

**CREER, INFERIR Y ACEPTAR:**  
**UNA DEFENSA DE LA INFERENCIA A LA MEJOR EXPLICACIÓN**  
**APTA PARA INCRÉDULOS\***

**Eleonora Cresto**

[Publicado en *Revista Latinoamericana de Filosofía*, Vol. XXVIII, N° 2, 2002)]

**ABSTRACT:** My aim in this paper is to offer an elucidation of the idea of Inference to the Best Explanation (IBE). I suggest that we understand IBE as a complex process involving two steps: an abductive stage, where the agent chooses among the pool of elements previously identified as candidates for explanation, and a selective stage. I give an account of the selective stage in terms of so-called cognitive decision theory. In addition, I distinguish between believing a hypothesis and accepting it for a particular context of research, and argue that typical cases of IBE are *contextual* in this sense — hence an IBE typically yields a new contextual acceptance, but not necessarily a new belief. Finally, I show that the use of IBE, so understood, does not commit us to scientific realism.

## **1. Introducción**

Mi objetivo principal en este trabajo es ofrecer una elucidación de la idea de Inferencia a la Mejor Explicación (de ahora en más, IME). En la sección 2 propongo entender a una IME como un proceso complejo, que involucra el uso de herramientas de teoría de

---

\* Una versión anterior de este trabajo fue presentada en las Conferencias de Invierno del Instituto de Filosofía de la Universidad de Buenos Aires, en julio de 2002. Agradezco a todos los asistentes, y especialmente a Eduardo Barrio por haberme invitado a participar en dichas conferencias. Deseo agradecer también a Alejandro Cassini, quien leyó cuidadosamente el primer manuscrito y me advirtió sobre diversas imprecisiones, así como a Gustavo Ortiz-Millan, por sus sugerencias sobre la naturaleza de las creencias. Durante el período de realización de este trabajo conté con un subsidio de la Fundación Antorchas.

la decisión cognitiva, y que en general culmina con un cambio en el estado epistémico del agente, en un sentido que habrá que explicitar. Argumento además que la idea de IME debe distinguirse del concepto de abducción. En la sección 3 explico en qué consiste la teoría de la decisión cognitiva, y presento brevemente una de las versiones más conocidas de dicha teoría. En la sección 4 sugiero que toda investigación se lleva a cabo en un determinado contexto, y que, en los casos típicos, el objetivo de los agentes consiste en ampliar el conjunto de elementos aceptados para el contexto en cuestión. Esto, sin embargo, no necesariamente lleva a la adquisición de nuevas *creencias*. Con base en estas sugerencias, en la sección 5 esbozo una teoría formal de cambio epistémico basada precisamente en la distinción entre aceptar y creer. Argumento luego que la modificación epistémica que introduce una IME debe entenderse a la luz de esta distinción: en los casos típicos, una IME lleva al investigador a abrazar una nueva aceptación contextual, pero no necesariamente una nueva creencia. Finalmente, en la sección 6 señalo algunas conclusiones interesantes que el presente análisis entraña para la perspectiva epistemológica conocida como “realismo científico”.

## **2. La inferencia a la mejor explicación**

¿Qué es una inferencia a la mejor explicación? Mucho se ha escrito desde que Gilbert Harman acuñó el término por primera vez en (1965).<sup>1</sup> Sin embargo, todavía no existe consenso acerca de cómo debe entenderse exactamente este concepto.

Si nos atenemos a un punto de vista filológico, una IME es, antes que nada, un tipo de inferencia. Pero aquí comienzan los problemas, porque la noción de inferencia

---

<sup>1</sup> Véase también Harman (1968), para algunas aclaraciones importantes sobre su idea original.

es ambigua. Puede referir tanto a un *argumento* como al *proceso* por el cual un agente pasa de creer o aceptar ciertas proposiciones<sup>1</sup> a creer o aceptar alguna otra. De manera correspondiente una IME podría consistir o bien en un tipo especial de argumento, o bien en un proceso, tal vez complejo, por el cual un agente pasa de creer o aceptar que cierto fenómeno ha ocurrido, a creer o aceptar cierta explicación para dicho fenómeno. En el primer caso, se presupone que una IME puede ser capturada por una forma de razonamiento característica, o tal vez por un conjunto de tales formas. En el segundo caso, el énfasis está puesto en el conjunto de operaciones que el agente realiza para llegar a la mejor explicación. Dichas operaciones bien podrían recurrir a estructuras formales bien definidas, pero ello no es una la idea de IME como proceso.

Formalmente, podría pensarse que la manera más sencilla de capturar una IME sería a través de una falacia de afirmación del consecuente: si la hipótesis  $H$  fuera cierta, el fenómeno  $F$  se seguiría;  $F$  ocurre; por lo tanto,  $H$  es cierta. De acuerdo con otro análisis posible, una IME equivaldría a una inducción por eliminación, o tal vez a una falacia de afirmación del consecuente complementada con una inducción eliminativa.<sup>2</sup> En algunos autores encontramos también la sugerencia de que una IME encubre un silogismo disyuntivo con premisas inciertas, esto es, un razonamiento deductivo de la forma:  $(H_1 \vee H_2 \dots \vee H_n) \ \& \ \neg H_1 \ \neg H_2 \dots \ \& \ \neg H_{n-1} \vdash H_n$ , donde  $H_1 \dots H_n$  son explicaciones posibles para cierto *explicandurn*, pero tal que no tenemos total certeza de que la serie de disyunciones sea completa.<sup>3</sup> Por otra parte, desde el punto de vista de una caracterización completamente informal, una IME simplemente describiría cómo, frente a un determinado fenómeno que resulta problemático por alguna razón, el ingenio del

---

<sup>2</sup> Harman (1965), pp.88-89, afirma que una IME equivale aproximadamente, entre otras cosas, a lo que otros autores han llamado “inducción eliminativa”.

<sup>3</sup> Véase Josephson y Josephson (1994), p.12.

investigador combina de manera más o menos intuitiva los distintos aspectos que hacen a una buena explicación (por ejemplo, simplicidad, plausibilidad, etc.), hasta arribar a la mejor explicación de dicho fenómeno que es capaz de construir. Alternativamente, sería la operación según la cual el investigador, guiado otra vez de manera más o menos intuitiva por los aspectos mencionados, selecciona la explicación que cree mejor a partir de un grupo dado de explicaciones posibles.<sup>4</sup>

Tenemos pues al menos dos maneras de entender una IME: la idea de IME como forma (o formas) de razonamiento, y la idea de IME como proceso que puede involucrar tanto aspectos formales como informales. La identificación de IME con un conjunto de reglas formales es más común en el campo de la inteligencia artificial,<sup>5</sup> mientras que las descripciones provenientes del terreno de la filosofía son más propensas a una caracterización en términos de procesos complejos.<sup>6</sup> En este trabajo optaré por esta última perspectiva.

En segundo lugar, nuevamente desde un punto de vista filológico, una IME debe proporcionar alguna manera de seleccionar la *mejor* explicación para cierto evento. (Obsérvese que esto descarta, entre otras cosas, la posibilidad de que una IME pueda ser capturada *con el solo recurso* de una falacia de afirmación del consecuente.) Claro está que nadie supone que pueda haber recetas para ofrecer la mejor explicación *simpliciter*; una “mejor explicación” forzosamente hace referencia a las posibilidades consideradas por cierto investigador en cierto momento. Por ello creo conveniente exigir que un

---

<sup>4</sup> Véase Thagard (1978) para la sugerencia de cómo combinar informalmente diferentes criterios de selección, a partir de algunas consideraciones pragmáticas. Lipton (1991) constituye un buen ejemplo de una perspectiva totalmente informal sobre las IME.

<sup>5</sup> Tal identificación es hartamente común considerando el presupuesto adicional de que IME es un nombre alternativo para hablar de abducción. Véase más abajo, en esta misma sección, la discusión sobre los diversos significados de “abducción”.

<sup>6</sup> Para una pequeña muestra de cómo pueden combinarse factores formales e informales véase el artículo de Richard Fumerton, “Inference to the Best Explanation”, en Dancy y Sosa (1993), pp. 207-209.

análisis de la idea de IME sea capaz de ofrecer algún mecanismo explícito para elegir una hipótesis explicativa, o tal vez más de una, entre un conjunto pre-seleccionado de explicaciones potenciales. La presentación explícita del problema como un problema de selección de hipótesis equivale tan sólo a traer a la luz un elemento que está ya más o menos dado por supuesto en la mayoría de los estudios sobre IME. En este trabajo voy a proponer una elucidación de las IME que satisfaga este requisito básico.

Más precisamente, sugiero entender una IME como un proceso que consta de dos etapas bien definidas. En un primer momento el investigador propone un conjunto de hipótesis explicativas. Cada una de ellas brinda información específica que permite mejorar la comprensión que el agente tiene de cierto fenómeno; expresado de otra manera, cada una de ellas constituye una solución característica al problema (o perplejidad) que inició la investigación. En un segundo momento, el investigador selecciona una de tales hipótesis, o tal vez suspende el juicio entre dos o más de ellas. Al primer momento, siguiendo algunos escritos de Peirce,<sup>7</sup> lo llamaré *abducción*; no me ocuparé demasiado de la etapa abductiva en este trabajo.<sup>8</sup> A la segunda etapa la llamaré *etapa selectiva*, o etapa de selección de hipótesis, y haré una elucidación de la misma en términos de la teoría de la decisión cognitiva. Entiendo que dicha teoría proporciona la manera más eficiente de elegir elementos entre un conjunto predeterminado de hipótesis.

Hay otro ingrediente que es importante mencionar a la hora de ofrecer un análisis de la idea de IME. Desde Harman en adelante, todos los autores que han escrito sobre este tipo de inferencia parecen estar de acuerdo en que el objetivo último de una

---

<sup>7</sup> Peirce (1908). Esto es, me refiero a la segunda época de Peirce; véase más abajo en esta misma sección.

<sup>8</sup> Por ende, no me ocuparé de qué propiedades constituyen una *buena* explicación. Para un análisis de IME que se concentra precisamente en este problema, véase Lipton (1991).

IME es provocar algún tipo de cambio epistémico en el agente que la realiza. De allí el atractivo que esta idea despierta en los ámbitos de teoría del conocimiento o filosofía de las ciencias. Así pues, respetando el sentido habitual de IME, entenderé aquí que el resultado de la etapa de selección de hipótesis será, al menos en los casos típicos, la modificación del estado epistémico del agente, en un sentido que habrá que elucidar. En este punto voy a proponer un esquema de cambio epistémico, deudor de las llamadas “teorías de revisión de creencias”,<sup>9</sup> que hace uso de una distinción fundamental entre los conceptos de *aceptación* y *creencia*. La distinción entre creer y aceptar goza actualmente de cierta popularidad en diversos ámbitos de la filosofía,<sup>10</sup> pero ninguna de las interpretaciones conocidas recoge las ideas que la presente propuesta intenta capturar. Retomaré este tema en las secciones 4 y 5.

Una pregunta posible a esta altura es por qué restringir el uso de “abducción” sólo a la primera etapa de una IME. En efecto, mi uso del término se separa de aquellos autores que tienden a *identificar* IME con abducción.<sup>11</sup> Creo que llamar abducción a todo el proceso inferencial es un error, o, para ser más exactos (puesto que cada uno es libre de usar las palabras como más le guste), es simplemente el fruto de una confusión histórica. Lamentablemente, es una confusión bastante extendida, de modo que tal vez sea apropiado dedicar algunos párrafos a este tema.

Virtualmente todos los autores que hacen uso del término abducción dicen rescatar el sentido usado por Peirce. Pero el asunto es complicado, porque Peirce dio por lo menos dos caracterizaciones diferentes, correspondientes a dos épocas distintas de su

---

<sup>9</sup> Para un panorama sistemático de tales teorías véase Gärdenfors (1988) o Hansson (1999).

<sup>10</sup> Cf. por ejemplo Cohen (1992), van Praassen (1989), o Maher (1993), entre muchos otros.

<sup>11</sup> A título de ejemplo, véanse van Fraassen (1989), cap. 7; Josephson y Josephson (1994); o S. Psillos, “Abduction: between conceptual richness and computational complexity”, en Flachs y Kakas (2000), pp. 59-74.

pensamiento. De acuerdo con el primer sentido, cronológicamente más antiguo, abducción o hipótesis es un tipo de razonamiento ampliativo que concluye con una explicación de algunos de los eventos descritos en las premisas.<sup>12</sup> Peirce entiende que es una manera posible de invertir el orden de premisas y conclusión en un razonamiento deductivo. Curiosamente, dentro de los deductivos también ubica lo que hoy llamaríamos silogismos estadísticos. Por ejemplo, estaríamos haciendo una abducción si tenemos premisas que dicen que “el 80% de los usuarios intensivos de computadoras presentan el síndrome del ojo seco, dolores metacarpianos, y contracturas cervicales”, y que “este individuo tiene los tres síntomas descritos”, y a partir de ellas concluimos que “este individuo es un usuario intensivo de computadora”. En otra terminología, una abducción en este sentido equivale o bien a una falacia de afirmación del consecuente, llamada también “*modus ponens* inverso” [*backward modus ponens*], o bien a la inversión de un silogismo estadístico que imite la forma en que dicha falacia invierte al *modus ponens*.

La idea de abducción como forma de razonamiento ampliativo sobrevive actualmente sobre todo en el campo de la inteligencia artificial, donde podemos encontrar numerosas ofertas de reglas abductivas que, al menos en principio, podrían ser implementadas por una máquina.<sup>13</sup> En muchos de estos casos leemos además que lo que se ofrece es en verdad lo mismo que, en otros ámbitos, suele llamarse “IME”. Pero, según hemos visto, una IME es algo bastante más complejo que una mera falacia de afirmación del consecuente. Requiere que al investigador se le ocurran hipótesis explicativas apropiadas, y requiere también algún tipo de mecanismo de comparación

---

<sup>12</sup> Peirce (1878).

<sup>13</sup> Para un panorama general del tratamiento de la abducción en inteligencia artificial véase Flachs y Kakas (2000); en Pagnucco (1996) puede encontrarse un estudio detallado de la idea de abducción desde la perspectiva de las teorías de revisión de creencias.

entre las posibles hipótesis ya propuestas. Por supuesto, si por “reglas abductivas” entendemos un conjunto lo suficientemente variado y sofisticado de reglas, no habría en principio inconvenientes en llamar abducción a una IME; simplemente obtendremos una perspectiva puramente formal de IME al estilo de las que mencionamos al comienzo de esta sección. Pero es claro que esto ya no es ‘abducción en el (primer) sentido de Peirce’.

De todos modos, creo que la raíz del problema (esto es, el *origen* de la identificación entre abducción e IME) debe buscarse en el segundo sentido que da Peirce a esta palabra. En una segunda acepción, abducción es el proceso de formación de hipótesis a partir de una observación que requiere explicación. Pertenece, por tanto, a lo que hoy llamaríamos “contexto de descubrimiento”. Es el sentido de abducción rescatado, entre otros, por Hanson (1958). Peirce afirma que dicho proceso, visto globalmente, responde a la siguiente forma lógica:

“El hecho sorprendente *C* es observado;  
Pero si *A* fuera verdadero, *C* resultaría obvio,  
Por ende, hay razones para sospechar que *A* es verdadero.”<sup>14</sup>

Es decir que nos topamos nuevamente con una falacia de afirmación del consecuente, esto es, con la forma que anteriormente había clasificado como abducción o hipótesis (ello explica, por supuesto, que utilice el mismo término<sup>15</sup>). Pero aquí hay una diferencia fundamental respecto de la primera época. Una abducción ahora *no es un*

---

<sup>14</sup> Peirce (1958), 5.189.

<sup>15</sup> En sentido estricto, en sus escritos tempranos Peirce usa únicamente el término “hipótesis”. Sin embargo, cuando más tarde critique su clasificación temprana de razonamientos, se referirá a su anterior concepto simplemente como al de una idea confusa de abducción. Véase Peirce (1958), 2.102.

*procedimiento ampliativo*: las hipótesis resultantes de un proceso abductivo deben contrastarse primero, antes de estar en condiciones de incorporarse a un cuerpo de creencias. En esta segunda época Peirce entiende por *inducción*, justamente, el proceso de contrastación de hipótesis con la experiencia a través de predicciones;<sup>16</sup> esta etapa, posterior a la abductiva, pertenece claramente a lo que hoy llamamos “contexto de justificación”.

Para Peirce, además, la formulación de hipótesis no es una actividad que esté sujeta a reglas, sino que depende de la intuición del investigador. Ahora bien, no cualquier hipótesis posible es admisible como resultado de una abducción: dicha hipótesis debe ser explicativa, económica y capaz de evaluación experimental. En verdad, ello significa que ya hay algún tipo de elección o descarte implícito entre las hipótesis lógicamente posibles que puede concebir un agente. Sin embargo, esta selección, como ya se ha dicho, no culmina con la aceptación definitiva de una hipótesis, sino con la propuesta de que la hipótesis en cuestión es digna de ser puesta a prueba. Estas afirmaciones parecen haber confundido a muchos autores. En efecto, algunos análisis de abducción e IME parecen confundir los mecanismos por los cuales Peirce indica que una hipótesis puede ser considerada una buena *candidata* para explicar un fenómeno, con aquellos mecanismos que permiten decidir si cierta candidata es la mejor de un conjunto previamente constituido.<sup>17</sup> Para evitar estos problemas terminológicos es que Laudan, por ejemplo, propone hablar del *contexto de prosecución* [*context of pursuit*], para referir al terreno intermedio entre descubrimiento y justificación.<sup>18</sup>

---

<sup>16</sup> Véase Peirce (1908), pp. 367-373.

<sup>17</sup> De manera paradigmática, véase Josephson y Josephson (1994), p. 5.

<sup>18</sup> Laudan (1981).

No está claro si la mencionada confusión puede atribuirse al propio Harman, aunque sí parece haber tocado a muchos de sus lectores. Harman (1965) expresa que “la IME corresponde *aproximadamente a lo que otros han llamado ‘abducción’, ‘el método de hipótesis’, ‘inferencia hipotética’, ‘el método de eliminación’, ‘inducción eliminativa’ e ‘inferencia teórica’*” (pp. 88-89, énfasis mío). Pero, justamente porque todos estos términos están cargados de connotaciones indeseables es que, según nos dice, prefiere su propia terminología. De modo que no sería correcto culpar a Harman por haber identificado abducción con IME, porque, en sentido estricto, él no hace semejante cosa: la identificación, en todo caso, correría por cuenta de aquellos “otros” a quienes Harman refiere.

Independientemente de cuál haya sido la idea original de Harman, es claro que “IME” no puede ser un sinónimo de “abducción” en el segundo sentido de Peirce, a menos que se rechace la idea de IME como mecanismo de cambio epistémico, lo cual ningún autor parece estar dispuesto a conceder. Y, como ya hemos visto, tampoco puede equipararse a “abducción” en su primera acepción. Desde luego, cada uno es libre de estipular el uso que prefiera; mi intención en estos párrafos fue tan sólo la de observar que no sería correcto, sin embargo, atribuir dicho uso a Peirce.

Para recapitular, voy a entender a una IME en términos de un proceso que consta de dos etapas, y que, por lo general, culmina con la modificación del estado epistémico de un agente. Durante la primera etapa (fase abductiva) tiene lugar la generación de respuestas potenciales, y durante la segunda (etapa selectiva) tiene lugar la selección de hipótesis previamente propuestas en la primera etapa. En los casos típicos, el elemento seleccionado es *aceptado* por el agente para cierto contexto de investigación, lo cual modifica su estado epistémico en un sentido particular que habrá que clarificar. Para

ello argumentaré que se hace necesario distinguir entre las nociones de aceptación y creencia.

Como ya he indicado, en este trabajo no me ocuparé ni del contexto de descubrimiento propiamente dicho, ni del de “prosecución” (en terminología de Laudan). Es decir que no me ocuparé de examinar qué hace que una hipótesis sea en verdad explicativa, ni tampoco de detallar qué otros rasgos, aparte de su capacidad explicativa, hacen que una hipótesis merezca ser considerada una posibilidad seria. Centraré mis esfuerzos, en cambio, en estudiar cómo procede la elección entre hipótesis rivales una vez que éstas han sido seriamente propuestas como resultado de una abducción. Más precisamente, analizaré el caso particular en que dichas hipótesis son todas empíricamente adecuadas, esto es, compatibles con la evidencia observacional disponible en un momento dado. ¿Cómo se realiza la selección en este caso? Entiendo que la teoría de la decisión cognitiva brinda una respuesta adecuada para esta pregunta. En la siguiente sección explicaré someramente en qué consiste dicha teoría, y presentaré una de las versiones más conocidas, debida a Isaac Levi.

### **3. La teoría de la decisión cognitiva**

Supongamos que un agente considera un conjunto determinado de diferentes acciones posibles, y que debe decidir cuál de dichas acciones llevar a cabo. La teoría de la decisión indica cómo efectuar la elección de manera racional, dada una serie de supuestos básicos. Por ejemplo, se supone que el agente puede ordenar las consecuencias de sus acciones por grado de preferencia. Es decir que un agente preferirá que ocurran ciertas consecuencias más que otras. Esta escala de preferencias no es simplemente ordinal, sino que se supone, además, que el agente puede asignar utilidades

específicas a cada consecuencia.<sup>19</sup> También se presupone que cada agente es capaz de hacer asignaciones coherentes de probabilidad subjetiva sobre los estados posibles del mundo, lo cual determina a su vez que cada una de las consecuencias posibles de una acción tendrá asociada cierta probabilidad. Con todos estos elementos, la utilidad esperada (*UE*) de cada acción considerada por el agente es la sumatoria de las utilidades asociadas con cada consecuencia, ponderadas por la probabilidad de cada consecuencia. Si  $c_i$  es la consecuencia de la acción *A* dado que ocurre el estado *i*,  $u(c_i)$  la utilidad de  $c_i$ , y  $P(c_i)$  la probabilidad de  $c_i$ , y además, según el agente, hay *n* estados posibles, tenemos que  $UE(A) = \sum^n [u(c_i) P(c_i)]$ . La teoría, en su uso normativo, indica que los agentes, para actuar racionalmente, deben actuar de modo tal que maximicen su utilidad esperada.<sup>20</sup>

Pues bien, la teoría de la decisión cognitiva consiste en la aplicación de estos conceptos básicos al campo epistémico. Así, en lugar de calcular la utilidad esperada que resulta de llevar a cabo diversas acciones posibles, la teoría calcula la utilidad *epistémica* esperada de distintas hipótesis. La utilidad epistémica esperada de una hipótesis equivale en verdad a la utilidad esperada que tendría para el agente el acto de creer o aceptar la hipótesis en cuestión.

Posiblemente la versión más detallada de dicha teoría se deba a Isaac Levi.<sup>21</sup> La versión de Levi será el punto de partida de mi análisis del contexto selectivo de una

---

<sup>19</sup> La teoría requiere que el agente asigne utilidades, debido a que ello permite medir la distancia relativa entre los elementos de la jerarquía. No se necesita ninguna otra información más allá del orden y la distancia relativa entre dichos elementos. Las escalas de utilidad son pues lo que se conoce como “escalas de intervalo”: en sentido estricto, la escala de utilidades de un agente determina un conjunto infinito de funciones de utilidad, todas ellas equivalentes bajo transformación positiva lineal.

<sup>20</sup> Es importante señalar que hay usos no normativos de la teoría de la decisión. Por ejemplo, los autores que se comprometen con alguna versión de la llamada “teoría de la preferencia revelada” proponen interpretar la acción de un individuo como aquella que maximiza su utilidad esperada, con lo cual las escalas de preferencia y probabilidad son reconstruidas a *posteriori*. De hecho una de las formulaciones estándar de la teoría, debida a Savage (1954), incurre en dicho compromiso.

<sup>21</sup> Por ejemplo en Levi (1984), capítulos 5, 6 y 7, o Levi (2001).

IME, de modo que la expondré con cierto detalle, aunque muchos aspectos deberán ser inevitablemente simplificados por razones de espacio.

Levi recurre a la idea de decisión cognitiva a la hora de caracterizar lo que él llama una *expansión inductiva*. En una expansión inductiva típica, el estado epistémico de un agente se ve ampliado por el agregado de un nuevo elemento  $H$ , tal que los elementos previamente presentes hacen razonable (*i.e.*, apoyan inductivamente) la incorporación de  $H$ .  $H$ , sin embargo, no se *deduce* de los elementos previos, y por lo tanto, según la perspectiva del agente, aceptar  $H$  entraña cierto riesgo de error.

La oportunidad de realizar una expansión inductiva aparece cuando el agente desea obtener una respuesta para cierta pregunta. Sea  $K$  una representación del estado epistémico inicial del agente. La representación más usual de dicho estado epistémico consiste en un conjunto deductivamente cerrado de enunciados de un lenguaje formal debidamente regimentado. Es habitual referirse a  $K$  como el “conjunto de creencias” del agente. Con base en  $K$ , se define luego una *partición última*  $U_k$ , o conjunto de hipótesis  $h_1...h_n$ , tal que: (i) cada una de las hipótesis provee una respuesta a la pregunta que inicia la investigación; (ii) todas las hipótesis son excluyentes entre sí; (iii) todas son compatibles con  $K$ ; (iv) ninguna de ellas se deduce de  $K$ ; y (v) una y sólo una de ellas es verdadera dado  $K$ . Para la aplicación que nos interesa en este artículo, podemos interpretar a  $h_1...h_n$ , como explicaciones potenciales rivales, empíricamente adecuadas, y capaces de satisfacer cierta demanda de información por parte del agente.

El siguiente paso consiste en evaluar cuál es el elemento preferido de dicha partición última; en caso de duda, más de un elemento podría ser seleccionado; en algunas situaciones, ninguno de ellos lo será. Cada uno de estos resultados determinará un tipo de modificación característica de  $K$ , sobre la que volveremos más adelante. Para

llevar a cabo dicha evaluación, Levi calcula la utilidad epistémica esperada de cada una de las hipótesis. Veamos cómo.

En primer lugar, debemos definir una función de probabilidad  $P_K$ , relativa a  $K$ , sobre los elementos de la partición última. (En lo que sigue eliminaré el subíndice  $K$  de  $P_K$  toda vez que no haya riesgo de confusión).

En segundo lugar, debemos evaluar la *utilidad* de cada uno de dichos elementos. Establecer funciones de utilidad adecuadas para los elementos de  $U_K$  no es una tarea sencilla. Reinterpretando una vieja idea de William James, Levi sugiere que en toda investigación coexisten dos objetivos o metas en pugna entre sí; dos *desiderata*.<sup>22</sup> Una función de utilidad apropiada deberá tenerlos en cuenta a ambos. Uno de ellos consiste en evitar el error, o evitar creer cosas erróneas. Diremos que  $T(h,x)$  es una función que arroja diferente resultado dependiendo de si el error fue evitado ( $x = v$ ) o no ( $x = f$ ). Sea  $T(h,v) = a$  y  $T(h,f) = b$ , donde  $a > b$ . Para simplificar, estipularemos que  $a = 1$  y  $b = 0$ .

El otro objetivo es adquirir nueva información valiosa. Sea  $C(h,x)$  una función que especifica el incremento en valor informacional obtenido. Se presupondrá que este valor es el mismo ya sea que  $x = v$  o que  $x = f$ ; esto es,  $C(h,v) = C(h,f) = C(h)$ . No entraré aquí en los pormenores de cómo calcular  $C$ . Solamente diré, de manera un tanto imprecisa, que  $C$  mide el *contenido* de la hipótesis.

En síntesis, de acuerdo con la teoría de los dos *desiderata* de la investigación, por un lado deseamos salir de situaciones de incertidumbre y adquirir nueva información, pero por otro lado, también deseamos evitar el error. La utilidad cognitiva  $V(h,x)$  de una hipótesis, entonces, debería expresar cuantitativamente el grado en el cual la hipótesis satisface ambos objetivos. Pero salta a la vista que éstos no pueden

---

<sup>22</sup> James (1897), p. 17 y ss.

satisfacerse máximamente de manera simultánea: maximizar el deseo de evitar el error equivale a minimizar el deseo de adoptar nuevas creencias dudosas, y viceversa. De modo que una función de utilidad lo más que puede hacer es expresar el grado en que una hipótesis es capaz de satisfacer *cierto compromiso* entre ambos objetivos. Sea pues  $V(h,x) = \alpha T(h,x) + (1-\alpha)C(h,x)$ . El coeficiente  $\alpha$ , o índice de cautela,<sup>23</sup> puede verse como una manera de representar la importancia relativa atribuida a la preocupación por evitar el error, y a la preocupación por maximizar el valor informacional de nuestro cuerpo de creencias.

Con esto ya tenemos todas las herramientas necesarias para calcular la utilidad cognitiva esperada (*UE*) de cada elemento de la partición última. Es claro que, para cada hipótesis, hay dos únicos estados posibles del mundo que son relevantes: o bien la hipótesis es verdadera, o bien es falsa. Entonces, la utilidad cognitiva esperada de una hipótesis será la probabilidad de que la hipótesis sea verdadera, multiplicada por la utilidad asociada a que sea verdadera, más la probabilidad de que sea falsa, multiplicada por la utilidad asociada a que sea falsa. Es decir:

$$\begin{aligned} UE(h) &= P(h) V(h,v) + P(-h) V(h,f) = \\ &= P(h) [\alpha T(h,v) + (1-\alpha)(C(h))] + (1-P(h)) [\alpha T(h,f) + (1-\alpha)(C(h))] = \\ &= \alpha P(h) + (1 - \alpha)(C(h)).^{24} \end{aligned}$$

De acuerdo con esto, Levi propone una regla específica para la modificación de *K*. No daremos aquí los detalles de dicha regla.<sup>25</sup> Sólo diré que Levi propone rechazar todos

---

<sup>23</sup> En sentido estricto, lo que Levi llama “índice de cautela” (o de audacia, según se mire) es  $(1-\alpha)/\alpha = q$ .  $q$  tiene ciertas propiedades muy interesantes, sobre las que lamentablemente no me puedo explayar aquí. Véase Levi (2001), pp. 32 y ss.

<sup>24</sup> Como ya he advertido, esta presentación es tan sólo una versión simplificada de la propuesta de Levi. Entre otras cosas, éste presupone que los agentes raramente cuentan con una única función *P* de probabilidad subjetiva. En los casos típicos, debemos suponer que disponen de un *conjunto convexo* de tales funciones de probabilidad. Sobre este tema véase “On Indeterminate Probabilities”, reimpreso en Levi (1997), cap. 6.

<sup>25</sup> Véase Levi (2001), p. 32.

los elementos que no cumplan con cierto nivel mínimo de utilidad cognitiva esperada. Luego, se incorporará a  $K$  la disyunción de todos los elementos  $h_i \in U_K$  que no han sido rechazados, para finalmente realizar la clausura deductiva correspondiente. Así, si la disyunción de elementos no rechazados es representada por  $H$ , tendremos como resultado un nuevo estado epistémico  $K'$ , conformado por  $K$ ,  $H$ , y las consecuencias lógicas de ambos. En símbolos,  $K' = Cn(K \cup \{H\})$ , donde  $Cn$  es el operador tarskiano de consecuencia lógica. La idea es que, si hay más de un elemento no rechazado (esto es, si  $H$  no es simplemente idéntica a algún  $h$  de la partición última) el agente suspende el juicio entre ellos. En caso de duda severa,  $H$  podría consistir en la disyunción de todos los elementos de  $U_K$ , en cuyo caso en realidad no estaríamos haciendo modificación alguna del estado epistémico del agente, puesto que  $H$  ya pertenecería a  $K$  (dado que  $K$  es un conjunto deductivamente cerrado, y dado que inicialmente al menos un elemento de la partición última es juzgado verdadero). Por otra parte, rechazar *todos* los elementos de  $U_K$  supondría introducir una contradicción en el conjunto de creencias. De modo que, de obtenerse este resultado, el agente se vería forzado a retirar algunos elementos de su conjunto inicial  $K$ . Pero aquí ya no estaríamos tratando con expansiones inductivas, sino con *contracciones*. En este trabajo no me ocuparé de examinar las posibles estrategias de contracción de un conjunto de creencias.

Obsérvese que el mecanismo de selección descrito en esta sección puede muy bien aplicarse a un caso de IME, en el cual la partición última consiste en un conjunto dado de explicaciones rivales empíricamente adecuadas. Para elegir la mejor explicación, o disyunción de mejores explicaciones, se calcula la utilidad cognitiva esperada de cada hipótesis de la manera ya señalada, para lo cual deben tenerse en cuenta las funciones  $P$  y  $C$  del investigador, así como cierto índice de cautela. Luego de

realizada la selección, la teoría presentada aquí requiere que la mejor explicación, o disyunción de mejores explicaciones, se incorporen al conjunto  $K$  de creencias del agente, modificándolo de la manera detallada en el párrafo anterior.

Entiendo que el mecanismo de selección descrito cumple con todo lo que podemos pedir de un mecanismo selectivo en una IME. Sin embargo, creo también que el modelo general de cambio epistémico en el que dicho mecanismo de selección está inserto no permite capturar toda la gama de situaciones que pueden ocurrir en el marco de una IME, En la próxima sección argumentaré que las ideas básicas presentadas aquí deberían adoptarse en combinación con un modelo de cambio epistémico un tanto más complejo.

#### **4. Compromisos epistémicos contextuales**

De acuerdo con una distinción usual en la bibliografía reciente, hay básicamente dos maneras opuestas de encarar la problemática de la investigación racional y la creencia. Por un lado, podemos optar por un modelo cartesiano de investigación, llamado a veces el “modelo Duda- Creencia”; por otro lado, podemos adoptar un modelo peirceano, o modelo “Creencia-Duda”.<sup>26</sup> El primero se identifica con la teoría del conocimiento tradicional, y viene en todo tipo de variantes (fundacionismo, coherentismo, externalismo, internalismo, etc.). Adviértase que en este contexto el término “cartesiano” no refiere a la variedad o *especie* de modelo dada por Descartes mismo, sino a su *género*. Un rasgo común de todas las variantes de este modelo es la exigencia de identificar algún criterio para separar las creencias que están justificadas de aquellas

---

<sup>26</sup> Es habitual identificar a Peirce (1877) como el escrito fundacional del modelo Creencia-Duda. Para una descripción sencilla de ambos modelos, véase la introducción de Fuhrmann (1997).

que no lo están. Por el contrario, el segundo modelo surge de raíces pragmatistas, y mantiene que sólo los *cambios* de creencias deben justificarse. Es este último modelo el que está habitualmente presupuesto en la mayoría de las discusiones contemporáneas sobre revisión de creencias o teoría de la decisión cognitiva.

Podríamos tratar de capturar el corazón de una epistemología pragmatista de tipo peirceano por medio de algún principio apropiado. ¿Qué podría contar como una formulación satisfactoria de tal principio? He aquí una posible sugerencia:

*EP: La racionalidad requiere que los agentes no realicen cambios de creencias a menos que haya justificaciones adecuadas para tales cambios. Conversamente, prescribe que los agentes lleven a cabo dichos cambios toda vez que haya buenas razones para ello. Además, también exige que los agentes no se ocupen de justificar sus creencias presentes.*

Este es sólo un principio esquemático, pues deja abierto el problema acerca de qué cosa sea una justificación adecuada. Esto es exactamente lo que deseamos; el mero compromiso con una línea epistemológica pragmatista no tiene por qué forzarnos a adoptar un criterio particular de justificación.

Trataré de aclarar ahora el sentido de “creencia” que parece darse por sentado en muchas de las discusiones contemporáneas sobre cambio epistémico. Una posible interpretación es la de creencia como “compromiso epistémico” [*doxastic commitment*].<sup>27</sup> En la manera más intuitiva de entender esta idea, decir que un agente tiene una determinada creencia significa que el agente se ha hecho a sí mismo la promesa de actuar de cierta manera, donde “actuar” ha de ser tomado en un sentido

---

<sup>27</sup> Cf. Levi (1991), cap. 2.

amplio. Así, si  $X$  cree que  $p$ , y le preguntamos si ocurre que  $p$ , y además  $X$  quiere revelarnos su compromiso,  $X$  deberá contestar la pregunta afirmativamente; si  $X$  se encuentra en la disyuntiva de decidir si llevar a cabo  $A$  o  $B$ , deberá tomar su decisión bajo el supuesto de que  $p$ ; de manera análoga, si  $X$  se propone deliberar sobre alguna cuestión (independientemente de si luego hará algún uso práctico del resultado de la deliberación),  $X$  deberá proceder bajo el supuesto de que  $p$  es verdadero. Como puede verse, “compromiso” es un concepto normativo, que indica lo que el agente *debe* hacer, y no lo que de hecho hace en cada caso.

Así pues, decir que un agente cree que  $p$  significa que, toda vez que el agente actúe o delibere, deberá dar por supuesto que  $p$ . Por supuesto, nada impide que  $X$  cambie de idea sobre  $p$ , pero mientras ello no suceda, el agente se ha resuelto a tomar a  $p$  como verdadero para toda acción o deliberación futuras, cualesquiera que éstas sean. Por supuesto, no todo lo que  $X$  haga en el futuro requerirá tomar a  $p$  en consideración. Pero esto es irrelevante; lo importante es que cada vez que la consideración de la verdad o falsedad de  $p$  sea capaz de producir algún tipo de diferencia, el agente deberá actuar como si  $p$  fuera cierto.

En síntesis, la creencia es interpretada como una actitud que el agente deberá tener en *todo contexto posible* (para el agente) *de acción o deliberación*. Sin embargo, adviértase también que la idea de compromiso, por sí mismo, no acarrea esta connotación. Un agente podría muy bien estar comprometido a actuar de cierta manera sólo bajo ciertas circunstancias. Por lo tanto, “creencia” no es simplemente sinónimo de “compromiso epistémico”, sino de “compromiso epistémico para todo contexto”. De modo que, a la hora de dar una definición rigurosa, deberíamos explicitar este último punto. Para ser precisos, entonces, digamos que un agente  $X$  cree que  $p$  si y sólo si  $X$  está comprometido con la verdad de  $p$  (junto con todas sus consecuencias lógicas) para

todo contexto de acción o deliberación. Entiendo que esta definición captura la idea de creencia implícita en buena parte de la bibliografía sobre revisión de creencias y teoría de la decisión cognitiva.

Con este marco teórico, volvamos ahora a considerar la elección de hipótesis, o “mejor explicación”, a través de un mecanismo de decisión cognitiva. Es claro que la versión ofrecida en la sección anterior debe ser entendida dentro del marco teórico de los modelos Creencia-Duda: su mecanismo central consiste en utilizar creencias ya presentes, que no se cuestionan, para evaluar la legitimidad de incorporar otras nuevas. Efectivamente, dicha propuesta puede verse como un medio para proveer un *criterio* que el agente puede usar para saber si cierta expansión es obligatoria o prohibida (dado cierto problema, cierto conjunto de respuestas potenciales, un determinado conocimiento antecedente, una función de probabilidad subjetiva, un índice de riesgo, etc.). En otras palabras, dicha propuesta completa el principio esquemático (EP) con un criterio sustantivo, aunque parcial,<sup>28</sup> sobre la justificación de los cambios.

Hay, sin embargo, un aspecto crucial de (EP) completado con el mecanismo de selección de hipótesis de la sección anterior que encuentro demasiado restrictivo.

Repasemos una vez más la interpretación de “creencias” como “compromisos para todo contexto de acción y deliberación”. Entiendo que, a la hora de tomar parte en un proceso de decisión cognitiva, los agentes no siempre tienen como meta la adopción de nuevas creencias en este sentido. En numerosas ocasiones, todo lo que un agente quiere es comprometerse epistémicamente *para cierto contexto determinado*. Consideremos el enunciado que dice que la enfermedad *X* resulta causada por un único factor *Y*. Un agente puede encontrar razonable aceptar dicho enunciado como verdadero

---

<sup>28</sup> El criterio es parcial puesto que no trata específicamente el caso de las *contracciones* de un conjunto de creencias.

en el contexto de una investigación sobre la génesis de otras enfermedades, y, sin embargo, a la hora de diseñar una política de salud pública, tal vez piense que la cautela que debe ejercerse es mayor, y que es más adecuado suspender el juicio, *i.e.*, actuar como si su verdad o falsedad no estuviera aún establecida. En este ejemplo, evidentemente el agente *no* ha decidido *creer* que *X* es causada por *Y*, en el sentido de creencia ofrecido más arriba. Sin embargo, ha decidido dar por supuesta dicha relación causal en un ámbito importante de su vida. Considero que, lejos de ser una excepción, esta es una situación con la que nos topamos habitualmente. En otras palabras, en numerosas ocasiones los agentes llevan a cabo decisiones cognitivas *contextuales*

De aquí en más voy a reservar el nombre de “aceptaciones” para referirme a los compromisos que un agente adquiere para *algún* contexto (no necesariamente para todos); las “creencias”, como ya dijimos, referirán a los compromisos que el agente mantiene para todo contexto. Obsérvese que, bajo esta definición, una creencia es un tipo especial de aceptación.

Esta definición entraña una importante consecuencia. Puesto que es posible para un agente aceptar un elemento para cierto contexto pero no para todos, ello implica que hay contextos para los cuales el agente tiene mayor cantidad de compromisos epistémicos que creencias. Así, el conjunto de elementos que un agente toma como “base de acción y deliberación” puede cambiar según el contexto. En nuestro ejemplo anterior sobre la causa de la enfermedad *X*, podemos suponer que, para llevar a cabo la decisión racional que finalmente culminó con la aceptación de un nuevo compromiso, nuestro agente no solamente usó un índice de cautela propio para dicho contexto, sino que dio por sentados elementos con los que no se comprometería en otros ámbitos, esto es, elementos que no son en sentido estricto “creencias”.

Podría argumentarse que hablar de realizar problemas de decisión cognitiva contextuales simplemente equivale a poner un nuevo nombre a la vieja idea de que los agentes a veces llevan a cabo razonamientos hipotéticos. Decir que hemos adoptado el nuevo compromiso de que  $p$  para un contexto en el cual damos a  $q$  por sentado, cuando  $q$  no es un compromiso para todo contexto, no sería sino otra manera de decir que, razonando bajo el supuesto (falso, en lo que a nosotros concierne) de que  $q$  es el caso, hemos llegado a ser capaces de afirmar  $p$ .

En verdad, no tengo mayores inconvenientes en redescibir la existencia de múltiples contextos de decisión como contextos de razonamiento hipotético. Sin embargo, hablar de decisiones tomadas bajo supuestos hipotéticos es un poco engañosos por cuanto parece dar a entender que dichas decisiones constituyen un mecanismo de cambio epistémico provisorio. Parece implicar que alguna vez el agente buscará, digamos, “cancelar” los supuestos; parece implicar también que las únicas decisiones cognitivas realmente importantes o que realmente cuentan, son las que atañen a modificaciones de un conjunto de creencias propiamente dicho. Pero ninguna de estas connotaciones es apropiada. Las decisiones epistémicas contextuales no son necesariamente provisorias; no son decisiones que un agente tome temporariamente hasta tanto decida qué creencias adoptar. Por el contrario, considero que la existencia permanente de múltiples contextos de decisión es parte esencial del estado epistémico habitual de un agente. Por supuesto, cómo referirse a dichos contextos constituye una cuestión terminológica menor, siempre y cuando no se generen connotaciones indeseables. Para evitar dichas connotaciones es que prefiero evitar hablar de “decisiones cognitivas hipotéticas” o de “decisiones bajo supuestos”. Hablaré en cambio de “contextos de aceptación”, o “contextos de decisión cognitiva”, que son expresiones más neutrales, así como de “elementos aceptados contextualmente”.

En vista de la existencia de múltiples contextos, el principio (EP) formulado más arriba resulta demasiado restringidos puesto que un problema de decisión cognitiva no siempre dará como resultado un *cambio de creencias*. En la próxima sección trataré con un poco más de detalle la distinción entre creencia y aceptación que acabo de sugerir, y propondré un nuevo principio fundamental para el modelo peirceano de epistemología. El resultado será una concepción más flexible de la expansión inductiva y de la inferencia a la mejor explicación.

## 5. Aceptar y creer

En esta sección ofreceré el esbozo de una teoría formal basada en las ideas sugeridas en la sección 4.

A la hora de describir el estado epistémico de un agente, propongo distinguir entre un conjunto  $K$  de creencias en sentido estricto y un conjunto más amplio de elementos aceptados, al que llamaremos  $M$ . Tanto “creencias” como “aceptaciones” son representadas por oraciones de un lenguaje proposicional  $L$ .  $M$  es consistente e incluye a  $K$  como parte propia. Diremos que  $M$  es el conjunto de elementos que el agente se ha comprometido a tomar como base de acción y deliberación para *algún* contexto.

Sea  $A_c$  el conjunto de aceptaciones para el contexto  $c$ .  $A_c$  está conformado por aquellos elementos con los que el agente se ha comprometido para  $c$  (esto es, que toma como verdaderos en  $c$ ). Se requerirá además que  $A_c$  sea deductivamente cerrado y consistente.  $K$  es la intersección de todos los conjuntos de aceptaciones, para todo contexto, y  $M$  es la unión de tales conjuntos. Formalmente:

- $A = Cn(A) \subseteq M$  es el conjunto de elementos aceptados como verdaderos en el contexto  $c$ .
- $K = \bigcap_c A_c$ ; y  $M = \bigcup_c A_c$ , para todo  $c$ .

Ahora bien, en muchas teorías estándar de dinámica epistémica, se supone que a la hora de expandir un conjunto de *creencias*, un agente necesita ordenar de alguna manera los elementos inciertos, esto es, los elementos compatibles con  $K$  pero que están fuera de  $K$ . Esto se consigue, por ejemplo, si suponemos que el agente cuenta con una función de probabilidad subjetiva dado  $K$ . Se recordará que en la sección 3 llamamos a esta función  $P_K$ . Es claro que todos los elementos de  $K$ , o creencias, tienen probabilidad 1 según  $P_K$ . Algunos de los elementos fuera de  $K$ , pero compatibles con  $K$ , podrían también tener probabilidad 1 según  $P_K$ , aunque ello no es necesario.

Pues bien, en el nuevo esquema que sugiero aquí, supondremos que el agente cuenta con distintas funciones de probabilidad para cada contexto. Diremos que  $P_{Ac}$  es la función de probabilidad subjetiva *relativa al conjunto de aceptaciones del contexto  $c$* . El hecho de contar con diferentes funciones para cada contexto, evidentemente, tendrá consecuencias a la hora de llevar a cabo un ejercicio de decisión cognitiva y evaluar un posible cambio epistémico.

Así, propongo analizar el cambio epistémico de la siguiente manera: al participar en un problema de decisión cognitiva, un agente busca expandir  $A$ , para cierto contexto  $c$ , lo cual puede llevar a expandir también  $M$ . Más precisamente a la hora de participar en un problema de decisión cognitiva el agente usará como base para su deliberación algún conjunto  $A \subseteq M$ . Por consiguiente, el agente utilizará alguna función de probabilidad  $P_{Ac}$  que asigne probabilidad 1 a los elementos de  $A$ , y que asigne algún número en el intervalo  $[0,1]$  a los elementos compatibles con pero que no pertenecen a  $A$ . El resultado de una expansión por  $H$ , para  $H$  consistente con  $M$ , será

entonces un nuevo conjunto  $A'c = Cn(Ac \cup \{H\})$ . En este caso, si  $H$  no estaba ya en  $M$  se habrá ampliado el conjunto total de aceptaciones del agente, pero no necesariamente su conjunto de creencias (a menos que se trate del caso especial en que  $A_c = K$ )

En fin, la aceptación es contextual, y el agente no sabe cómo reaccionará en otro contexto  $c'$  frente a la posibilidad de incorporar  $H$ , no solamente porque  $P_{AC'}$  puede diferir de  $P_{AC}$ , sino porque el índice de cautela que el agente decidió usar en  $c$  puede no ser el mismo que está dispuesto a usar en  $c'$ .

La idea de que un agente puede tener más de una función de probabilidad subjetiva no es demasiado novedosa. Sin ir más lejos, en la propuesta de Levi el agente cuenta en realidad con un *conjunto convexo* de tales funciones.<sup>29</sup> Ello intenta recoger la intuición de que los agentes raramente asignan valores numéricos precisos a sus incertidumbres. Supongamos que Juan tiene un alto grado de confianza en que mañana lloverá, aunque no tiene certeza total. Nos parecería extraño, sin embargo, si Juan nos dijera que tiene una confianza de grado 0,957 en que mañana habrá lluvias. La confianza de los agentes más bien se refleja en *intervalos* numéricos; Juan podría afirmar, por ejemplo que su nivel de confianza en que mañana lloverá se ubica entre 0,8 y 1. Pues bien, Levi muestra que esto equivale a reconocer que Juan cuenta en verdad con un *conjunto* de funciones de probabilidad, tal que cada una de ellas atribuye un número preciso entre 0,8 y 1 a la proposición de que habrá lluvia mañana. Sin embargo, en el esquema de Levi todas estas funciones son relativas al conjunto  $K$ . Es decir que la “base de acción y deliberación” es siempre el conjunto  $K$ . Yo propongo en cambio permitir que dicha base cambie de acuerdo a los contextos.

---

<sup>29</sup> Un conjunto  $\Delta$  no vacío de funciones de probabilidad subjetiva es *convexo* si y sólo si: si  $P, P' \in \Delta$ , entonces todo  $P''$  que resulte de un promedio ponderado entre  $P$  y  $P'$  también pertenece a  $\Delta$ . Véase Levi (1997), cap. 6, p. 127.

En verdad, adaptando la idea de Levi, bien podríamos suponer que el agente cuenta con un conjunto convexo de funciones (en lugar de una única función) *para cada contexto*. Si realizamos esta suposición adicional, podemos obtener maneras alternativas de elucidar la idea de aceptación contextual. Por ejemplo, podríamos tomar *como punto de partida* el conjunto de probabilidades subjetivas del agente, y con base en ellas determinar cuáles son los contextos posibles para el agente, y cuáles los elementos aceptados para cada contexto. Así, cada contexto podría muy bien ser caracterizado por cierto conjunto de probabilidades. Veamos cómo:

Sea  $\Delta$  un conjunto convexo<sup>30</sup> de funciones de probabilidad subjetiva  $P_n$  sobre  $L$ , y sea  $\mathbf{T}$  el conjunto de todas las teorías  $T_i$  (conjuntos deductivamente cerrados) de  $L$ . Entonces, puede definirse una función  $f: \Delta \rightarrow \mathbf{T}$  tal que  $f(P_n) = T_i$  si y sólo si para todo elemento  $a \in T_i$ ,  $P_n(a) = 1$ , y para toda otra teoría  $T_j$ , si  $P_n(a) = 1$  para todo  $a \in T_j$ , entonces  $T_j \subseteq T_i$ . Esto es, cada elemento  $P_n$  de  $\Delta$  selecciona algún miembro de  $\mathbf{T}$ , a saber, la mayor teoría  $T$  para la cual  $P_n(a) = 1$ , para todo  $a$  en  $T$ . Típicamente, algunos miembros de  $\Delta$  seleccionarán el mismo elemento en  $\mathbf{T}$  (y algunas teorías en  $\mathbf{T}$  no serán seleccionadas en absoluto, por supuesto). Un conjunto de aceptaciones contextuales será entonces una teoría  $T$  de  $\mathbf{T}$  que es seleccionada por algún miembro de  $\Delta$ . Un contexto  $c$  puede caracterizarse mediante el subconjunto  $\Delta' \subseteq \Delta$  compuesto por todos aquellos elementos  $P_n$  que seleccionan la teoría  $T = A_c$ . Dicho de otra manera, cada contexto puede identificarse con el conjunto de funciones de probabilidad subjetivas del agente que determinan unívocamente cierto conjunto de aceptaciones (a saber, el conjunto de aceptaciones para dicho contexto).

---

<sup>30</sup> Para una explicación de la propiedad de convexidad, véase la nota anterior.

Luego podemos definir  $K$  como la mayor teoría incluida en todo contexto. Es claro que entonces toda función  $P_n$  asigna probabilidad 1 a todo elemento de  $K$ , como debe ser. Por otra parte, no es necesario que haya una función  $P_n$  que atribuya probabilidad 1 a todo elemento de  $M$ , si bien todo elemento de  $M$  recibe probabilidad 1 por parte de alguna (no necesariamente la misma) función  $P_n$ .

En este escenario, registrar un cambio en un conjunto de elementos aceptados es en verdad una manera diferente de expresar que ha ocurrido un cambio en el conjunto de probabilidades subjetivas del agente, esto es, que ha tenido lugar un cambio en  $\Delta$ . Dicho cambio, sin embargo, no necesariamente será el resultado de un cambio en  $K$ , esto es, de un cambio en sus *creencias* propiamente dichas.

La posibilidad de esta última elucidación es de alguna importancia, según creo. Por un lado, permite simplificar los elementos primitivos del sistema: todo lo que necesitamos es suponer que un agente cuenta con un conjunto de funciones de probabilidad. Por otro lado, disipa el posible recelo que la idea de contexto puede generar, puesto que muestra que hablar en términos de aceptaciones contextuales no nos compromete con entidades dudosas, ni está reñido con la posibilidad de ofrecer un análisis riguroso.

Si el análisis anterior es plausible, entonces debemos considerar un nuevo principio fundamental de la epistemología peirceana. A saber:

*EP': Los agentes pueden modificar su conjunto de elementos aceptados sólo si existe justificación adecuada para dicho cambio. Conversamente, los agentes deberán modificar tal conjunto toda vez que haya razones de peso para ello. Además, los agentes no deben preocuparse por justificar elementos previamente aceptados.*

Obsérvese que obtenemos a (EP) como caso límite de (EP').

¿Qué podemos concluir de toda esta discusión respecto de nuestro problema original, esto es, el problema de elucidar la idea de Inferencia a la Mejor Explicación? En la sección 2 propuse entender una IME como un proceso que culmina con una modificación del estado epistémico de un agente. Creo que ahora es claro cómo ha de interpretarse dicha modificación. Típicamente, una IME se formula para resolver un problema específico, y *en el marco de un contexto específico*, lo que supone un conjunto particular de funciones de probabilidad subjetivas, y, como consecuencia, un conjunto particular de aceptaciones. Además, en general el índice de cautela que el agente está dispuesto a usar no será trasladable a otros contextos, aún si se tratara de evaluar, en un contexto diferente, el *mismo* conjunto de hipótesis, esto es, la misma partición última. Por todo ello, el resultado de una IME será una nueva aceptación del agente, aunque no necesariamente una nueva creencia.

## **6. Inferencia a la mejor explicación y realismo científico**

La mayoría de los estudios filosóficos sobre 1MB aparecidos recientemente realiza algún tipo de evaluación sobre la relación entre 1MB y la posición epistemológica conocida como “realismo científico”. De hecho, en tiempos recientes raramente encontramos publicaciones de alguna de las dos temáticas sin referencia a la otra. El presente estudio no estaría completo, pues, sin al menos una breve mención sobre un posible vínculo con el realismo. Por otra parte, algunos autores han argumentado que la relación entre 1MB y realismo brinda un motivo de sospecha sobre la legitimidad de las 1MB, de modo que dicho vínculo nos presenta en verdad con un problema al que hay

que dar alguna solución.<sup>31</sup> En los párrafos que siguen espero mostrar que la elucidación propuesta en este trabajo permite evaluar el supuesto compromiso realista de las IME desde una nueva perspectiva.

Ante todo, es conveniente introducir algunas aclaraciones. La posible conexión entre IMB y realismo puede aludir a dos cuestiones muy diferentes que no deben confundirse entre sí. Lamentablemente, rara vez han sido enunciadas con claridad como cuestiones independientes. Una de ellas es si la idea misma de IME se compromete con una posición realista; otro problema es si es posible defender el realismo científico usando una estructura argumentativa que responda al esquema de una IMB. Como ejemplos de autores involucrados en la primera polémica, podemos citar a Bas van Fraassen (en calidad de defensor de la estrecha conexión entre IMB y realismo)<sup>32</sup> y a Arthur Fine (en el papel de escéptico respecto del compromiso realista de una IME).<sup>33</sup> Como ejemplos de la segunda polémica, encontramos a Richard Boyd<sup>34</sup> (defendiendo la legitimidad de una estructura de IME para defender al realismo) y a Peter Lipton<sup>35</sup> (en

---

<sup>31</sup> Hay otro problema importante asociado con las IMB; a saber, se ha argumentado que las IMB constituirían un patrón incoherente de cambio epistémico. Esta objeción se debe esencialmente a van Fraassen (1989, p. 166 y ss.), y hace uso de un argumento más general que se conoce como “*Dutch Book* diacrónico”. Dicho argumento aparece por primera vez en Teller (1973), quien recoge una idea original de David Lewis. Lamentablemente no tengo espacio en este artículo para desarrollar el problema con el detalle que merece, de modo que quedará para otra oportunidad. Una estrategia posible para desactivar la objeción consiste en intentar invalidar la legitimidad de los *Dutch Books* diacrónicos en general. Existe una muy amplia bibliografía sobre el tema. Como muestra, véanse Douven (1999), Machina (1991), Maher (1993), cap. 5, o Levi (1987).

<sup>32</sup> Van Fraassen (1989). Como es bien conocido, van Fraassen mantiene una perspectiva crítica hacia el realismo científico y, por consiguiente, hacia las IMB.

<sup>33</sup> Fine ha expresado en diversas ocasiones que el resultado de una IMB bien podría ser una hipótesis que “salve los fenómenos”, al estilo instrumentalista (cf. por ejemplo Fine (1984)). Fine, sin embargo, no distingue ambas polémicas del modo en que las hemos presentado aquí, y también podría ser ubicado como escéptico en la segunda polémica que menciono en el trabajo.

<sup>34</sup> Por ejemplo en Boyd (1983).

<sup>35</sup> Véase por ejemplo Lipton (1991), cap. 9. Cabe observar que Lipton sí cree que hay argumentos para defender al realismo científico. Pero concede que un antirrealista no podría ser convencido por medio de un argumento que apele a la idea de que la verdad de las teorías exitosas constituye *la mejor explicación* de su éxito (pp. 160-168).

el papel de escéptico respecto de dicha legitimidad). Para asegurarnos que ambas discusiones son independientes, basta notar que, al menos en principio, resultaría legítimo apelar a una IME para apuntalar al realismo *si y sólo si* mantenemos la neutralidad epistémica de las IME, so pena de circularidad. Aquí me ocuparé brevemente de la primera polémica, pero no diré nada acerca de la segunda.

Consideremos, pues, la posibilidad de que la IME se comprometa intrínsecamente con el realismo científico. Adoptar un compromiso realista no es necesariamente un problema, por supuesto. Sin embargo, si la conexión es tan fuerte como algunos autores parecen creer, la suerte de toda IMB está atada a la del realismo; todo aquel que encuentre a este último insatisfactorio rechazará igualmente las IME.

Esta es precisamente la posición de van Fraassen. He aquí una reconstrucción de uno de los argumentos presentados por van Fraassen en (1989):

- (1) El realismo científico es la posición que se compromete con la verdad de las mejores explicaciones disponibles, o, al menos, con su verosimilitud.
- (2) Una IME es esencialmente una regla de cambio epistémico según la cual las mejores explicaciones disponibles deben incorporarse al conjunto de creencias de un agente, y por ende deben adoptarse como verdaderas (para todo contexto).
- (3) Pero no hay garantías para afirmar que las mejores explicaciones disponibles sean verdaderas, ni para afirmar que puedan estar remotamente cerca de la verdad en algún sentido interesante: el realismo científico es una posición infundada.
- (4) Por ende, toda regla que recomiende *creer* las mejores explicaciones disponibles es una mala regla de cambio epistémico.
- (5) La IME es una mala regla de cambio epistémico.<sup>36</sup>

---

<sup>36</sup> Más adelante van Fraassen considera una regla de IME probabilística, e intenta desacreditarla mediante el llamado argumento del *Dutch Book* diacrónico. Véase la nota 31 de este trabajo.

Pero de acuerdo con el análisis propuesto en este artículo, la premisa (2) es falsa, de modo que el argumento queda desactivado, independientemente de que aceptemos o rechacemos la premisa (3). El vínculo entre IME y realismo sólo se sostiene bajo el supuesto de que el producto de un problema de decisión cognitiva es siempre la recomendación de agregar una nueva creencia al cuerpo epistémico del agente. Pero se ha argumentado que un problema de decisión cognitiva no necesariamente culmina con un nuevo conjunto de creencias; de hecho, en los casos típicos tan sólo se obtiene una nueva aceptación contextual. En síntesis, el mecanismo de una IME no fuerza a un agente a adoptar una nueva *creencia*, si bien tampoco lo prohíbe. El uso de IMB como una regla para el cambio epistémico no nos compromete con una epistemología realista.

## **7. Conclusión**

El concepto de inferencia a la mejor explicación se relaciona estrechamente con muchos de los temas centrales de la epistemología contemporánea, lo cual lo transforma en un objeto de estudio particularmente interesante. Comencé por observar que en numerosas ocasiones lo encontramos confundido con la idea de abducción. Argumenté aquí que la identificación entre ambos conceptos debe ser evitada. Sugerí en cambio reservar el nombre de abducción para una de las dos etapas que componen el proceso de una IME.

En la primera etapa, o etapa abductiva, se generan las explicaciones potenciales que pueden dar respuesta a algún interrogante del agente. Quedará para otra oportunidad un análisis más pormenorizado de este ámbito. En la segunda etapa, o etapa de selección de hipótesis, se lleva a cabo la elección de la mejor (o mejores) hipótesis disponibles. Propuse llevar a cabo dicha selección con ayuda de algunos instrumentos de teoría de la

decisión cognitiva. Para ello me basé en la propuesta de expansión inductiva de Isaac Levi. Argumenté que la teoría de Levi presupone cierto modo peculiar de entender los lineamientos básicos de una epistemología pragmatista de tipo peirceano, y argumenté también que dicha interpretación de la epistemología peirceana es demasiado restringida. Más precisamente, argumenté que un principio básico de epistemología pragmatista no debe ocuparse de regular el cambio racional de *creencias*, sino que debe limitarse a reglamentar el cambio racional de elementos aceptados, o *aceptaciones*.

La razón es la siguiente. Comencé por defender la conveniencia de elucidar la idea pragmatista de creencia en términos de un “compromiso epistémico para todo contexto de acción o deliberación”. Ahora bien, a la hora de llevar a cabo una investigación, nos encontramos con que los agentes muchas veces toman en consideración elementos esencialmente contextuales, que raramente pueden ser generalizados para todo contexto. Es decir que los agentes típicamente evalúan las ventajas o desventajas de comprometerse con cierta hipótesis para el contexto particular en el cual se lleva a cabo la investigación. De ello resulta que, en general, una explicación potencial podrá ser aceptada o rechazada *para el contexto en cuestión*.

De modo que propuse una manera más amplia de entender la epistemología peirceana y, correspondientemente, una versión un tanto más compleja de cómo procede un mecanismo de selección de hipótesis o explicaciones potenciales. Para ello, esbocé un modelo epistémico que toma como central la idea de conjuntos de elementos aceptados para diferentes contextos.  $K$ , o el conjunto de creencias propiamente dichas del agente, es la intersección de la totalidad de tales conjuntos. Dado que un proceso de decisión cognitiva es típicamente relativo a cierto contexto  $c$ , el resultado de llevar a cabo una inferencia a la mejor explicación es un nuevo conjunto de aceptaciones para algún  $c$ . También sugerí que cada contexto puede identificarse con un conjunto de

funciones de probabilidad subjetivas del agente, las que determinan unívocamente cierto conjunto de aceptaciones (a saber, el conjunto de aceptaciones para dicho contexto). Argumenté, finalmente, que mi propuesta permite revelar que una inferencia a la mejor explicación es en principio un mecanismo de cambio epistémico neutral respecto del realismo científico.

Por supuesto, aún queda mucho por hacer. Por ejemplo, no me ocupé de dar cuenta de las posibles contracciones del estado epistémico de un agente. Por otra parte, el presente trabajo podría ser suplementado con un análisis de cuándo una hipótesis es en verdad *explicativa*, lo cual correspondería, de acuerdo con el esquema sugerido, a un estudio de la etapa abductiva de las IME. Confío, sin embargo, en que este trabajo pueda verse como un primer paso hacia la construcción de una teoría general de cambio epistémico, que sea además capaz de tratar satisfactoriamente el problema de la elección entre hipótesis rivales empíricamente adecuadas.

## **Bibliografía**

- BOYD, Richard (1983), "On the Current Status of the Issue of Scientific Realism", *Erkenntnis*, 19, pp. 45-90. COHEN, Jonathan (1992), *An Essay on Belief and Acceptance*, Oxford, Clarendon Press.
- DANCY, Jonathan, y SOSA, Ernest (1993), *A Companion to Epistemology*, Oxford, Blackwell.
- DOUVEN, Igor (1999), "Inference to the Best Explanation Made Coherent", *Philosophy of Science*, 66 (Proceedings), pp. S424-S435.

- FLACHS, Peter y KAKAS, Anthony (2000), *Abduction and Induction: Essays on their Relation and Integration*, Dordrecht, Kluwer. FINE, Arthur (1984), “The Natural Ontological Attitude”, reimpresso en Papineau, David (ed.) *The Philosophy of Science*, Oxford, 1996, pp. 21-44.
- FUHRMANN, André (1997), *An Essay on Con traction*, Stanford, California, CSLI Publications.
- GÁRDENFORS, Peter (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Cambridge, Mass., The MIT Press.
- HANSON, Norwood R. (1958), *Patterns of Discovery*, Cambridge, Cambridge University Press.
- HANSSON, Sven Ove (1999), *A Textbook on Belief Dynamics: Theory Change and Database Updating*, Dordrecht/ Boston! London, Kluwer.
- HARMAN, Gilbert (1965), “The Inference to the Best Explanation”, *The Philosophical Review*, vol. 74, 1, pp. 88-95.
- HARMAN, Gilbert (1968), “Enumerative Induction as Inference to the Best Explanation”, *The Journal of Philosophy*, 65(18), pp. 529-533.
- JAMES, William (1897), *The Will to Believe and Other Essays in Popular Philosophy*. Reimpresso en New York, Dover, 1956.
- JOSEPHSON, John y JOSEPHSON Susan (1994), *Abductive Inference: Com putation, Philosophy, Technology*, Cambridge, Cambridge University Press.
- MACHINA, Mark (1991), “Dynamic Consistency and Non-Expected Utility”, en: Bacharach, Michael y Hurley, Susan, *Foundations of Decision Theory: Issues and Advances*, Oxford, Blackwell.
- MAHER, Peter (1993), *Betting on Theories*, Cambridge, Cambridge University Press.

- LAUDAN, Larry (1981), "Why was the Logic of Discovery Abandoned?", en: *Science and Hypothesis: Historical Essays on Scientific Methodology*, Dordrecht, D. Reidel, cap. 11, pp. 181-191.
- LEVI, Isaac (1984), *Decisions and Revisions*, Cambridge, Cambridge University Press.
- LEVI, Isaac (1987) "The Demons of Decision", *The Monist*, 1987, pp.193- 211.
- LEVI, Isaac, (1991), *The Fixation of Belief and Its Undoing: Changing Beliefs Through Inquiry*, Cambridge, Cambridge University Press.
- LEVI, Isaac (1997), *The Covenant of Reason*, Cambridge, Cambridge University Press.
- LEVI, Isaac (2001), "Inductive Expansions and Non-monotonic Reasoning", en: Rott, Hans y Williams, Mary Ann, *Belief Revision*, Dordrecht, Kluwer.
- LIPTON, Peter (1991), *Inference to the Best Explanation*, London, Routledge.
- PAGNUCCO, Maurice (1996), *The Role of Abductive Reasoning within the Process of Belief Revision*, Tesis de Doctorado, Basser Department of Computer Science, University of Sydney, Australia.
- PEIRCE, Charles Sanders (1877), "The Fixation of Belief", reimpresso en: Houser, N. y Kloesel, Ch., *The Essential Peirce*, Vol.I, Indiana University Press, cap. 7, 1992, pp. 109-123.
- PEIRCE, Charles Sanders (1878), "Deduction, Induction and Hypothesis", reimpresso en: Houser, N. y Kloesel, Ch., *The Essential Peirce*, vol. 1, Indiana University Press, cap. 12, 1992, pp. 186-199.
- PEIRCE, Charles Sanders (1908), "A Neglected Argument for the Reality of God", reimpresso en: *Selected Writings*, New York, Dover, 1966, pp. 358-379.
- PEIRCE, Charles S. (1958), *Collected Papers of Charles Sanders Peirce*, Cambridge, Mass., Harvard University Press. Editado por C. Harstshorne, P. Weiss. y A. Burks.

SAVAGE, Leonard J. (1954), *The Foundations of Statistics*, 2da. edición, New York, Dover, 1972.

TELLER, Paul (1973), “Conditionalization and Observation”, *Synthese*, 26, pp. 218-258.

THAGARD, Paul (1978), “The Best Explanation: Criteria for Theory Choice”, *Journal of Philosophy*, vol. 75, 2, pp. 76-92.

VAN FRAASSEN, Bas (1989), *Laws and Symmetry*, Oxford, Clarendon Press.