

La abucción como cambio epistémico: C. S. Peirce y las teorías epistémicas en inteligencia artificial

Atocha Aliseda

UNAM, México
Analogía 12 (1998), 125-144

*Doubt, usually, perhaps always,
takes its rise from surprise
(CP 5.166, 1903)*

1 Introducción

El proyecto intelectual de Charles S. Peirce (1839-1914), en un sentido muy amplio, consistió en desarrollar una teoría semiótica para explicar el pensamiento y el lenguaje. Fuertemente influenciado por la filosofía de Immanuel Kant, Peirce propone extender sus categorías y corregir su lógica. Con respecto a nuestros fines, la pregunta fundamental que le preocupa concierne al razonamiento sintético:

“De acuerdo con Kant, la pregunta central de la filosofía es ‘¿Cómo son posibles los juicios sintéticos *a priori*?’ Pero anteriormente a ésta, surge la pregunta acerca de cómo son posibles los juicios sintéticos en general y, de modo aún más general, cómo es posible en absoluto el razonamiento sintético. Cuando se haya obtenido respuesta al problema general, el particular será comparativamente más simple. Ésta es la llave de la puerta de la filosofía” (CP 5.348, 1868).

El análisis que Peirce hace de este tipo de razonamiento tiene fundamentalmente dos fines: por un lado, justificar la posibilidad del razonamiento sintético, y por el otro, desarrollar un método para su adquisición. La respuesta a lo primero la encuentra Peirce en su metafísica; lo segundo es justamente lo que nos ocupa en este trabajo, a saber, su *lógica abductiva* (cf. Hookway 1992 para un análisis del primer aspecto).

Peirce propone la abducción como lógica del razonamiento sintético: un modo de inferencia *ampliativo* cuyo resultado añade ideas nuevas al conocimiento¹. Sin

¹En contraste, la deducción es analítica; es un modo de inferencia explicativa, ya que la información de la conclusión está ya sugerida en las premisas. La inducción es también considerada por Peirce como un modo de inferencia ampliativa, aunque no puramente sintética (CP 5.170, 1903).

embargo, cabe señalar que las nociones de inferencia y validez lógica propuestas por Peirce van más allá del entendimiento actual de lo que es la lógica, están ligadas a su epistemología, a una visión dinámica del pensamiento como indagación lógica. Por lo tanto, para un entendimiento de la lógica abductiva es necesario un estudio a fondo del sistema filosófico peirceano.

En este trabajo nuestro interés fundamental es concentrarnos en la forma lógica de la abducción propuesta por Peirce, en particular, en el aspecto que el elemento *sorpresa* juega en ella. Nuestro objetivo consiste en proponer a la abducción como un proceso epistémico de adquisición de conocimiento. Una consecuencia natural de esta interpretación es que la forma lógica de la abducción es la de un proceso y no la de un argumento, como generalmente se ha entendido. Esta interpretación tiene una conexión natural con las teorías de cambio epistémico en inteligencia artificial, en particular, con la teoría propuesta por Peter Gardenfors (Gardenfors 1988) en el campo de representación del conocimiento en inteligencia artificial (IA).

El artículo está organizado de la siguiente manera: la sección 2 la dedicamos a Peirce; ofrecemos una reconstrucción de su noción de abducción así como de los elementos epistémicos que modelan al pensamiento como acción dinámica. En la sección 3 presentamos a la abducción en el campo de la IA y ofrecemos una descripción breve de las teorías epistémicas. En la sección 4 proponemos a la abducción como cambio epistémico y presentamos un modelo que combina elementos de Peirce y de la IA. Finalmente en la sección 5 ofrecemos nuestras conclusiones.

2 Charles S. Peirce

2.1 El desarrollo de la noción de abducción

Charles S. Peirce fue el primer filósofo que propuso una forma lógica para el razonamiento abductivo. Sin embargo, su noción de abducción es muy difícil de descifrar. Por un lado, en su pensamiento filosófico se desarrollaron diversas versiones, y por el otro, su noción de abducción está enredada con muchos otros aspectos de su filosofía que son igualmente complejos. Tocaremos algunos puntos claves en el desarrollo de su noción de abducción para luego concentrarnos en la forma lógica.

El desarrollo de una *lógica de la indagación* ocupó el pensamiento de Peirce desde el inicio de su trabajo intelectual. En un principio esta lógica está compuesta por tres modos de razonamiento: *deducción, inducción e hipótesis*², cada uno de los cuales es un proceso independiente de prueba y corresponde a una forma silogística, que ilustramos en el siguiente ejemplo, frecuentemente citado (CP 2.623, 1877):

²Las diferentes concepciones de la abducción van acompañadas de diversa terminología. Al principio Peirce usó *hipótesis* (CP 2.623, 1878) y pasó luego a usar *abducción* y *retroducción* (CP 1.68, 1896; 2.776, 1901; 7.97, 1910).

Deducción

Regla	Todas las alubias de este saco son blancas.
Caso	Estas alubias son de este saco.
Resultado	Estas alubias son blancas.

Inducción

Caso	Estas alubias son de este saco.
Resultado	Estas alubias son blancas.
Regla	Todas las alubias de este saco son blancas.

Hipótesis

Regla	Todas las alubias de este saco son blancas.
Resultado	Estas alubias son blancas.
Caso	Estas alubias son de este saco.

De estos tres, la deducción es el único tipo de razonamiento completamente certero, que infiere su ‘Resultado’ como conclusión necesaria. La inducción produce una ‘Regla’ que se valida solamente “a la larga” (CP 5.170, 1903), y la hipótesis, la menos certera de las tres, simplemente sugiere que algo puede ser ‘el caso’ (CP 5.171, 1903).

Posteriormente Peirce considera a estos tres tipos de razonamiento como tres etapas en un método para la indagación lógica, en donde la hipótesis, ahora denominada abducción, es la primera de ellas: De su sugerencia (abductiva), la deducción puede inferir una predicción que puede ser puesta a prueba mediante la inducción. (CP 5.171, 1903). La noción de abducción se hace más compleja y se convierte en: “*el proceso de construir una hipótesis explicativa*” (CP 5.171, 1903) y la forma silogística se complementa con la siguiente forma lógica:

“Se observa un hecho sorprendente, *C*; Pero si *A* fuera verdadera, *C* sería una cosa corriente, Por lo tanto, hay razón para sospechar que *A* es verdadera” (CP 5.189, 1903).

Peirce propone dos aspectos más para determinar qué tan prometedora es una hipótesis abductiva: se debe poder *poner a prueba* y debe ser *económica*. Así, una abducción es una explicación si da razón de los hechos conforme a la forma lógica arriba citada; su estatus es el de una sugerencia hasta que no se pone a prueba, lo cual explica el segundo criterio. Las motivaciones del criterio de economía son dos: la respuesta al problema práctico de manejar un sinnúmero de hipótesis explicativas, así como la necesidad de contar con un criterio para seleccionar la mejor explicación dentro de las que son verificables.

Para Peirce, el razonamiento abductivo es fundamental en toda pesquisa humana. La abducción juega un papel en la percepción: “La sugerencia abductiva nos viene como un destello” (CP 5.181, 1903) y también está presente en el proceso general de la invención: “Ella [la abducción] es la única operación lógica que incorpora nuevas ideas” (CP 5.171, 1903).

Así, la abducción parece ser tanto “*un acto de intuición como uno de inferencia*”, como ha propuesto Anderson (1986), quien sugiere un doble aspecto en la abducción, el intuitivo y el racional.

2.2 Interpretaciones de la abducción peirceana

La contraposición de los dos aspectos de la abducción, el intuitivo y el racional, ha confundido invariablemente a los estudiosos de Peirce. En general, se toma uno de los dos para su análisis (Kapitan 1990, Sharpe 1970, Thagard 1977). Algunos críticos han interpretado esta dualidad como “el dilema de Peirce” y concluyen que Peirce no tenía una visión coherente sobre la naturaleza de la abducción (Frankfurt 1958). Existe también la postura que se une a la arriba mencionada, que trata de dar sentido a estos dos aspectos y propone a la abducción peirceana como “*el instinto racional*” (Ayim 1974).

Con respecto a la forma lógica de la abducción, mientras que algunos estudiosos de Peirce le han dado un análisis que la identifica con la inducción (Reilly 1970), otros han preferido darle la interpretación de *modus ponens* invertido (Anderson 1986); y finalmente otros la han visto como una forma de heurística (Kapitan 1990).

Hasta aquí dejamos nuestra reconstrucción de la evolución de la noción de abducción en Peirce. Para un estudio muy detallado, donde claramente se distinguen tres etapas en la evolución de la noción de abducción, se recomienda consultar el trabajo de K. T. Fann (1970). Otra referencia clave es la de Anderson (1987) para un análisis sobre abducción y creatividad.

2.3 La epistemología en Peirce

En la epistemología de Peirce, el pensamiento es un proceso dinámico, esencialmente una acción que oscila entre los estados mentales de *duda* y *creencia*. Mientras que la esencia del segundo es la “*instauración de un hábito que determina nuestras acciones*” (CP 5.388, 1877), con la cualidad de ser un estado satisfactorio y apacible en el que todo humano quisiera permanecer, el primero “*nos estimula a indagar hasta autodestruirse*” (CP 5.373, 1877) y se caracteriza por ser un estado turbulento e insatisfactorio del que todo humano lucha por liberarse: “la irritación de la duda provoca una contienda para alcanzar el estado de creencia” (CP 5.374, 1877)

Nótese que Peirce habla de estado de creencia y no de conocimiento. Así, la pareja ‘duda-creencia’ es en realidad un ciclo entre dos estados diametralmente opuestos; mientras que la creencia es un hábito, la duda es la privación del mismo. Sin embargo, la duda, escribe Peirce, no es un estado que se genere voluntariamente haciendo una pregunta, así como una oración no es interrogativa por el simple hecho de ponerle una marca especial, debe existir una duda *real* y *auténtica*:

“La duda auténtica siempre tiene un origen externo, usualmente viene de la *sorpresa*; siendo tan imposible para un hombre producir una duda

auténtica voluntariamente como lo sería el imaginarse la condición de un teorema matemático, y como sería el generar una sorpresa mediante un simple acto de voluntad” (CP 5.443, 1905).

De esta forma, Peirce no sólo argumenta que para romper un hábito debe existir una duda auténtica, sino que la identifica con la sorpresa, y de hecho, parece usar estos dos términos indistintamente: “La creencia, mientras dura, es un hábito fuerte, y como tal, fuerza al hombre a creer hasta que una *sorpresa* rompe el hábito”. Más aún, Peirce distingue dos formas de romper el hábito: “El rompimiento de una creencia solo puede deberse por una *experiencia novedosa*” (CP 5.524, c.1905) o “. . . hasta que nos encontramos a nosotros mismos confrontados con alguna *experiencia contraria a las expectativas*” (CP 7.36, c.1907; el énfasis es mío).

“La sorpresa tiene sus variedades activa y pasiva; la primera se da cuando uno percibe positivamente *conflictos* dada una expectativa, la segunda cuando no se tiene una expectativa positiva sino sólo la ausencia de cualquier sospecha de que ocurra algo fuera de lo común totalmente inesperado, —como un eclipse total de sol que uno no había previsto” (CP 8.315, 1909).

El modelo epistémico peirceano propone a la sorpresa como el detonante de toda pesquisa; sorpresa que puede darse por *novedad* o por *anomalía*. En Aliseda (1997) he denominado estos dos aspectos los *detonadores abductivos*. Más adelante veremos cómo se relacionan con el enfoque de la abducción en IA.

2.4 Abducción y epistemología

La conexión entre la lógica abductiva y la transición epistémica entre los estados mentales de duda y creencia se ve muy claramente en el hecho de que la *sorpresa* sea la detonadora tanto del razonamiento abductivo, —como lo indica la primera premisa de la formulación lógica de la abducción—, como del estado de duda al romper un hábito de creencia.

El proceso cognitivo que integra a la inferencia abductiva con el proceso epistémico puede describirse como sigue: una experiencia novedosa o anómala da lugar a un *hecho sorprendente*, el cual genera un estado de duda que rompe un hábito de creencia, y así *dispara* el razonamiento abductivo. Este consiste justamente en explicar el hecho sorprendente y así “apaciguar” el estado de duda. Digo *apaciguar* y no *destruir* porque una explicación abductiva no necesariamente desemboca en una creencia. La explicación abductiva es simplemente una sugerencia que debe ser puesta a prueba antes de convertirse en creencia.

3 Inteligencia artificial

3.1 Abducción

Las investigaciones sobre la abducción en el campo de la inteligencia artificial se remontan a 1973 (Pople 1973), pero no es hasta fechas muy recientes cuando este tema ha cobrado gran interés. La formulación lógica de Peirce ha sido el punto de partida de muchos estudios en diversos campos de la IA, entre los que se cuentan los siguientes: en programación lógica (Kakas et al 1995), en asimilación del conocimiento (Kakas 1990) y en procesamiento del lenguaje natural (Hobbs et al 1990). Sin embargo, estos enfoques no han prestado atención suficiente a los elementos de la formulación abductiva y menos aún a lo que Peirce propuso en otros escritos.

En general, la interpretación que se da a la formulación de Peirce es la del argumento lógico:

$$\frac{C}{\frac{A \rightarrow C}{A}}$$

En donde el *status* de la conclusión A es *tentativo*. Se asume que la segunda premisa $A \rightarrow C$ es parte de la teoría de trasfondo θ en cuestión y el criterio económico peirceano se incorpora como un proceso posterior para seleccionar “*la mejor explicación*”, ya que puede haber muchas fórmulas que cumplan con esta forma lógica sin ser por ello hipótesis idóneas. Cabe decir, sin embargo, que en IA sólo se reconoce como peirceana la forma lógica de la abducción y no así los criterios adicionales de verificación y economía.

A pesar de resultar una interpretación muy intuitiva, esta formulación lógica no refleja ni el que C sea sorprendente, ni ninguno de los otros criterios que Peirce propuso. Más aún, nótese que la interpretación de la segunda premisa no tiene porque verse como una implicación material clásica, pues podría ser una implicación lógica de otra naturaleza o incluso un proceso computacional en donde A sea la entrada y C la salida.

Es imposible ofrecer un análisis detallado del desarrollo de la abducción en todos los campos de la IA. Nos limitaremos a dar una descripción breve sobre la abducción como inferencia lógica, para luego pasar a describir las teorías epistémicas en inteligencia artificial y finalmente establecer la conexión entre estas dos³.

3.2 Abducción como inferencia lógica

La interpretación más usual de la abducción como inferencia lógica es como *deducción para atrás más condiciones adicionales*. A continuación damos la versión

³Existen muchos otros enfoques abductivos, entre los que se cuentan estudios en redes bayesianas, conexionismo, y muchos otros enfoques computacionales no lógicos, en los que el énfasis se encuentra más bien en la construcción de algoritmos para generar abducciones.

estándar que combina algunos requisitos comunes a diversos enfoques (Mayer et al 1993, Konolige 1996, Kakas 1995):

Dada una teoría θ (un conjunto de fórmulas) y una fórmula ϕ (una fórmula atómica), α es una explicación abductiva si:

1. $\theta, \alpha \rightarrow \phi$
2. α es consistente con θ
3. α es mínima
4. α tiene una restricción en su forma sintáctica (usualmente de fórmula atómica).

Así, una fórmula α es una explicación de ϕ con respecto a una teoría de trasfondo θ si se infiere a ϕ como consecuencia lógica (condición 1) y la explicación es consistente con la teoría (condición 2). Además, la mejor explicación es la mínima en el sentido de que es la fórmula más simple (condición 3), lo que en general se interpreta como aquella que es implicada por todas las otras posibles explicaciones. Sin embargo, no existe una sola caracterización de minimalidad lógica (cf. Aliseida 1997 para una discusión sobre la noción de minimalidad como criterio lógico). Finalmente, la condición 4 es una restricción que evita que formulas triviales como $\theta \rightarrow C$ cuenten como explicaciones abductivas.

Para ilustrar esta definición, veamos el ejemplo más común de abducción en IA:

Sea θ la siguiente teoría:

El patio se moja cuando llueve ($ll \rightarrow m$)

El patio se moja cuando los aspersores de agua están prendidos ($a \rightarrow m$)

Sea ϕ : El patio está mojado (m).

Si solo consideramos la condición 1, las siguientes son algunas de las fórmulas que cumplen con la condición de consecuencia lógica: ll , a , $ll \wedge a$, $ll \wedge \neg ll$, $\theta \rightarrow \phi$. La condición 2 de consistencia es necesaria ya que en la noción de consecuencia lógica clásica tarskiana de un conjunto inconsistente de premisas se deriva lo que sea. Así, tomamos en cuenta todas las condiciones, eliminamos las explicaciones inconsistentes ($ll \wedge \neg ll$) y las de forma compuesta ($ll \wedge a$, $\theta \rightarrow \phi$), quedándonos tan solo con las dos primeras (ll, a) como explicaciones posibles.

Una condición adicional, que no se hace siempre explícita, es que $\neg(\theta \rightarrow \phi)$ y $\neg(\theta \rightarrow \neg\phi)$. Esto dice que ni el hecho a explicar ni su negación deben inferirse de la teoría. Esta condición aparece a veces como precondition para un *problema abductivo*.

La abducción como inferencia lógica es sólo una de las múltiples interpretaciones en inteligencia artificial. Flach (1996) sugiere que esta multiplicidad se

remonta a las dos teorías de la abducción en Peirce: la temprana teoría silogística y la tardía teoría inferencial, que corresponden a los dos períodos del desarrollo de la abducción en Peirce (cf. sección 2.1).

3.3 Teorías epistémicas en inteligencia artificial

La motivación principal de las teorías epistémicas en inteligencia artificial es la de desarrollar mecanismos lógicos y computacionales para incorporar nueva información a una teoría científica, base de datos o conjunto de creencias. Diferentes tipos de cambio epistémico son apropiados en diferentes situaciones. El trabajo pionero que cabe mencionar en este sentido es el de Peter Gardenfors (Gardenfors 1988), quien propuso una teoría normativa de cambio epistémico caracterizada por las condiciones que un operador de creencia debe satisfacer.

Los elementos fundamentales de estas teorías son los siguientes. Se parte de una teoría θ , consistente y cerrada bajo consecuencia lógica, llamada el “estado de creencia” y de una oración ϕ , la creencia de entrada a incorporar en la teoría. Existen tres *actitudes epistémicas* de θ con respecto a ϕ : es aceptada ($\phi \in \theta$), rechazada ($\neg\phi \in \theta$), o es indeterminada ($\phi \in \theta, \neg\phi \notin \theta$), es decir, la teoría no tiene ninguna postura con respecto a ella. Además, se definen tres operaciones para incorporar ϕ en θ y así modificar el conjunto de creencias. Estas son:

Expansión: ($\theta + \phi$)

Una oración aceptada o indeterminada ϕ se añade a θ .

Revisión: ($\theta * \phi$)

Para poder incorporar a θ una fórmula ϕ que es rechazada y asegurar consistencia en el sistema de creencias resultante, primero se borran las fórmulas de θ que están en conflicto con ϕ y solo entonces se puede agregar ϕ a θ .

Contracción: ($\theta - \phi$)

La fórmula ϕ de θ se borra junto con algunas otras fórmulas que la implican.

De estas operaciones, la revisión es la más compleja de las tres. De hecho se puede definir como la composición de las otras dos: primero borra las creencias de θ que estén en conflicto con ϕ y luego expande la teoría modificada con ϕ . Mientras que la expansión tiene una caracterización única, este no es el caso ni para la contracción ni para la revisión, ya que diversas fórmulas se pueden borrar y obtener el mismo efecto. Veamos el siguiente ejemplo:

Sea θ la siguiente teoría:

El patio se moja cuando llueve ($ll \rightarrow m$)

Llueve (ll)

Sea ϕ : El patio no está mojado ($\neg m$).

Para poder incorporar $\neg m$ en la teoría sin alterar su consistencia, la teoría debe ser revisada. Pero hay dos posibilidades para este fin: borrar cualesquiera de $l \rightarrow m$ ó l nos permitiría expandir la teoría revisada con $\neg m$ consistentemente. Es claro que con la operación de contracción no se puede determinar en puros términos lógicos cuál de las dos debe ser borrada de la teoría. Por lo tanto, se necesita un criterio adicional para tal efecto⁴.

Esta es la base de las teorías epistémicas en inteligencia artificial. En la práctica, sin embargo, estos sistemas son muy diversos. Difieren al menos en tres aspectos: (a) la representación de sus estados de creencia (conjuntos, bases o mundos posibles), (b) operaciones de cambio epistémico (postulados o métodos constructivos), y (c) su posición epistémica con respecto a la cualidad epistémica a preservar. Mientras que los “fundamentalistas” argumentan que toda creencia debe estar justificada, los “coherentistas” consideran que lo más importante es mantener la consistencia del sistema sin importar si las creencias están o no justificadas.

Hasta aquí dejamos la breve descripción de las teorías de cambio epistémico en inteligencia artificial.

3.4 Abducción y teorías de cambio epistémico

La conexión entre la abducción y las teorías de cambio epistémico ya se ha estudiado en la literatura. En general se considera el papel de la abducción como el mecanismo para generar las explicaciones de las creencias a incorporar, proceso necesario de acuerdo con el enfoque fundamentalista.

Nuestro argumento con respecto a esta conexión es aún más fuerte: el razonamiento abductivo es una forma de cambio epistémico por medio del cual puede modelarse la incorporación de nuevas creencias. Aunque está claro que en este contexto una oración de entrada no es una observación, sino una creencia para la que se busca una explicación. Veamos algunas razones que justifican este argumento.

En primer lugar, analicemos a que corresponden los detonadores abductivos (cf. 2.2) en el modelo epistémico. La condición de *novedad* puede formalizarse como sigue:

$$\neg(\theta \notin \phi) \text{ y } \neg(\theta \notin \neg\phi)$$

ϕ es novedoso con respecto a θ ya que no se infiere ni él ni su negación. Este caso corresponde a la condición de indeterminación epistémica:

$$\phi \notin \theta \text{ y } \neg\phi \notin \theta$$

ya que θ es cerrada bajo consecuencia lógica. En otras palabras, el hecho ϕ es sorprendente justamente porque la teoría no tiene ninguna posición acerca de él.

⁴Hay varias formas de lidiar con este problema. En Gardenfors (1988) se propone la noción de “atrincheramiento”, un orden preferencial que lista a las fórmulas de acuerdo con su importancia epistémica.

El otro detonador abductivo, el de anomalía, puede expresarse formalmente de la siguiente manera:

$$\neg(\theta \notin \phi) \text{ y } (\theta \notin \neg\phi)$$

Esto es, la teoría no explica el hecho sorprendente sino más bien a su negación. En este caso, la actitud epistémica correspondiente es la de rechazo:

$$\neg(\phi \in \theta)$$

Además de esta correspondencia, con la abducción se pueden modelar las operaciones epistémicas para incorporar creencias, pero esto lo veremos en detalle en la propuesta que a continuación presentamos.

4 Propuesta: la abducción como cambio epistémico

Nuestro argumento en este artículo consiste en interpretar al razonamiento abductivo como cambio epistémico. Para ello combinamos elementos lógicos y epistémicos de Peirce y de propuestas en IA expuestos en las secciones anteriores. En Aliseda (1997) se propone una taxonomía para la abducción, dentro de la cual se enmarca el argumento que ahora nos ocupa. A continuación presentamos brevemente esta taxonomía para luego concentrarnos en los elementos esenciales de nuestro argumento.

4.1 La taxonomía de la abducción

La abducción es un proceso de razonamiento, cuyo producto son explicaciones con cierta estructura inferencial. El esquema lógico de la abducción puede verse como una relación entre tres elementos:

$$\theta, \alpha \Rightarrow \phi$$

la observación o creencia ϕ (p. e., m), la explicación abductiva α (p. e., ll), y la teoría θ de trasfondo (p. e., ll \rightarrow m, a \rightarrow m). Además, he propuesto tres parámetros principales que determinan tipos de abducción. (i) el parámetro inferencial (\Rightarrow) que determina la relación lógica entre el *explanandum* ϕ y el *explanans* (la teoría de trasfondo θ y la explicación α). Esta relación puede ser la de consecuencia lógica clásica, inferencia estadística o incluso algún tipo de inferencia no clásica. (ii) los detonadores abductivos: ϕ puede ser un fenómeno novedoso, como ver el patio mojado en el ejemplo 1, o ϕ bien puede estar en conflicto con la teoría, como el observar que el patio no está mojado (\neg m), en el ejemplo 2. (iii) Finalmente, las ‘salidas’ α son los diversos productos de los procesos abductivos: hechos, reglas e incluso teorías. Además de estos parámetros, las condiciones adicionales de consistencia y minimalidad, se toman en cuenta para ciertos tipos de abducción (cf.

Aliseda 1997 para un análisis de caracterizaciones lógicas de diversas *lógicas abductivas* que se generan al instanciar estos parámetros y condiciones adicionales).

Esta taxonomía generaliza la definición estándar de abducción como inferencia lógica en IA (cf. 3.1) ya que no se limita a interpretar la relación inferencial como clásica ni tampoco restringe la forma de las explicaciones a átomos. Más importante, sin embargo, es que este esquema general propone a la revisión como caso de razonamiento abductivo. Aunque existen diversos trabajos que consideran este caso como de abducción (e.g. Aravindan et al 1994), no se ha tratado desde un punto de vista lógico. Incorporar la revisión como caso de razonamiento abductivo reconoce que los cambios en la teorías no se dan solo por acumulación de nuevos descubrimientos, sino que se dan también como consecuencia de anomalías. Así, nos acercamos a la noción de abducción peirceana en donde un hecho es sorprendente por ser una experiencia *novedosa* o una *contraria a las expectativas* (cf. 2.2).

En este trabajo, nos hemos concentrado principalmente en el parámetro que concierne a los detonadores abductivos, así que pasamos a describirlo en detalle.

4.2 Detonadores abductivos

Como hemos visto, de acuerdo con la formulación de abducción de Peirce, el razonamiento abductivo se dispara por un *hecho sorprendente*. Sin embargo, la noción de sorpresa es relativa, ya que un hecho es sorprendente sólo con respecto a una teoría θ de trasfondo que provee “expectativas”. Lo que es sorprendente para mi (p. e., que las luces del cuarto de la fotocopidora se prendan en cuanto entro), puede no serlo para ti.

Así, un hecho sorprendente lo interpretamos como aquél que requiere de explicación, esto es, aquél que es novedoso o anómalo, lo que expresamos formalmente como sigue:

Novedad abductiva: ϕ es novedosa. No puede ser explicada $\neg(\theta \Rightarrow \phi)$, ni tampoco su negación $\neg(\theta \Rightarrow \neg\phi)$.

Anomalía abductiva: ϕ es anómala. No puede ser explicada $\neg(\theta \Rightarrow \phi)$, y de hecho la teoría explica a su negación $(\theta \Rightarrow \neg\phi)$.

Los hechos no sorprendentes, aquellos que ya son explicados ($\theta \Rightarrow \phi$) no debieran ser candidatos para explicación. Sin embargo, uno pudiera especular que si la teoría de trasfondo θ infiere a ϕ en forma probabilística, tal vez se pudiera necesitar una explicación de algún tipo para hacer “más certera” la relación entre la teoría θ y el hecho ϕ . Más aún, aunque un hecho ya sea explicado, puede haber una razón para buscar otra explicación más fuerte en algún sentido, con más capacidad explicativa.

A continuación presentamos las operaciones que inducen cada uno de los detonadores abductivos. Estas son las operaciones abductivas de cambio epistémico:

Expansión abductiva: Dada una novedad abductiva ($\neg(\theta \Rightarrow \phi), \neg(\theta \Rightarrow \neg\phi)$), una explicación consistente α se calcula de tal forma que $\theta, \alpha \Rightarrow \phi$. Así se añaden ϕ y α a la teoría θ por medio de una simple expansión.

Revisión abductiva: Dada una anomalía abductiva ($\neg(\theta \Rightarrow \phi), \theta \Rightarrow \neg\phi$), una explicación consistente α se calcula de la siguiente manera: Primero se revisa θ de tal forma que no explique a $\neg\phi$. Esto es, se obtiene θ' de tal forma que $\neg(\theta' \Rightarrow \neg\phi)$ donde $\theta' = \theta - \{\beta_1, \dots, \beta_n\}$ ⁵. Una vez obtenida θ' , se obtiene una explicación α consistente con θ' de tal forma que $\theta', \alpha \Rightarrow \phi$. Así, el proceso de revisión involucra los procesos de contracción y de expansión.

Una vez definidas las nociones anteriores, podemos describir el proceso abductivo de cambio epistémico como sigue: el razonamiento abductivo se dispara por una sorpresa, la cual genera una duda que puede ser de dos tipos: novedad o anomalía. En el primer caso el fenómeno a explicar es totalmente nuevo y consistente con la teoría, por lo que su explicación se calcula y se incorpora a la teoría por la operación de extensión. En el segundo caso, como el hecho es anómalo, la operación de revisión es necesaria para incorporarlo. Así, la teoría se revisa de tal forma que su modificación no esté en conflicto con el hecho a explicar, a continuación se calcula la explicación y se incorpora a la teoría revisada por expansión.

5 Conclusiones

Las nociones de inferencia lógica y abducción que Peirce propuso van más allá de formulaciones lógicas. Estas nociones están ligadas a su epistemología, una visión dinámica del pensamiento como una indagación lógica, y corresponden a una preocupación filosófica muy profunda, la de descifrar la naturaleza del razonamiento sintético. Sin embargo, hemos mostrado que las nociones peirceanas de sorpresa, novedad y anomalía pueden expresarse como condiciones lógicas muy precisas. Con ellas se ha propuesto un modelo que tiene conexiones muy íntimas con enfoques actuales en representación del conocimiento en IA.

Nuestro análisis de la abducción combina elementos epistémicos del modelo peirceano y de enfoques en IA y muestra que la abducción es un fenómeno muy complejo que requiere de un análisis muy minucioso. Por un lado, este tipo de razonamiento puede tomar diversas formas lógicas que dependen de la relación de consecuencia seleccionada para su representación. Además, la forma de la explicación puede ser muy diversa. Por el otro, como argumentamos con detalle en este trabajo, no hay una única forma de razonamiento abductivo; hay al menos dos tipos de detonadores abductivos, novedad y anomalía, y cada uno de ellos induce operaciones diferentes para incorporar la nueva creencia y su explicación a la teoría.

⁵En muchos casos, varias formulas y no solo una deben sacarse de la teoría. P.e., $\theta = \{\alpha \rightarrow \beta, \alpha, \beta\}$ y $\phi = \neg\beta$. Para hacer $\theta, \neg\beta$ consistente, uno debe quitar $\{\beta, \alpha\}$ o $\{\beta, \alpha \rightarrow \beta\}$.

Una consecuencia de este análisis es que la interpretación de la formulación abductiva peirceana va más allá del argumento lógico, como generalmente se ha propuesto. La formulación abductiva toma la forma de un proceso ya que la modificación de la teoría está involucrada.

Referencias

- [1] Aliseda, Atocha (1997) *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. Amsterdam: Institute for Logic, Language and Computation, Universidad de Amsterdam.
- [2] Anderson, Douglas (1986) The Evolution of Peirce's Concept of Abduction. *Transactions of the Charles S. Peirce Society* 22: 145-164.
- [3] Anderson, Douglas (1987) *Creativity and the Philosophy of C.S. Peirce*. Dordrecht: Nijhoff.
- [4] Aravindan, C. y Dung, P. M. (1994) Belief Dynamics, Abduction and Databases. En *Logics in Artificial Intelligence. European Workshop JELIA'94*, editado por C. MacNish, D. Pearce y L. Moniz Pereira, 66-85. New York: Springer-Verlag.
- [5] Ayim, Maryann (1974) Retroduction: The Rational Instinct. *Transactions of the Charles S. Peirce Society* 10: 34-43.
- [6] Fann, Kuang Tih (1970) *Peirce's Theory of Abduction*. La Haya: Nijhoff.
- [7] Flach, P. (1996) Abduction and Induction: Syllogistic and Inferential Perspectives. En *Abductive and Inductive Reasoning Workshop Notes, ECAI'96*, Budapest.
- [8] Frankfurt, Harry G. (1958) Peirce's Notion of Abduction. *The Journal of Philosophy* 55: 593-597.
- [9] Gardenfors, Peter (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, MA: MIT Press.
- [10] Hanson, Norwood Russell (1961). *Patterns of Scientific Discovery*. Cambridge: Cambridge University Press.
- [11] Hobbs, J. R., Stickel, M., Appelt, D. y Martin P. (1990) *Interpretation as Abduction*. Menlo Park, CA: SRI International Technical Note 499.
- [12] Hookway, Christopher (1992) *Peirce*. London: Routledge.
- [13] Kakas, A. y Mancarella, P. (1990). Knowledge Assimilation and Abduction. En *Proceedings of the European Conference on Artificial Intelligence, ECAI'90*. International Workshop on Truth Maintenance, Stockholm: Springer-Verlag.

- [14] kakas, A. y Kowalski, Toni (1995) Abductive Logic Programming. *Journal of Logic and Computation* 2: 719-770.
- [15] Kapitan, Tomis (1990) In What Way is Abductive Inference Creative? *Transactions of the Charles S. Peirce Society* 26: 499-512.
- [16] Konolige, Kurt (1996) Abductive Theories in Artificial Intelligence. En *Principles of Knowledge Representation*, editado por G. Brewka. Stanford, CA: CSLI Publications.
- [17] Mayer, M. C. y Pirri, F. (1993) First Order Abduction Via Tableau and Sequent Calculi. *Bulletin of the IGPL* 1: 99-117.
- [18] Pople, H. E. (1973) On the Mechanization of Abductive Logic. En *Proceedings of the Third International Conference on Artificial Intelligence, IJCAI-73*, 147-152. Stanford, CA.
- [19] Reilly, Francis E. (1970) *Charles Peirce's Theory of Scientific Method*. New York: Fordham University Press.
- [20] Sharpe, Robert (1970) Induction, Abduction, and the Evolution of Science. *Transactions of the Charles S. Peirce Society* 6:17-33.
- [21] Thagard, Paul R. (1977) The Unity of Peirce's Theory of Hypothesis. *Transactions of the Charles S. Peirce Society* 13: 112-123.