

**Peirce desde la inteligencia artificial:
la abducción y la condición de consistencia.¹**

Alejandro Ramírez Figueroa

alamire@uchile.cl

1

En sus *Conferencias sobre el pragmatismo*, Peirce estipula que una “buena” hipótesis inferida abductivamente es aquella que cumple con su fin; y el fin propio de una hipótesis abducida es explicar ciertos hechos observados, y que son considerados extraños. Ello puede interpretarse, como veremos, como una condición de consistencia para la abducción, pero una condición “externa” a la inferencia abductiva misma. Por su parte, los diversos enfoques actuales acerca del razonamiento abductivo, especialmente los que se desarrollan desde la inteligencia artificial, sostienen explícitamente que una de las condiciones de la abducción, necesaria pero en absoluto suficiente, es la consistencia. Sin embargo, en dichos enfoques también se exige la condición de inconsistencia, tal que, incluso aseguran radicalmente algunos autores, tal condición le sería “esencial” al razonamiento abductivo. Esta ponencia, como parte del proyecto general cuya meta es comprender las relaciones entre inferencia y conocimiento, se propone analizar y tratar de resolver esta aparente divergencia de condiciones que regiría el razonamiento de generación de nuevas hipótesis. La tesis central que se defiende aquí es que: a) las condiciones de consistencia e inconsistencia se dan en diferentes niveles en la abducción, por lo que ambas en realidad se complementan. Pero b), además ambos constreñimientos cumplen diferentes roles al interior de la estructura de la abducción, por lo que ninguno de los dos podría eliminarse en beneficio del otro. Al menos *prima facie*, los diversos

¹ Esta ponencia es parte del proyecto de investigación financiado por el Departamento de Investigación de la Universidad de Chile, por el período 2005 a 2006, código SOC 04/01-2. Su presentación en estas II Jornadas ha sido posible gracias al apoyo y patrocinio que presta dicho proyecto.

enfoques sobre la abducción consideran como consistente a un sistema si no contiene contradicciones, sea internamente, sea en referencia con otro sistema.

En la sección 2 siguiente se analiza la proposición de Peirce acerca de la inferencia abductiva ; la sección 3 está dedicada a comprender el requerimiento de consistencia y, en la sección 4 y final, el de inconsistencia.

2

Peirce formaliza de dos maneras el acceso intuitivo a las nuevas ideas, a las nuevas hipótesis. La primera de ellas, formulada en 1878, es la forma silogística, S, que Peirce llamó “Hipótesis”, y la segunda, establecida en 1903, es la forma explicativa, que se denominará, aquí, E. La formalización silogística es : (i) X es B; (ii) Todo A es B; (iii) de lo cual se abduce que X es A, hipótesis que da cuenta de por qué X es B. (Peirce , 1965, *CP* 2.620). La forma explicativa E en cambio, muestra la conocida estructura: (i) *Se observa el hecho sorprendente C*; (ii) *Pero si A fuese verdadero, C sería materia corriente*; (iii) *luego, hay razones para sospechar que A es verdadero*. (Peirce, *CP* 5.189). Peirce no dice si tales formas las consideraba equivalentes o no. Tal vez por ello, la tradición posterior ha considerado que no hay diferencias entre ambas y las trata indistintamente. A partir de Harman (1965), a la forma E se le agrega la condición de que la hipótesis se infiere si constituye *la mejor* explicación. ²

Refiriéndose al modo E de su formalización, Peirce expone su criterio de buena abducción de la siguiente manera: “Desde luego, debe explicar los hechos. ¿Pero, qué otras condiciones tiene que cumplir para ser buena? La cuestión de la bondad de cualquier cosa es que ella cumpla con sus fines. ¿Cuál es, entonces, el fin de una hipótesis explicativa? Su fin es, a través de la sujeción a un test experimental, conducir a la evitación de toda sorpresa... Cualquier hipótesis, por lo tanto, puede ser admisible, en ausencia de cualquier

² No es el tema de esta ponencia, pero hay razones para pensar que ambas formas no son equivalentes. Autores como A. Aliseda (1998), por ejemplo, afirma que ambas son *complementarias*; Niiniluoto, (1999) a su vez , afirma que la forma E o IME, es *mas general* que la forma S. Hay que considerar que en la forma E, la hipótesis A ya está enunciada en la segunda premisa, y como lo afirma el mismo Niiniluoto y Hanson (1958, 1960), la forma en que la hipótesis A se establece en la segunda premisa puede ser por medio de muchos tipos inferenciales, incluso el deductivo. Una distinción también la hacen Hendricks y Faye (1999).

razón para lo contrario, si es susceptible de verificación experimental... ¿Qué hemos de entender por verificación experimental? La respuesta a esto implica el conjunto de la lógica de la inducción”. (Peirce *CP* 5.197). Se interpreta, aquí, la condición de verificación de Peirce, como una correspondencia con el modo hipotético-deductivo, HD, que es esencialmente inductivo. Y dicha forma HD posee como condición positiva de la contrastación una condición de consistencia: la consistencia que debe darse, si la hipótesis ha de ser provisionalmente aceptada, entre ella y los enunciados contrastadores o de observación en general, esto es, entre hipótesis y evidencia. Esa es una de las formas en que se manifiesta la cuestión consistencia-inconsistencia en las ciencias fácticas. (Priest, 2002; Feyerabend, 1970). Ella puede darse entre a) teoría y teoría (por ejemplo, entre la teoría heliocéntrica y la geocéntrica; entre la teoría de la evolución por selección natural y la teoría termodinámica, específicamente respecto de la edad de la tierra); b) al interior de una teoría (por ejemplo, la teoría del átomo, de Bohr) o c) entre una teoría y las observaciones que ella predice (por ejemplo, el perihelio de Mercurio y la mecánica newtoniana).

Sin embargo, se debe observar que la condición de consistencia especificada para una buena abducción según Peirce, se da *no* al interior de la inferencia abductiva propiamente tal, sino que, podría afirmarse, al *exterior* de ella. Ocurre que la verificación de una hipótesis debe sobrevenir a la generación misma de la hipótesis, esto es, una vez que ésta ya ha sido formada. No es esto, sin embargo, lo que ocurre en los desarrollos actuales acerca de la abducción, en que las condiciones consistencia e inconsistencia pueden caracterizarse como de *internas* a la inferencia abductiva misma.

3

Sobre la base de la forma S de Peirce, los enfoques de inteligencia artificial IA, en general suponen como condición básica de la abducción a la consistencia. La forma standard de la inferencia abductiva en estos enfoques es la siguiente (por ejemplo, Aliseda 1998, Flach y Kakas, 2000, Flach 2002, Bessant 2000, Psillos 2000, Magnani 2001):

- T = una teoría de fondo acerca de ciertos fenómenos, que puede contener, por supuesto, varios enunciados de tipo universal o general;

- O = observación de un hecho, expresada en un enunciado particular, “extraño”, en la expresión de Peirce.

- Entonces, una hipótesis H es abducida de T y O ($T, O \triangleleft H$) si es que explica la observación O, y ello, a su vez, ocurre bajo las tres siguientes condiciones necesarias (hay otras condiciones que concurren pero que no tienen un papel relevante para el presente problema, como la condición de ser “mínima”, esto es no descomponible” en sub hipótesis que cumplan la misma función explicativa):

- $T, H \vdash O$, condición de deducción reversa.³
- (ii) T y H son *consistentes*.
- (iii) $T \not\vdash H$; $T \not\vdash \neg H$, protección de relevancia.

La condición (ii) indica que la hipótesis explicativa, si ha de serlo, no puede estar en contradicción con lo que dice la teoría de fondo. Pero, también Flach (2000, p.92) añade otra condición de consistencia, relacionada con ésta, pero que alude a la relación entre observación y la hipótesis que la explica: si $O \triangleleft H$ (con T), entonces $\not\vdash (O \vdash \neg H)$, esto es, para que O sea explicada por H, O no puede ser inconsistente con H.

La condición (iii) establece una protección a la relevancia de H, esto es, si no se cumpliera (iii), entonces H no sería abducida propiamente tal, y no explicaría O.

Por ejemplo: T = (“la segregación produce pobreza”; “la desigualdad produce pobreza”); O = “S, en el tiempo t, es una sociedad pobre; las siguientes hipótesis cumplen con explicar O, según las condiciones anteriores: H1 = “S es segregada”; H2= “S es desigual”; H3= “S es segregada y desigual”. Una hipótesis H4= “S es segregada y no es segregada”, también cumple con las condiciones señaladas: sin embargo, al no ser *consistente* en sí misma, produce la trivialidad de la condición (i), de acuerdo con el principio de Cornubia. Por tanto H4 se elimina como hipótesis explicativa. La condición de

³ Esta primera condición evidentemente se cumple en la formulación silogística de Peirce, en la que el *caso* y la *regla* implican el *resultado*. (CP 2.623). Por ello, en Peirce, también se cumplirían las otras condiciones, aunque Peirce no lo explicita. La diferencia con la formulación standard actual en IA, es que en ésta la teoría de fondo T es un complejo de enunciados, y no sólo una *regla*.

consistencia actúa aquí como *evaluación* formal de hipótesis plausibles. Puede constatar, también, que una hipótesis $H_5 = \text{“S es segregada y no es pobre”}$ debe eliminarse, pues no cumple con la condición (ii).

Pero, nota Psillos (2000, p.70), que la condición (iii) posee, además de la ya señalada, otra interpretación: H no puede ser implicada por T, pues, entonces, todas las hipótesis $H, H_1 \dots H_n$, serían consistentes entre sí, y ocurre, afirma el autor, que “Es esencial a la abducción ser capaz de tratar con explicaciones mutuamente inconsistentes”.

Así, en la formulación Standard IA de la abducción, hay condición de consistencia en dos niveles distintos: uno más bien formal y otro a nivel formal-evaluativo de una explicación, que permite eliminar, por ejemplo, aquellas hipótesis que sean contradictorias en sí mismas. Esta condición de consistencia es *interna*, diferente por tanto a la condición *externa* de Peirce. Además de la consistencia, como se vio, también hay una condición de *inconsistencia*, dada por la condición (iii), según Psillos.

4

Peirce pone, pues, a la forma E de la abducción, una condición de consistencia, que resulta ser externa. Sin embargo, los desarrollos clásicos posteriores de esa forma E, denominada hoy como Inferencia a la mejor explicación, IME, (por ejemplo, Harman 1965, Fumerton 1980; así como los enfoques en IA de E, por ejemplo Josephson y Josephson 1996, Thagard 1978, 1993) consideran que la abducción maneja hipótesis incompatibles. Algunos autores (Psillos, 2000 p.70) consideran que la inconsistencia es esencial a la abducción, como ya se mencionó. Tomando la expresión de Fumerton (1980, p.594) como modelo de la forma E o IME, se tiene:

(i) Q

(ii) *Del conjunto de hipótesis disponibles e incompatibles $P_1, P_2 \dots P_n$, capaces de explicar Q , P_1 es la que mejor explica Q de acuerdo con criterios $C_1, C_2 \dots C_k$*

(iii) P_1

Una razón que se puede esgrimir para fundar la idea de que las hipótesis plausibles deben ser incompatibles es que si no lo fueran no tendríamos realmente alternativas de explicación; se reducirían a modificaciones de la misma hipótesis. Así, por ejemplo, las diferentes hipótesis abducibles que Kepler sucesivamente manejó aparecen como incompatibles entre sí: a) la órbita de Marte es circular ; b) la órbita posee un epiciclo; c) la órbita es elíptica. Efectivamente a) con c) son contradictorias; pero también lo es a) con b), en el sentido de que es contradictorio afirmar que: “La órbita tiene un epiciclo y no tiene un epiciclo”.

Sin embargo, esta condición de inconsistencia no es asumida uniformemente en los diversos enfoques vigentes. Josephson y Josephson (1996) establecen que las relaciones entre las hipótesis plausibles (que contienen los *abducibles* que completan la teoría de fondo), pueden ser tanto de mutua compatibilidad como de inconsistencia. Por tanto, de ninguna manera la inconsistencia sería esencial a la abducción. Josephson y Josephson trabajan desde las aplicaciones de la abducción al diagnóstico médico, y , allí, es posible encontrar diagnósticos plausibles para un síntoma que no sean necesariamente incompatibles entre sí. Pueden ser, por el contrario, mutuamente soportantes, si están causalmente asociadas. Pero, también, las hipótesis pueden ser mutuamente excluyentes, si se pretende que expliquen distintos subtipos de la misma enfermedad. La inconsistencia es definida en términos semánticos por los autores, con la siguiente formulación: “Que h_j sea incompatible con h_k significa que si h_j es verdadera, entonces h_k no puede serlo, y viceversa” (Josephson y Josephson, 1996, p.71). Los autores clasifican en 6 los tipos de las relaciones entre hipótesis. Las principales van desde ser mutuamente compatibles, a ser una hipótesis un subconjunto de otra; o que el uso de una hipótesis sugiera el uso de la otra; hasta, en el otro extremo, ser mutuamente incompatibles. Así, los autores y su equipo crearon dos sistemas de abducción de diagnósticos, (*máquinas abductivas*) específicamente para abducir explicaciones acerca de qué anticuerpos están con más plausibilidad presentes en un organismo dados determinados antígenos detectados.⁴ El programa RED-1 sólo considera hipótesis consistentes; pero ello resulta limitado, por lo que RED-2 se flexibiliza y considera también explicaciones mutuamente incompatibles. Por ello, la condición de

⁴ Ver el desarrollo de tales máquinas abductivas en Josephson y Josephson, 1996, cap. 3.

inconsistencia se convierte en una virtud. Es más real, en este caso en el campo de diagnóstico médico, considerar que se tendrán todo tipo de hipótesis, sean o no consistentes entre sí.

De acuerdo con Psillos (2000) la abducción equivale a un mecanismo de manejo de inconsistencias, pues, a diferencia de Josephson, para él la inconsistencia es esencial a la abducción. Para ver dicho concepto de manejo de inconsistencias se propone recurrir a la forma en que P.Thagard (1993) establece las condiciones de la abducción del tipo E de Peirce. La cuestión ahí es cómo elegir la hipótesis sobre un conjunto de explicaciones plausibles disponibles. Una de las relaciones para tal evaluación de la mejor explicación (IME), es la concilianza (*consilience*) entre hipótesis (otras condiciones son la simplicidad y la capacidad analógica). Una hipótesis es más conciliante que otra si logra explicar más clases de hechos (no más hechos de la misma clase). En el ejemplo de Thagard se tiene: H1= “el sonido es onda”; H2 = “el sonido es un conjunto de partículas”. Según la concilianza, se tiene que:

- H1 explica (i) el sonido atraviesa sólidos”; (ii) “el sonido se propaga”; (iii) “el sonido se refleja”
- H2 explica : (i) “el sonido se propaga”; (ii) “el sonido se refleja”.

H1 es más conciliante que H2, pues explica fenómenos pertenecientes a una clase de la que H2 no puede dar cuenta. H1 se la puede entender como incompatible con H2 puesto que H2 afirma que el sonido atraviesa cuerpos, en tanto H2 afirma que el sonido no atraviesa cuerpos, o al menos equivale a esa afirmación, por defecto. Por tanto, el criterio de concilianza opera como una solución de hipótesis que explican el sonido pero son entre sí inconsistentes. Pero, sobre esto, hay otro aspecto que recalcar. Ambas hipótesis, H1 y H2 pueden descomponerse cada una en un conjunto de enunciados, 3 en el caso de H1 y 2 para H2. En tal caso, H1 y H2 son inconsistente entre sí, pero sólo en relación con el enunciado “el sonido atraviesa sólidos”; para los otros dos, son mutuamente consistentes. Una de las estrategias esgrimidas en la filosofía de las ciencias para entender la inconsistencia en ciencias empíricas es justamente analizar la teoría inconsistente y considerar sus partes como separadas. O también, de dos teorías inconsistentes respecto de un enunciados,

considerarlas siempre como independientes. (se deben tratar *no adjuntivamente*, como es una de las estrategias de las lógicas paraconsistentes). Por ejemplo, como ya se mencionó, la teoría darwiniana de la evolución y la teoría termodinámica implican enunciados contradictorios respecto de la edad de la tierra. Más, desde un punto de vista no adjuntivo, que sean inconsistentes entre sí no implica que una de ellas deba rechazarse. (Priest, 2002). De lo contrario tendríamos que esperar una consistencia absoluta entre todo enunciado posible respecto de cualquier fenómeno fáctico, lo que parece un constreñimiento demasiado fuerte, e irreal, para la razón.

5

La exposición precedente permite formular al menos las siguientes conclusiones:

- La abducción tiene como condiciones, entre otras, la consistencia y la inconsistencia.
- Mas ellas no se dan en el mismo nivel en la inferencia y, además, cumplen roles diferentes. En la forma abductiva S de Peirce la consistencia cumple, principalmente, el papel de permitir la inferencia misma; en tanto que en la forma E la condición de consistencia opera como un factor de evaluación entre las diferentes hipótesis plausibles. Además de ello, en la forma E la consistencia no parece ser una condición necesaria. En la forma S sí lo es.
- Aunque Peirce no la explicitó, es evidente que la condición de consistencia “interna” ya está implicada en su originaria formulación silogística de la abducción.
- La abducción, en su forma E, en los casos en que las hipótesis plausibles son mutuamente incompatibles, equivale a un mecanismo manejador de inconsistencias.

Referencias

- Aliseda A, 1998, “Abducción como cambio epistémico: C.S.Peirce y las teorías epistémicos en inteligencia artificial”, *Analogía*, 12, UNAM, Mexico, 125-144.
- Aliseda A., 2006, *Abductive Reasoning. Logical Investigations into Discovery and Explanation*, Springer, Dordrecht.
- Bessant B, 2000, “On Relationships between Induction and Abduction: a Logical Point of View”, en Flach y Kakas edits, *Abduction and Induction*, Kluwer, Dordrecht.
- Feyerabend P, 1974, *Contra el método*, Ariel, Barcelona.
- Fumerton, R.A., 1980, “Induction and Reasoning to the Best Explanation”, en *Philosophy of Science*, vol. **47** N°4, 589-600.
- Hanson, N.R., 1958, “The Logic of Discovery”, *The Journal of Philosophy*, vol **LV**, N°25, 1073-1089.
—1960, “More on The Logic of Discovery”, *The Journal of Philosophy*, vol **LVII**, N°6, 182-188.
—1961, “¿Is there a Logic of Discovery?”, en Feigl and Maxwell edits, *Current Issues in the Philosophy of Science*, Holt,Rinehart and Winston, N.York.
- Harman G., 1965, “ Inference to the Best Explanation”, en *The Philosophical Review*, vol **XXIV**, N°1.
- Hendricks V y Faye J, 1999, “Abducting Explanation”, en Magnani, Nersessian, Thagard, edits., *Model-Based Reasoning in Scientific Discovery*, Kluwer, Dordrecht.

- Josephson J y Josephson S, 1996, *Abductive Inference*, Cambridge University Press, Cambridge.
- Flach P, Kakas A, 2000, “Abductive and Inductive Reasoning: Background and Issues”, en Flach y Kakas edits.
- Flach P, 2002, “Modern Logic in the Study of Knowledge”, en Dale Jacquette, *A Companion to philosophical logic*, Blackwell, Oxford / Mass.
- Magnani L, 2001, *Abduction, Reason and Science*, Kluwer, Dordrecht.
- Niiniluoto I, 1999, “Defending Abduction”, *Philosophy of Science*, Supplement to vol **66**, N°3, S436-S451.
- Peirce Charles S, 1965, *Collected Papers of Charles Sanders Peirce*, Ch. Hartshorne y P.Weiss, edits., Cambridge, The Belknap Press of Harvard University Press.
- Psillos S., 2000, “Abduction: between Conceptual Richness and Computational Complexity”, en Flach y Kakas edits.
- Priest G, 2002, “Inconsistency and the Empirical Science”, en J,Meheus edit., *Inconsistency in Science*, Kluwer, Dordrecht., 119-128.
- Thagard P., 1993, *Computational Philosophy of Science*, The MIT Press, Mass.
 —1978, “The Best Explanation: Criteria for Theory Choice”, *The Journal of Philosophy*, **75**, 76-92.