencias cerrados por deducción y que *“el caso finito puede ser modelado sin problemas por una función de selección particular. Visto desde la perspectiva del nivel del conocimiento, la única información adicional necesitada por una revisión de creen-*

<sup>91</sup> Las referencias bibliográficas que detalla abarcan casi todas las publicaciones de los autores de la teoría, desde el artículo sobre contracción de Alchourrón y Makinson de 1982, hasta un trabajo de Gardenfors de ese mismo año 1989, pasando por el presentado en la TARK’88.

<sup>92</sup> Cuando se desea eliminar una conjunción de una teoría, al menos uno de los dos conjuntos debe ser extraído (en otro caso, dado que el resultado está clausurado por consecuencia obtendríamos la conjunción nuevamente). El postulado de inclusión conjuntiva expresa que si uno de ellos fuera suprimido, es esperable que también sean eliminadas todas las fórmulas que hubieran sido suprimidas al realizar una contracción explícita sólo de ese conjunto.

*cias es una relación de preferencia en conjuntos de proposiciones*<sup>93</sup> y hacía notar que, de todos modos, para que se cumpliera *inclusión conjuntiva* se requería que la función de selección estuviera ligada a una relación transitiva.

Al momento de escribir su artículo, Nebel<sup>94</sup> contaba con un vínculo de intercambio con Gardenfors (que le facilitó borradores de sus trabajos y le aportó comentarios sobre los avances del propio Nebel) y también estaba conectado con Borgida y Dalal, a quienes agradecía sus *"hints e ideas"*. En cuanto al modo en que conoció AGM, Nebel afirma que la conexión vino a través de la relación entre su director de tesis y el filósofo H. Rott. *"I heard about 'belief revision', when I was writing my Ph.D. thesis, probably in 1987 or 1988. My thesis topic was to attack the problem of changing terminologies –or what is called ontologies these days–. My supervisor, Wolfgang Wahlster, told me at that time about the Ph.D. thesis work of Hans Rott, which he knew because Wolfgang Wahlster and Hans Rott are both in the 'Deutsche Studienstiftung'. In any case, I believe, that we set up a meeting and talked about 'belief revision'."*<sup>95</sup>

El atractivo de AGM residía, para Nebel, en el hecho de que encabraba formalmente cuestiones centrales para la IA, cuyo tratamiento al momento se hacía a través de mecanismos ad hoc. *"I later started to look into the existing literature from philosophical logic and was fascinated, in particular because AGM addressed the problem of how logical theories could evolve. At that time, there existed a few approaches to model similar things in Computer Science, but they were all more pragmatic and ad hoc... (Also) it seemed to be an interesting operation missing from databases and seem to be similar to what so-called Truth-Maintenance-System did. Further, belief revision appeared to be dual to reasoning in non-monotonic logics"*<sup>96</sup>

También Katsuno y Mendelzon en "A Unified View of Propositional Knowledge Base Updates"<sup>97</sup> se inspiraron directamente en AGM. En la Introducción, luego de presentar la cuestión de la revisión de una KB (concebida como un conjunto finito de sentencias en un lenguaje L) y de definir –informalmente– las operaciones de revisión, contracción, eliminación ("borrar" una sentencia y sus equivalentes lógicas de la KB) y

<sup>93</sup> Bernhard Nebel, "A Knowledge Level...".

<sup>94</sup> Nebel desarrolla el trabajo en el contexto de un proyecto Esprit de gerencia del conocimiento.

<sup>95</sup> Bernhard Nebel, en una comunicación personal a los autores

<sup>96</sup> Bernhard Nebel, en una comunicación personal a los autores.

<sup>97</sup> "A Unified View of Propositional Knowledge Base Updates", H.Katsuno y A. Mendelzon, *Proceedings of the 11th IJCAI* (1989), pp. 1413-1419.

retracción (deshacer el resultado de una operación previa), los autores remarcan que “*Gardenfors y sus colegas han hecho un trabajo fundacional en revisión de bases de conocimiento*”.<sup>98</sup> Al referirse a los postulados de la teoría, señalan que “*no suponen ninguna representación concreta de la KB, de hecho ésta es modelada como un conjunto de sentencias de un lenguaje no especificado, cerradas por deducción. Cuando consideramos una KB computacional, necesitamos fijar un formalismo y una representación sintáctica finita de la KB... la cuestión que se plantea es si el resultado de una actualización dependerá del conjunto particular de sentencias que forman la KB o sólo de los mundos que son descriptos por ellas*”.<sup>99</sup> En consecuencia consideraban que cualquier método orientado a Bases de Conocimiento finitas debía satisfacer el Principio de Irrelevancia de la Sintaxis de Dalal, autor al que consideraban “*el primero en relacionar su enfoque con los postulados de Gardenfors, y en mostrar que su propia propuesta para un operador de revisión satisfacía los mismos*”.<sup>100</sup> Katsuno y Mendelzon se proponían avanzar hacia una caracterización más general, en teoría de modelos, de los operadores de revisión para bases finitas que satisficieran los postulados AGM. Su principal resultado fue un teorema de caracterización de las operaciones de revisión que cumplieran los seis (ocho) postulados AGM, basado en un pre-orden parcial (total) entre modelos que sólo dependía de la KB. Usando esta idea de caracterizar operaciones de revisión en términos de órdenes entre modelos de la vieja y nueva información, analizaban diferentes propuestas de revisión, bien conocidas en la literatura de IA, mostrando que todas ellas podían ser capturadas con esta semántica, con sólo considerar en cada caso cuál era el orden entre modelos, tanto de lo viejo como de lo nuevo. Entre los métodos estudiados estaban el de Dalal y el de Satoh, (aunque en este último caso la comparación debió restringirse al caso proposicional, ya que Satoh trabajaba con una KB con sentencias de primer orden).

El encuentro de Katsuno y Mendelzon con AGM vino de la mano de una charla sobre la cuestión de la actualización de Bases de Conocimiento sostenida por los autores con Ray Reiter, destacado investigador de la IA, conocido por su ‘Lógica Default’, quien había asistido a la TARK’88.

*“The story goes back to 1988. At that time, I worked for NTT (Nippon Telegraph and Telephone Corporation), and NTT kindly gave me a chance for a kind of sabbatical. Then, I stayed at Toronto for one year as*

<sup>98</sup> H.Katsuno y A.Mendelzon, “A Unified View...”.

<sup>99</sup> H.Katsuno y A.Mendelzon, “A Unified View...”.

<sup>100</sup> H.Katsuno y A.Mendelzon, “A Unified View...”.

*a visiting scientist, and Alberto hosted me. I arrived at Toronto in the end of August 1988. In the beginning of September, Alberto arranged a short meeting with Ray Reiter, because Alberto knew that I was very much interested in a series of Ray's works. Then, I introduced myself to Ray, and I probably said to him that I was interested in update of knowledge bases. At the end of the meeting, Ray gave us a copy of two papers: Peter Gardenfors and David Makinson, "Revisions of Knowledge Systems Using Epistemic Entrenchment", 83-95, Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, CA, March 1988, and Morgan Kaufmann 1988 and Dalal M., "Updates in Propositional Databases", Technical Report, DCS-TR222, Department of Computer Science, Rutgers University, February 1988. I knew for the first time the existence of AGM paper through the above Gardenfors and Makinson paper, and I guess that Alberto also didn't know AGM work until then. At least, Alberto didn't say to me that he knew Alchourrón at that time. I remember that, later, Alberto informed me that Alchourrón was a famous philosopher in Argentina, but I cannot recall when he said so".<sup>101</sup>*

También en 1988, el mismo año de la segunda TARK, se realizó en Alemania el *Second Workshop on Non Monotonic Reasoning*.<sup>102</sup> Los principales actores del mismo fueron destacados investigadores de IA, que presentaron allí sus propuestas, algunas de las cuales usaban mecanismos ad hoc, mientras que otras recurrían a las herramientas de las lógicas modales. Pero lo que distinguió a este Workshop fue el hecho de que, como reflejo del mismo tipo de inquietudes dentro de la IA que venimos recalcando, haya sido invitado a exponer David Makinson y que la única ponencia en la sección de Aspectos Generales del Razonamiento No Monótono (RNM) haya sido su trabajo.<sup>103</sup> La presentación de un esquema conceptual que permitía caracterizar las propiedades de la noción de consecuencia lógica subyacente en los distintos formalismos ad hoc del RNM (como la Lógica Default de Reiter) produjo una fuerte impresión, probablemente, por las mismas razones de fondo que facilitaron el impacto de AGM.<sup>104</sup> Karl Schlechta, investigador en IA presente en el Workshop, afirma que, pro-

<sup>101</sup> Hirofumi Katsuno, en una comunicación personal a los autores. Lamentablemente el notable investigador argentino residente en Canadá, Alberto Mendelzon, falleció prematuramente en 2005.

<sup>102</sup> *Second Workshop on NMR*, Grassau, Alemania, junio de 1988.

<sup>103</sup> David Makinson, "General Theory of Cumulative Inference", en Michael Reinfrank, Johan de Kleer, Matthew L. Ginsberg, Erik Sandewall (Eds.): *Non-Monotonic Reasoning, 2nd International Workshop*, Grassau, FRG, June 13-15, 1988, *Proceedings. Lecture Notes in Computer Science*, 346 Springer 1989.

<sup>104</sup> D. Makinson, en una comunicación personal a los autores.

bablemente, su primer conocimiento de AGM fue a través de la charla de D. Makinson en Grassau.<sup>105</sup> El año siguiente, Schlechta presentó “Some Results on Theory Revision”,<sup>106</sup> realizado bajo la impronta de AGM y, en particular, del “*epistemic entrenchment*”. Comienza afirmando que “*El problema de revisión de teorías consiste en ‘agregar’ una fórmula a una teoría, preservando consistencia y haciendo sólo cambios mínimos a la teoría original. Una forma natural de determinar en forma única este proceso es imponiendo un orden de ‘epistemic entrenchment’ sobre las fórmulas, como lo han hecho Gardenfors y Makinson*”.<sup>107</sup> Reconocía, como limitación del modelo, la dificultad en el abordaje de las revisiones consecutivas (iteradas) ya que éstas constituyen un fenómeno común en los sistemas cognitivos. La primera parte del trabajo apuntaba a una propuesta superadora de esa limitación (construir un orden para todas las teorías dentro de un mismo lenguaje). La segunda, independiente de la primera, trataba con contracciones de sistemas  $\langle K, A \rangle$  donde K era un conjunto de fórmulas cerrado por deducción y A un conjunto de axiomas para K. En estos casos, para Schlechta, una revisión para  $\langle K, A \rangle$  consistía, esencialmente, en elegir un adecuado subconjunto de A. En la sección final presentaba propuestas de métodos para poder seleccionar subconjuntos consistentes razonables de conjuntos –parcialmente ordenados– de información conflictiva, tanto ‘dura’ como default, en un contexto de lenguajes ordenados parcialmente, aumentados con información ‘default’, y donde la información más específica era más confiable, en caso de conflicto, y la incomparable era tratada en forma imparcial (‘fair’). Para la época en que elaboró el trabajo, Schlechta era colega de Nebel, tanto en IBM Alemania como en el proyecto Lilog de la Comunidad Europea, y estaba en contacto con Gardenfors y con Makinson, a los que agradeció por las lecturas previas de los borradores y por las sugerencias recibidas.

Desde una perspectiva simétrica a las TARK y en medio del crecimiento de la comunidad de cambio de creencias, en 1989 se reunió en Konstanz, Alemania, un Workshop denominado *La lógica del cambio de teorías*, cuyo centro era AGM y sus derivaciones.<sup>108</sup> Esa fue la única oportunidad en que el trío de autores AGM estuvo reunido físicamente. A dicho

<sup>105</sup> “I probably first heard about Theory Revision in a talk given by David Makinson at NMR Workshop 6/88 in Grassau, Germany”, Karl Schlechta, en una comunicación personal a los autores.

<sup>106</sup> “Some Results on Theory Revision”, Karl Schlechta, *Proceedings of the Workshop on The Logic of Theory Change*, Konstanz, 1989.

<sup>107</sup> Karl Schlechta, “Some Results...”.

<sup>108</sup> *The logic of theory change, Workshop*, Konstanz, Alemania, octubre de 1989.

evento fueron invitados varios investigadores de IA (Doyle, Martins, Brewka y Schlechta). De entre las ponencias presentadas por éstos, Martins no se refirió a la teoría AGM, Doyle la referenció, entre otra bibliografía, sólo para presentar como fenómenos análogos la actualización de Bases de Conocimiento, tal como había sido llevada en IA en esos años y la revisión de conjuntos de creencias tal como se la encaraba desde la filosofía y Brewka desarrolló, como los otros, su propio sistema y, al pasar, se refirió a AGM, haciendo el centro en la divergencia entre el tratamiento de conjuntos de creencias (cerrados por consecuencia lógica) y las necesidades de IA de operar con bases finitas. En esa oportunidad sólo Schlechta tomó AGM como una referencia sustancial en su ponencia, tal como hemos comentado anteriormente. Podríamos decir que Konstanz funcionó como una TARK'88 invertida y que su efecto fue "cerrar el círculo" de las conexiones entre campos de investigación. En los años siguientes, todos los investigadores de IA que participaron en Konstanz citaron o referenciaron a AGM y la mayoría de los invitados, más allá de sus orígenes disciplinares, trabajaron, en conjunto o por separado, tanto en temas de cambio de creencias como de no monotonía, bajo la influencia de la teoría AGM y sus derivaciones.

Como se reflejaba en las discusiones contenidas en estos primeros trabajos, existían abundantes propuestas que trataban con las actualizaciones de KB, pero con fuertes limitaciones, ya sea porque la consideración de la revisión por información inconsistente con la ya existente se dejaba al margen de las formalizaciones o bien porque se encaraba por procedimientos en el nivel simbólico, ad hoc, sin una clarificación adecuada, ya sea porque las propuestas no cumplían con el principio de la irrelevancia de la sintaxis, o no satisfacían el intuitivo principio de mínima pérdida. Estas situaciones llevaron a que AGM fuera adoptado como un modelo de cómo especificar en forma abstracta las propiedades de un mecanismo de revisión. En los años siguientes a 1990 AGM apareció citada por numerosos investigadores del área de IA, como R.Reiter, J.Doyle, M.Winslett, J.Halpern, Brewka, Shoham y J.P.Delgrande, entre otros. Si bien muchos cuestionaban sus limitaciones (como la dificultad para iterar cambios) o su carácter idealizado o excesivamente simplificado, sin embargo se convirtió en una referencia obligada, como lo muestra el siguiente pasaje de un artículo del año 2000, "Belief Revision: a Critique", uno de cuyos autores es J.Halpern, principal animador de las primeras TARK.<sup>109</sup>

<sup>109</sup> "Belief Revision: a Critique", N. Friedman y J.Halpern, *Journal of Logic, language and Information*, 2000.

*“Las aproximaciones al cambio de creencias comienzan típicamente con una colección de postulados, una argumentación de su razonabilidad y la prueba de algunas consecuencias de esos postulados... El mensaje principal de este artículo es que describir postulados y probar teoremas de representación no es suficiente...”*<sup>110</sup> Luego de aceptar que esto último podía haber sido razonable en los años ochenta del siglo pasado, y proponer que hoy es necesario también aportar una ontología o escenarios para el proceso de cambio de creencias, el trabajo afirma que *“nuestro foco estará puesto en las aproximaciones que toman como su punto de partida los postulados de cambio de creencias propuestos por Alchourrón, Gärdenfors y Makinson...”*<sup>111</sup>

## 5. El impacto de AGM hasta nuestros días

Sin ninguna duda el modelo AGM se ha convertido en un paradigma para los que intentan desarrollar sistemas que tengan “a raya” la inconsistencia y las fallas. Tanto en Inteligencia Artificial como en Ciencias de la Computación, casi todo trabajo que trate de abordar con un soporte formal el problema de operar con información inconsistente, utiliza las ideas generales de este modelo. Un caso particular donde se manifiesta este fenómeno es con los sistemas a prueba de fallas, donde se acepta que puede haber errores y el resultado de una situación errónea es “reparada” con una función de revisión. Curiosamente este puede verse como una reivindicación tardía (y no intencional) de algunas de las ideas del primer artículo de Alchourrón y Makinson *“Hierarchies of regulations...”*. Con el tiempo dichas áreas se fueron apropiando del modelo, adaptándolo, replanteándose varias de sus presunciones, abandonando algunas e imponiendo otras. Puede conjeturarse que ese nivel de aceptación se debe esencialmente a que AGM es el primero que plantea la construcción efectiva de operaciones de cambio y supera, a la vez, tanto lo meramente descriptivo y ad-hoc como lo puramente axiomático. En esta sección buscamos mensurar, cuantitativa y cualitativamente, el nivel de impacto e inserción del modelo AGM en ambas áreas.

El sistema de medición aplicado parte del análisis de citas bibliográficas. Esta decisión nos brinda un contexto preciso donde realizar las búsquedas y cuantificaciones. A su vez, dentro del espectro de potenciales selecciones que podrían hacerse, incluyendo los casos de referencias

<sup>110</sup> Friedman y Halpern, “Belief Revision...”.

<sup>111</sup> Friedman y Halpern, “Belief Revision...”.



indirectas al modelo, hemos decidido restringirnos exclusivamente a relevar los artículos y libros que mencionan explícitamente el artículo original de AGM de 1985, ya que ése es, a nuestro juicio, un indicador indudable de la presencia conceptual de AGM en el artículo en cuestión.

En este marco metodológico, se realizó una búsqueda de todos los artículos que contenían entre su bibliografía una referencia a dicho artículo, descartando las autorreferencias.<sup>112</sup> Como resultado de esa recopilación se confeccionó una planilla con un millar de entradas donde, para cada una, figura el título del trabajo, los autores, lugar de publicación y el año. Los datos volcados fueron validados y verificados cruzando diferentes fuentes de información. Posteriormente los trabajos fueron clasificados en cuatro categorías: Filosofía, Inteligencia Artificial, Computación y Lógica. Para la categorización se tuvo en cuenta el perfil de los autores, su área de incumbencia y el del lugar de publicación. En los casos de otras disciplinas, como Ciencias Cognitivas o Economía, fueron clasificados como de Filosofía.

El **Cuadro 5.1** sintetiza los resultados cuantitativos de la búsqueda y clasificación. La **Figura 5.1** muestra en forma gráfica la evolución de la cantidad de citas de AGM con el correr de los años. Para realizarlos, y sólo a los efectos de mejorar su visualización, se agruparon las categorías Lógica y Filosofía por un lado e Inteligencia Artificial y Computación por otro. Podemos observar que las menciones del artículo fundacional se incrementan en forma notable a partir del año 1994, en el que se duplican las citas de 1993. Luego experimentan un especial crecimiento desde 1996, que se sostiene hasta la actualidad (los datos del 2006 son parciales, pero, probablemente, al terminar el año se mantengan los guarismos del 2005). También observamos que la proporción de trabajos del área conjunta Computación más IA, no ha bajado, desde 1992, del 50% y, si descartamos tres años, está siempre por sobre el 60%, piso que se mantiene estable desde el 2001 hasta la fecha.<sup>113</sup>

El análisis cualitativo se realizó sobre las publicaciones del 2005, a modo de estudio de caso, con la intención de caracterizar el contexto y el tipo de uso que se hace de la teoría.

Antes de analizar los contenidos de los artículos vale la pena recordar que el modelo AGM supone que las teorías son estables y se corres-

<sup>112</sup> Las fuentes fueron: Web of Science, Citeseer y Google Scholar.

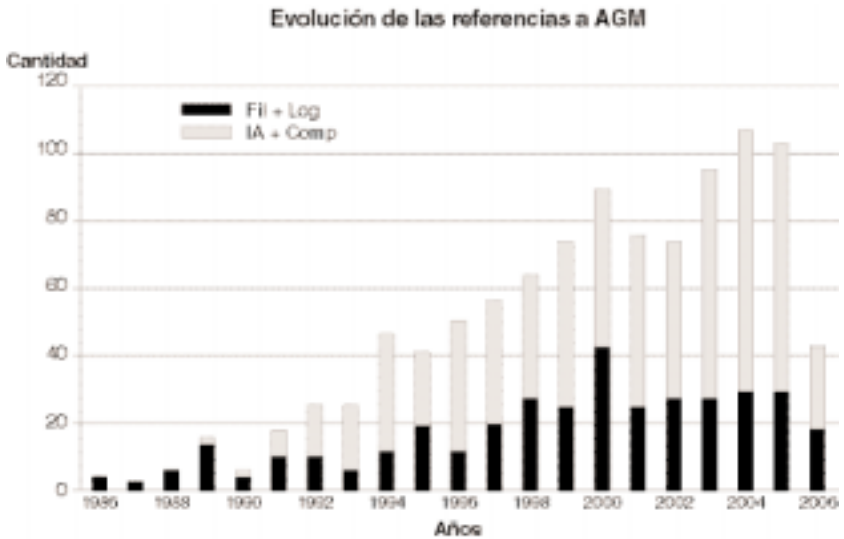
<sup>113</sup> Hay que señalar que, en los primeros años del período bajo análisis, cierta cantidad de trabajos no se difundió en versión electrónica ni fueron pasados más adelante a dicho soporte, por lo cual las citaciones de esos primeros años son claramente incompletas. Un caso notable es el trabajo de Rao y Foo comentado en la Sección 4.

<b>AÑO</b>	<b>Fil+Log</b>	<b>% Fil+Log</b>	<b>IA+ Comp</b>	<b>% IA+Comp</b>
<b>1986</b>	2	66,7	1	33,3
<b>1987</b>	1	100,0	0	0,0
<b>1988</b>	4	66,7	2	33,3
<b>1989</b>	12	85,7	2	14,3
<b>1990</b>	3	50,0	3	50,0
<b>1991</b>	9	56,3	7	43,8
<b>1992</b>	8	33,3	16	66,7
<b>1993</b>	6	25,0	18	75,0
<b>1994</b>	11	23,9	35	76,1
<b>1995</b>	18	46,2	21	53,8
<b>1996</b>	10	20,0	40	80,0
<b>1997</b>	18	32,1	38	67,9
<b>1998</b>	26	41,3	37	58,7
<b>1999</b>	24	32,9	49	67,1
<b>2000</b>	41	46,6	47	53,4
<b>2001</b>	24	32,0	51	68,0
<b>2002</b>	26	35,6	47	64,4
<b>2003</b>	25	26,9	68	73,1
<b>2004</b>	29	27,6	76	72,4
<b>2005</b>	29	29,0	71	71,0
<b>2006</b>	16	37,2	27	62,8
<b>342</b>	<b>34,3</b>	<b>656</b>	<b>65,7</b>	

**Cuadro 5.1.** *Cantidad de publicaciones por año (absoluto y relativo)*

ponden con conjuntos de fórmulas (posiblemente infinitos) de un lenguaje supraclásico, cerrados por un operador de consecuencia lógica, que incluye las tautologías, la regla de introducción de disyunción en las premisas, y compacidad. En el mismo no existe distinción entre creencias explícitas y las derivadas. La información que desencadena el cambio es representada por una fórmula proposicional, privilegiada respecto de la teoría existente. Las actitudes epistémicas son sólo tres: aceptación, rechazo o indiferencia, lo que condiciona a tener una igual cantidad de operaciones de cambio: Expansión, Contracción y Revisión. Estas son funciones que tienen por dominio pares ordenados teorías-fórmulas y por rango teorías, y están basadas en el “principio de mínimo cambio”.

Todos estos supuestos han sido cuestionados por diferentes autores y la mayoría de los artículos que aparecen en los últimos años proponen modelos alternativos que, en lo fundamental, plantean las siguientes cuestiones.



**Fig. 5.1.** *Evolución temporal de la cantidad de citas del artículo AGM 85*

- 1) Cambiar la elección del lenguaje de representación. Esto en general conlleva plantear nuevas nociones de “consistencia” en la información, contradicción, información derivada y vínculo entre los elementos del conjunto. Cualquiera sea el lenguaje elegido, surge la pregunta del modo de representación del corpus de información: mediante una única sentencia en el lenguaje (la conjunción de las sentencias independientes) o mediante un conjunto (quizás infinito) de ellas. En el caso de un conjunto, ¿éste debe ser cerrado bajo la noción de consecuencia lógica o como una simple enumeración de hechos “desnudos”? Esta segunda opción implica la necesidad de calcular luego de alguna manera las consecuencias de estos hechos y tomar el compromiso de diferenciar o no entre información explícita e implícita.
- 2) ¿Puede el corpus actualizarse “espontáneamente” o requiere de un “estímulo” externo?; es decir, ¿son estos internamente estables? En caso de considerar un corpus de información que sufre cambios en respuesta a estímulos externos: ¿debe haber diferencia entre el lenguaje de representación del corpus y el de la información que desata el cambio? Aun si el lenguaje fuera el mismo, ¿deben ser homogéneas las representaciones? ¿consistir ambas en fórmulas o en conjuntos de fórmulas? La primera postura es

- sostenida por la mayoría de las propuestas en informática y la segunda da lugar a las funciones de cambio múltiples.
- 3) ¿Los valores epistémicos asignados a las expresiones del lenguaje serán solamente aceptación, rechazo, indeterminación, o se deben considerar grados de aceptabilidad? ¿Qué tipo de información puede representarse en el corpus? ¿Puede ser representada información de cómo modificar el mismo corpus? Si el Orden Epistémico para una teoría K (*epistemic entrenchment*) lo caracteriza completamente, ¿por qué no plantear una revisión del orden directamente?
  - 4) ¿Debe relacionarse la noción de pertenencia de una sentencia al corpus y la validez de dicha sentencia? ¿Podemos contener en el corpus información falsa? ¿Las expresiones del lenguaje serán interpretadas como verdaderas o falsas o ni verdaderas ni falsas, o tendremos grados de verdad, inconsistencia, etc.?
  - 5) Por otro lado, parece fundamental definir operaciones que respondan a la noción de mínimo cambio, o máxima preservación del corpus de información. Es decir, se requiere alguna forma de “calcular el valor” de la información a ser descartada. ¿Existe un orden de preferencia que representa su credibilidad, solidez o valor informacional entre expresiones del lenguaje? ¿Este orden está incluido en el corpus de información o es intrínseco a la operación de cambio? ¿El mínimo cambio debe ser cuantitativo o cualitativo? ¿Cuántas y cuáles son las distintas formas en que un corpus de información puede ser modificado? ¿Son independientes o interdefinibles?
  - 6) El vínculo entre el corpus original y el actualizado: ¿es relacional o funcional? La función o relación: ¿es sólo entre el corpus original, la nueva información y el corpus modificado o deben considerarse otros parámetros más? ¿Las operaciones de cambio deben tener en cuenta la historia de los cambios producidos, o cada nueva operación se produce independientemente de las anteriores? ¿El proceso de actualización del corpus debe mantener la interpretación de las expresiones del lenguaje, o es pensable que un cambio altere las proposiciones asociadas a las expresiones del corpus? ¿Revisar un corpus por una secuencia de N sentencias es equivalente a N revisiones sucesivas? Este último problema es conocido como el de iteración.
  - 7) La operación desatada por la nueva información contingente ¿siempre debe ser exitosa? ¿Puede aceptarse parcialmente la nueva información?

Los 71 trabajos publicados en el año 2005 que, de acuerdo al criterio utilizado, hemos clasificado en las áreas de Inteligencia Artificial y Ciencias de la Computación, poseen las siguientes características.

- a) La mayor atención se la llevan los tendientes a resolver el problema de revisiones iteradas. En total hay nueve trabajos que, con diferentes propuestas, atacan el problema en forma directa. Esencialmente aquí la cuestión pasa por determinar cómo cambiar las preferencias en contraposición o la opción de cambiar las creencias mismas.
- b) Otro tema que resalta es la modelización del razonamiento abductivo a través de funciones de Agregado (que son casos especiales de expansiones),<sup>114</sup> tema que se refleja en siete trabajos que siguen la línea planteada por Pagnucco.<sup>115</sup>
- c) Un grupo de dieciséis artículos trata diferentes propuestas para generar rankings que representen las preferencias epistémicas que guían la operación de cambio. Varios procuran incorporar modelos empíricos donde se trata de determinar o capturar las preferencias humanas determinadas por distintas fuentes como emociones, estados de ánimo, etc. Algunas de estas propuestas también resuelven el problema de revisiones iteradas, dado que su representación de creencias se centra en el ranking y el cambio se asocia a un cambio en él, con lo que se obtiene un nuevo ranking a partir del ranking anterior y de una nueva sentencia.
- d) Varios trabajos se refieren directamente al problema de tratar con información inconsistente o conflictiva (dependiendo si hay o no noción de verdad involucrada). El espectro es amplio pero vale la pena mencionar algunos casos para dar un panorama de este grupo. Uno propone axiomatizar el nivel de inconsistencia al estilo AGM y haciendo un paralelo con aquella. Otro plantea la modelización de la interacción entre agentes de negociación y mediación a través de una variante de AGM. Otros plantean la tolerancia a la inconsistencia y, sin saberlo, reinventan la operación de 'delivery' que apareció en "*Hierarchies of regulations and their logic*". En algunos de los restantes artículos se indican técnicas para descubrir y resolver inconsistencias que, al pretender mostrar cierto grado de formalidad lógica, hacen referencia a

<sup>114</sup> Hans Rott, "Conditionals and Theory Change: Revision, Expansions and Additions", *Synthese*, 81:91-113, 1989.

<sup>115</sup> "The role of Abductive Reasoning within the process of Belief Revision", M. Pagnucco, PhD Thesis, Universidad de Sidney, 1996.

AGM como un patrón de comparación. En esta categoría podemos incluir unos ocho artículos.

- e) Una pequeña línea de investigación, que aparece como denominador común de otros cinco artículos, se refiere a las revisiones múltiples (donde no se revisa por una única fórmula sino por un conjunto) y, en ese contexto, a las operaciones de mezcla, donde la nueva información no es priorizada respecto de la vieja.
- f) Un común denominador de una veintena de artículos es la utilización de lenguajes formales alternativos a la lógica clásica. Así aparecen reductos de lenguaje natural, '*description logics*', OWL, Prolog, lenguajes modales, etc. La principal motivación corresponde al aumento de la demanda de lenguajes de ontologías, a raíz de las aplicaciones en 'Web Semantics' donde, dada la dinámica de estos sistemas, se incorporan funciones de cambio.
- g) Finalmente el resto de los trabajos no incluidos arriba refieren a aplicaciones estándares de AGM con construcción algorítmica de funciones de revisión y, salvo en muy pocos casos, la referencia es marginal.

## 6. Síntesis y Conclusiones

Carlos Alchourrón fue, ante todo, un destacado filósofo del derecho. Como tal desplegó investigaciones pioneras en tres tópicos inspirados directamente en el campo jurídico: la distinción entre normas y proposiciones normativas; el análisis lógico de la operación de derogación de un elemento de un corpus de leyes o regulaciones, y la cuestión de cómo formalizar en un sistema deóntico las obligaciones condicionales, habida cuenta de la posibilidad de que su aplicación a ciertas situaciones de hecho lleve a conclusiones (veredictos) en conflicto. Para cada uno de ellos realizó un análisis filosófico profundo y elaboró, en solitario o en conjunto con sus destacados socios intelectuales, diversos formalismos lógico-matemáticos cuya generalidad y oportunidad histórica lo llevaron a trascender el ámbito jurídico para proyectarse en el campo de la Inteligencia Artificial (IA), una disciplina que, en ese momento, estaba en auge. En el caso particular de la Teoría AGM, ésta tuvo y sigue teniendo un fuerte impacto en IA, donde se ha convertido en un clásico.

Sin embargo, y pese a la conexión intelectual entre sus líneas de investigación y cuestiones que eran centrales dentro de la IA, su actividad académica podría no haber traspasado las fronteras de la Filosofía del Derecho de no haber existido, desde fines de la década de 1970 y durante la de 1980, una coyuntura peculiar. Por un lado, se vivió en esa época

un período de esplendor de la IA, en la que este campo, fundado un cuarto de siglo atrás, se potenció con la revolución informacional naciente y se convirtió en un imán para practicantes de las más diversas disciplinas, particularmente filósofos, atraídos por sus rutilantes promesas. Por otro lado, este mismo crecimiento generó una crisis que llevó a muchos investigadores a revisar los cuestionamientos que se le habían hecho a la lógica y a los sistemas formales en los años anteriores. La TARK'88, donde AGM se presentó por primera vez en el mundo de la IA, puede verse como el fruto de ese doble movimiento: la convocatoria multidisciplinaria que tenían los problemas allí planteados y la búsqueda abierta de soluciones por parte de la comunidad de IA. En especial, la cuestión de cómo actualizar las Bases de Conocimientos frente a posibles inputs inconsistentes con el corpus previo, era un tema irresuelto, una "lata de gusanos" al decir de Halpern. Por eso es que, cuando aparece AGM, se topa con varios intentos en desarrollo que sintonizan perfectamente con ella, como es el caso de los trabajos de Dalal o de Satoh que hemos reseñado. Con esto queremos señalar que no debe creerse que había un vacío de propuestas formales. Pero parece evidente que la aparición de un modelo abstracto, "en el nivel del conocimiento", de acuerdo a la formulación de Newell, con una axiomática, una semántica y unos métodos constructivos que coincidían exactamente, produjo una fuerte atracción intelectual, al menos en los sectores más formalistas de la comunidad de IA.

En nuestro trabajo hemos validado este marco interpretativo general reseñando los escritos de destacados investigadores de la época y, al mismo tiempo, a través de la reconstrucción de las 'microhistorias' de cada uno de los que, dentro de la IA, fueron influenciados por AGM en los primeros momentos.

¿Pero podemos pensar que sólo estamos frente a un impacto unidireccional de las ideas de Alchourrón hacia la IA? Si pensamos en su curiosidad por develar los "misterios lógicos" que se escondían detrás de Prolog o de indagar en las posibilidades de la deducción automática en el campo jurídico, para lo cual promovió una integración interdisciplinaria que fue una "rara avis" en el panorama de la Universidad de Buenos Aires, podremos vislumbrar que él también fue parte de ese entrecruzamiento de filósofos, lingüistas, lógicos y psicólogos que buscaban respuestas en la IA y que, a la vez, le aportaban perspectivas nuevas para sus viejos problemas. En todo caso, desde su peculiar disposición personal, y a diferencia de otros filósofos que se "subsumieron" en el mundo de la IA, Alchourrón siempre se mantuvo fiel a las intuiciones provenientes del dominio legal, lo que lo llevó a inclinar sus preferencias por la construcción efectiva de funciones de contracción y, en particular, por enfoques del



tipo de la “safe contraction”, que le parecía más natural por tratarse de un orden entre las normas, o a cuestionarse los desarrollos de revisión que operaban con teorías y no con bases finitas (llegó a calificar a los primeros de “desvaríos metafísicos”) y, en el caso de los condicionales derrotables, a buscar su formalización en el nivel del lenguaje objeto.

En síntesis, si bien el nombre de Carlos Alchourrón quedó asociado en IA a la Teoría de Cambio, sus vínculos con este campo del conocimiento fueron recíprocos, ya que no fue un investigador encasillado en un rótulo disciplinario, sino un pensador abierto a los desafíos intelectuales de su época.

### **Agradecimientos**

No hubiéramos podido siquiera intentar reconstruir la historia de la introducción de la Teoría AGM en la Inteligencia Artificial, sin el invaluable aporte de muchos de los principales actores de aquellos años de efervescentes discusiones. Todos ellos se entusiasmaron con nuestras consultas, realizaron verdaderos esfuerzos de memoria y nos honraron con sus emotivos recuerdos. Por eso muchas gracias a: Cristina Bicchieri, Horacio Arló Costa, Alex Borgida, Mukesh Dalal, Norman Foo, Joseph Halpern, Gilbert Harman, David Israel, Hirofumi Katsumo, Hector Levesque, Bernhard Nebel, Ken Satoh, Karl Schlechta. Queremos hacer público nuestro reconocimiento a nuestro amigo Eduardo Fermé por las agudas observaciones que nos hizo al revisar el borrador de nuestro trabajo. Un agradecimiento muy especial a Gladys Palau, Eugenio Bulygin, David Makinson y Antonio Martino, quienes tomaron un compromiso personal muy directo en la reconstrucción, aceptando pacientemente nuestras reiteradas consultas, y compartieron generosamente con nosotros sus vivencias junto a Carlos Alchourrón.

Finalmente con este trabajo queremos honrar la memoria de Carlos Alchourrón de quien nos sentimos discípulos y amigos.