

# EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS XI JORNADAS

VOLUMEN 7 (2001), Nº 7

Ricardo Caracciolo

Diego Letzen

Editores



ÁREA LOGICO-EPISTEMOLÓGICA DE LA ESCUELA DE FILOSOFÍA  
CENTRO DE INVESTIGACIONES DE LA FACULTAD DE FILOSOFÍA Y HUMANIDADES  
UNIVERSIDAD NACIONAL DE CÓRDOBA



Esta obra está bajo una Licencia Creative Commons atribución NoComercial-SinDerivadas 2.5 Argentina



## Dinámica de Creencias en presencia de elementos incompatibles

Diego Letzen\*

La dinámica de creencias es un área de estudio originada recientemente, a partir de la convergencia de dos tradiciones de investigación: una relacionada con la computación y el desarrollo de la *inteligencia artificial*, y la otra con la investigación filosófica. El estudio de temas relacionados con el aspecto dinámico de sistemas de creencias ha estado presente tal vez, desde el comienzo de la pregunta por el conocimiento y por consiguiente, desde los mismos orígenes de la reflexión filosófica.

En esta área, el modelo AGM<sup>1</sup> ha permitido una presentación formal que permite dar cuenta de algunos principios intuitivamente asociados con las operaciones de cambio de creencias en agentes racionales. Sin embargo, los modelos están definidos de forma tal que, la existencia de algún tipo de inconsistencia en el conjunto de creencias del agente, produce su inmediata trivialización, al punto que se ha establecido como uno de los principales objetivos de un modelo, preservar la consistencia aunque esto resulte en muchos casos en la pérdida de información innecesariamente, con el solo propósito de mantener la consistencia.

Presentamos en este trabajo un conjunto de modificaciones sugeridas sobre la definición de operaciones de cambio en el modelo AGM, que permitan la representación de situaciones con inconsistencias, sin que aparezca el problema de las inferencias triviales, derivadas del carácter explosivo de la noción de consecuencia clásica, dando prioridad al criterio de la máxima información sobre el de la consistencia.

Las inconsistencias son, una zona límite en este marco de trabajo. La consistencia es considerada tradicionalmente una propiedad fundamental de los sistemas formales, relacionada con la calidad de las inferencias y por ende de la información disponible. Un sistema inconsistente determina en el contexto clásico inferencias triviales y con ello descalifica la credibilidad en los elementos que se pueden asumir a partir de él. Las operaciones de cambio están definidas en forma clásica y el mantenimiento de la consistencia es un elemento constitutivo de su funcionamiento.

El sentido de esta exigencia se pierde rápidamente cuando se la presenta desde el otro punto de vista: la inferencia clásica es explosiva en el sentido de tener una propiedad negativa, la de carecer de capacidad para restringir o paralizar el efecto devastador de la presencia de elementos inconsistentes. La situación en la que es necesario representar algún tipo de elemento inconsistente, que no se 'propague', trivializando el conjunto de creencias, es el objeto de consideración en el contexto de este trabajo. Esta situación excede la caracterización de los conjuntos de creencias (la parte estática del modelo), y llega hasta la definición de las operaciones de cambio (parte dinámica del modelo).

Entenderemos que una contradicción lógica introduce una restricción por la cual, si una creencia es considerada verdadera, su opuesta no puede ser verdadera. Frente a este con-

\* Universidad Nacional de Córdoba.

cepto de contradicción podemos considerar a las incompatibilidades como un par de afirmaciones cuya intersección es vacía, pero pueden considerarse en cierto modo como contradicciones no lógicas. En este trabajo presentaremos una opción que permita una representación de creencias con elementos en general incompatibles.

En trabajos anteriores hemos logrado mostrar que no basta con eliminar la clausura de los conjuntos de creencias en la representación de los estados de creencias y que deben modificarse también las operaciones de cambio en ese conjunto, a fin de establecer la posibilidad de conjuntos inconsistentes que no sean afectados tras los procesos de cambio.

Contraer una base de creencias por un enunciado, en este contexto, consiste en realizar la intersección de una selección de los subconjuntos maximales de la base que no implican el enunciado a contraer.<sup>2</sup>

Los subconjuntos maximales de la base que no implican al elemento a contraer deberán en primer lugar eliminar algún elemento del subconjunto de elementos inconsistentes (siempre que exista algún subconjunto tal) puesto que en otro caso este subconjunto inconsistente permitirá seguir infiriendo el elemento a contraer, por estar considerando las inferencias clásicas de un conjunto aún cuando este conjunto no esté cerrado lógicamente. Así, los subconjuntos inconsistentes de una base nunca podrán permanecer tras una contracción y por ende tampoco podrán hacerlo tras una revisión.

Una solución para lograr una dinámica de bases de creencias que tolere la representación de inconsistencias es la llamada de *cambio local*, propuesta por S.O. Hansson y R. Wassermann.<sup>3</sup> Consisten en definir operadores de cambio que a excepción del operador de expansión que tendrá alcance global,<sup>4</sup> actúen sobre una parte de la base de creencias sin afectar al resto. Para lograr esto se recurre a la noción de *compartimentalización* de una base de creencias en función de conjuntos de enunciados relevantes y por su intermedio, a la definición de un operador de consecuencia (las consecuencias locales) el que permite a su vez definir nuevamente las operaciones de tal manera que tengan alcance local (a una parte de la base) permitiendo la existencia de inconsistencias que no son afectadas tras el proceso de cambio por estar fuera del alcance del operador de consecuencia local utilizado.

Si bien esta propuesta logra el cometido de conservar subconjuntos inconsistentes de información que permanezcan tras las operaciones de cambio, como se muestra en el trabajo, esto se logra creando un operador de consecuencia especial, que opera como una restricción al operador clásico (del cual se sigue dependiendo, arrastrando las complicaciones derivadas de su uso), pero que carece de algunas propiedades importantes de un operador de consecuencia como por ejemplo la inclusión, que es junto con la monotonía y la idempotencia una de las tres propiedades básicas para un operador de consecuencia según la caracterización de Tarski.

Si suponemos un conjunto  $A$  compuesto de aquellos enunciados que corresponden o representan un cierto estado de creencias de un agente, deseamos saber qué sucede si deseamos representar el cambio de creencias del agente por la incorporación o eliminación de elementos del conjunto de enunciados que asumimos lo representaba, de forma tal que esta dinámica cumpla con algunas restricciones o criterios de racionalidad que caracterizan como sucede o debe suceder este cambio.

En teorías de este tipo se deben atender dos aspectos fundamentales: El aspecto estático que comprende el problema de la representación del conjunto de enunciados y su relación con el estado de creencias del agente. Y el aspecto dinámico correspondiente a la represen-

tación de las operaciones de cambio, de forma tal que acuerden con nuestros principios o criterios.

Las operaciones de cambio en estos modelos son susceptibles de ser presentadas, cada una, de acuerdo con un conjunto de postulados, los que permiten precisar o determinar, a modo de axiomas o principios en general, el comportamiento esperado de ellas; los que están a su vez asociados a principios de tipo general considerados guías del cambio racional de creencias. La definición de operaciones de cambio de creencias que resulten satisfactorias para un contexto con inconsistencias deberá, según los criterios expresados, poder ser caracterizada mediante un conjunto de postulados en gran medida equivalentes a los de sus contrapartes "clásicas".

Continuando los trabajos en los que proponíamos operaciones de cambio de creencias para contextos inconsistentes, presentaremos en este caso una caracterización del cambio de creencias para contextos con inconsistencias en términos de postulados, a fin de observar allí sus características más importantes, su relación con los postulados clásicos y con los principios generales que conforman la teoría de cambio de creencias en general.

Como ya se expresó, contraer una base de creencias por un enunciado consiste en extraer de la base aquellos elementos que implican el enunciado a contraer.

$$K - r = K \setminus (\sigma(K \# r))$$

Siempre que  $A \in (K \# r)$  si y sólo si:

$$A \setminus r$$

Si  $C \setminus r$  entonces  $C \notin A$

$$A \subseteq K$$

Y tenemos además que  $\sigma$  es una función llamada de incisión tal que:

$$\sigma(K \# r) \subseteq \cup(K \# r) \text{ y,}$$

si  $\emptyset \neq A \in (K \# r)$  entonces  $A \cap \sigma(K \# r) \neq \emptyset$

En el caso que nos ocupa, si  $K = \{a \rightarrow b, a, \sim b, r\}$

$$(K \# r) = \{\{a \rightarrow b, a, \sim b\}, \{r\}\}$$

por ejemplo puede ser  $\sigma_1(K \# r) = \{a, r\}$

$$\sigma_2(K \# r) = \{b, r\}$$

$$\sigma_3(K \# r) = \{\sim a, r\}$$

Pero lo que es seguro es que deberá eliminarse un elemento del conjunto inconsistente  $\{a \rightarrow b, a, \sim b\}$  en cualquier opción.

La opción pensada para obtener una dinámica que respete las restricciones pensadas para el cambio de creencias con inconsistencias es la siguiente:

$$K - r = K \setminus (\sigma(K \Downarrow r))$$

$A \in (K \Downarrow r)$  si y sólo si:

$$A \setminus r$$

Si  $C \setminus r$  entonces  $C \notin A$

$$A \subseteq K$$

$$\perp \notin Cn(A)$$

$\sigma$  es una función de incisión tal que.

$$\sigma(K \Downarrow r) \subseteq \cup(K \Downarrow r) \text{ y,}$$

$$\text{si } \emptyset \neq A \in (K \Downarrow r) \text{ entonces } A \cap \sigma(K \Downarrow r) \neq \emptyset$$

$$(K \Downarrow r) = \{\{r\}\} \text{ y } \sigma(K \Downarrow r) = \{r\} \text{ y}$$

$$K - r = K \setminus (\sigma(K \Downarrow r)) = \{a \rightarrow b, a, \sim b\}$$

Se obtiene así una base de creencias que no contiene el elemento a contraer, pero que si conserva los subconjuntos inconsistentes de la base original que sólo implican clásicamente y por efecto del principio *ex falso sequitur quodlibet*, el elemento a contraer.

Aunque el tratamiento propuesto logra neutralizar la expulsión de los elementos inconsistentes, no todos los conflictos en este contexto se limitan a las contradicciones lógicas como la antes descrita (el paciente trabaja con cal y no trabaja con cal -  $\{b, \sim b\}$ ). También podemos identificar otro tipo de contradicciones que los autores llaman no-lógicas que toman la forma:

*d* y *e* son incompatibles.

El tipo de restricción introducido con el concepto de incompatibilidad es más débil que el de contradicción. Este último corresponde a vinculaciones de situaciones en forma general y tal como caracterizan una actividad. Las restricciones introducidas por las incompatibilidades responden a circunstancias específicas, son introducidas en cada modelo particular posiblemente mediante la opinión de expertos, y por ello deben ser variables como lo es el conjunto de creencias del agente al que van asociadas.

Supongamos que a la base que expresaba que: *si el paciente es un albañil trabaja con cal* ( $a \rightarrow b$ ), que *no trabaja con cal* ( $\sim b$ ) y que *trabaja con cal* ( $b$ ). Lo que representamos de la siguiente manera:

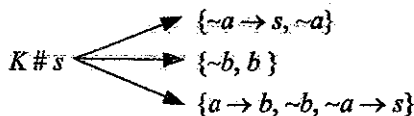
$$K = \{a \rightarrow b, \sim b, b\}$$

Le incorporamos la información siguiente: *si el paciente no es albañil, entonces tiene problemas de circulación* ( $\sim a \rightarrow s$ ), se nos informa que *el paciente no es albañil* ( $\sim a$ ), y además que *no ser albañil es incompatible con trabajar con cal* ( $b$ ), lo que representaremos mediante la relación  $\sim a \perp b$  ( $\sim a$  es incompatible con  $b$ ).

$$K = \{a \rightarrow b, \sim b, b, \sim a \rightarrow s, \sim a\}$$

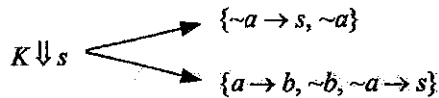
Nos informan además que el paciente *no tiene problemas de circulación*, lo que nos llevaría en primera instancia a eliminar  $s$  de la base de creencias.

En el enfoque clásico esto nos produciría uno de los siguientes conjuntos como resultado de esta operación:



En donde las opciones de cambio mínimo que satisfagan los requisitos de la función de selección serán  $\{\sim b, \sim a\}$  y  $\{b, \sim a \rightarrow s\}$ .

En nuestra propuesta tolerante a las inconsistencias:



Desaparece la exigencia de eliminar algún elemento del par contradictorio  $\{\sim b, b\}$  y las opciones de selección son

- $\sigma_1(K \Downarrow s) = \{\sim a \rightarrow s\}$
- $\sigma_2(K \Downarrow s) = \{\sim a \rightarrow s, \sim a\}$
- $\sigma_3(K \Downarrow s) = \{a \rightarrow b, \sim a \rightarrow s\}$
- $\sigma_4(K \Downarrow s) = \{\sim b, \sim a \rightarrow s\}$
- $\sigma_5(K \Downarrow s) = \{a \rightarrow b, \sim a\}$
- $\sigma_6(K \Downarrow s) = \{\sim b, \sim a\}$

Esta indeterminación quedaría fácilmente resuelta si atendemos a la información provista por la relación de incompatibilidad: si  $\sim a$  es incompatible con  $b$ , esto debe ser interpretado como que, estos elementos, y todo lo que de ellos se sigue, están previamente sospechados en presencia del otro elemento.

Estas soluciones, no cambian la noción de consecuencia clásica por otra (por ejemplo paraconsistente) sino que se limitan a restringir la aplicación de ésta, mediante los recursos formales expuestos, en aquellos casos en que podemos tener la presencia de inconsistencias o incompatibilidades.

De algún modo se puede decir, que el cambio de un estado de creencia a otro no se apoya sólo en la lógica sino en criterios de índole más general, que son propios de cada actividad o tipo de cambio que se este representando. Si es con inconsistencias o no en nuestro caso. Más aún, los criterios propiamente lógicos son de tipo convencional y no corresponden a cualquier ámbito.

La opción pensada para obtener una dinámica que respete las restricciones pensadas para el cambio de creencias con inconsistencias es la siguiente:

$$K - r = K \setminus (\sigma(K \Downarrow r))$$

$A \in (K \Downarrow r)$  si y sólo si:

$$A \vdash r$$

Si  $C \vdash r$  entonces  $C \not\subseteq A$

$$A \subseteq K$$

$$\perp \notin Cn(A)$$

$\sigma$  es una función de incisión tal que:

$$\sigma(K \Downarrow r) \subseteq \cup(K \Downarrow r) \text{ y,}$$

si  $\emptyset \neq A \in (K \Downarrow r)$  entonces  $A \cap \sigma(K \Downarrow r) \neq \emptyset$

$$(K \Downarrow r) = \{\{r\}\}, \sigma(K \Downarrow r) = \{r\} \text{ y}$$

$$K - r = K \setminus (\sigma(K \Downarrow r)) = \{a \rightarrow b, a, \sim b\}$$

Esta operación no satisface el postulado de consistencia para la contracción:

**Consistencia:**  $\perp \notin K - \alpha$

Pero tampoco el de éxito:

**Éxito:** Si  $\alpha \notin \text{Cn}(\emptyset)$ , entonces  $\alpha \notin \text{Cn}(K - \alpha)$

El postulado de éxito que resulta controvertido en el caso de la operación de revisión (puesto que no está claro qué es una revisión 'exitosa') resulta ineludible en alguna formulación en el caso de la operación de contracción.

Si se trata de una operación que debe retirar o quitar elementos de un conjunto de creencias o mejor dicho lograr un cambio tal que estos elementos (los elementos a contraer) dejen de inferirse, entonces los mencionados elementos deben dejarse de inferir después de realizada la operación, salvo casos extremos.

Un caso extremo está contemplado en la formulación ortodoxa del postulado: que se pretenda dejar de inferir una tautología lo que es imposible por el marco superclásico en el que está pensado el modelo (y que por lo demás resulta indeseable por las consecuencias contraintuitivas que tendría si bien su existencia resulta cuestionable también por referir al poco realista principio de omnisciencia epistémica de los agentes)

El otro caso extremo es el de las creencias que son autoinconsistentes. Precisamente, todos los elementos inconsistentes (en especial los que son autoinconsistentes) eran los primeros en resultar eliminados del conjunto de creencias tras una contracción (cualquier contracción) clásica. Tras la modificación operada, la nueva operación de contracción tolerante a las inconsistencias permite conservar elementos inconsistentes a excepción que explícitamente se los quiera eliminar, esta excepción no se aplica a los elementos autoinconsistentes  $\alpha$  tales que  $\alpha \leftrightarrow \perp \in \text{Cn}(\emptyset)$ , puesto que estos quedarán excluidos de su consideración al considerar los subconjuntos consistentes mínimos que implican el enunciado a eliminar y no serán removido, por ellos quedar satisfecha una versión más débil del postulado:

**Éxito- $\perp$ :** Si  $\alpha \notin \text{Cn}(\emptyset)$  y  $\alpha \leftrightarrow \perp \in \text{Cn}(\emptyset)$ , entonces  $\alpha \notin \text{Cn}(K - \alpha)$

O deberá en su defecto redefinirse la operación de manera tal que fuerce la eliminación de elementos mediante su unión al conjunto a ser 'recortado' del conjunto original  $K - \alpha = K \setminus (\sigma(K \downarrow \alpha) \cup \{\alpha\})$  opción que no ha sido explorada en este trabajo.

Hemos visto en este trabajo que el carácter usualmente descripto como 'explosivo', propio de la operación de consecuencia clásica, no debe ser trasladado innecesariamente a la dinámica apoyada en ella. Le dimos prioridad al criterio de la máxima información sobre el de la consistencia.

Soluciones de este tipo tienen la ventaja de que no cambian la noción de consecuencia clásica por otra (por ejemplo paraconsistente) sino que se limitan a restringir la aplicación de ésta, mediante los recursos formales expuestos, en aquellos casos en que podemos tener la presencia de inconsistencias o incompatibilidades.

## Notas

<sup>1</sup> Alchourrón, Gärdenfors & Makinson (1985).

<sup>2</sup> En el caso de una contracción de intersección parcial (*partial meet contraction*).

<sup>3</sup> S.O. Hansson y R. Wassermann (1998).

<sup>4</sup> Y que además es la única operación en la presentación clásica de AGM que es tolerante de las inconsistencias, ya que es compatible con la posibilidad de que  $\{\alpha, \sim\alpha\} \subseteq K \subseteq K + \beta$ .

## Bibliografía

- Alchourrón, C., Gärdenfors, P. y Makinson, D. (1985), "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions", *Journal of Symbolic Logic* 50, 510-530.
- Barwise, J. (1989), *The situation in logic*, Stanford: CSLI Lecture Notes.
- Gärdenfors, P. (1988), *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, Cambridge, Mass.: Bradford/MIT.
- Hansson S. O. y Wassermann R. (1998) "Local Change: A Preliminary Report", versión preliminar aparecida en *Fourth Symposium on Logical Formalizations of Commonsense Reasoning*.
- Harper, W. (1977), "Rational Conceptual Change", en *PSA 1976*, pp. 462-494.
- James, W. (1907), *Pragmatism – A New Name for Some Old Ways of Thinking*, New York: Longmans, Green & Co.
- Levi, I. (1980), *The Enterprise of Knowledge*, Cambridge, Mass.: The MIT Press.
- Peirce, C.S. (1982) *The Fixation of Belief (1877)*, en *Writings of Charles S. Peirce*, 1, M. Fisch (ed.), Bloomington: University of Indiana Press.